

2<sup>nd</sup> BSC International  
Doctoral Symposium

2015

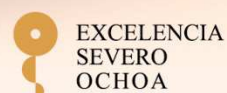
5th, 6th & 7th May, 2015

Book of abstracts



**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*



**EXCELENCIA  
SEVERO  
OCHOA**



*Book of Abstracts*  
**2nd BSC International Doctoral Symposium**

*Editor and Layout Design:*  
**Montserrat González Ferreiro**

*Graphic and Cover Design:*  
**Laura Bermúdez Guerrero**

*This is an open access book registered at UPC Commons  
(<http://upcommons.upc.edu>) under a Creative Commons license  
to protect its contents and increase its visibility.*

*This book is available at*  
[www.bsc.es/doctoral-symposium-2015](http://www.bsc.es/doctoral-symposium-2015)

*published by:*  
**Barcelona Supercomputing Center**

*supported by:*  
**The “Severo Ochoa Centres of Excellence” programme**

2<sup>nd</sup> Edition, 11<sup>th</sup> May 2015





## ACKNOWLEDGEMENT

The BSC Education & Training team gratefully acknowledges all the PhD candidates, Postdoc researchers, experts and specially the Keynote Speaker Nuria Oliver and the training lecturer Josep Lluís Pelegrí, contributing to this Book of Abstracts and participating in the 2nd BSC International Doctoral Symposium. We also wish to expressly thank the volunteers that supported the celebration of the event: Claudia Rosas, Luna Backes, Matías Rivero and Dani Gallart.

BSC Education & Training team  
education@bsc.es





## CONTENT

EDITORIAL COMMENT .....	8
WELCOME ADDRESS.....	10
PROGRAMME.....	12
KEYNOTE SPEAKER.....	18
LECTURER & TRAINING.....	19
TALK PRESENTERS .....	21
POSTER PRESENTERS.....	34
EXTENDED ABSTRACTS.....	49





## EDITORIAL COMMENT

We are proud to present the Book of Abstracts for the 2nd BSC International Doctoral Symposium.

During more than ten years, the Barcelona Supercomputing Center has been receiving undergraduate, master and PhD students, and providing them training and skills to develop a successful career. Many of those students are now researchers and experts at BSC and in other international research institutions.

In fact, the number of students has never decreased. On the contrary, their number and research areas have grown and we noticed that these highly qualified students, especially the PhD candidates, needed a forum to present their findings and fruitfully exchange ideas. As a result, in 2014, the first BSC Doctoral Symposium was born and held during March of the same year.

Last year, a total of 17 presentations were given, 10 posters were exhibited, 2 short trainings on legal protection of computer programmes and how to do a good presentation were conducted; and we reached more than 50 attendees.

In this second edition of the BSC Doctoral Symposium we have been even more ambitious. We have opened the participation to students all over the world (and thus, incorporated the International designation) and have succeeded enrolling students from different countries. Moreover, the talks scheduled have increased from 17 to 34, and the posters to be presented are 32 instead of 10. Also, we have planned a keynote speaker' talk, and an extensive training course on scientific writing.

The talks will be held in seven different sessions and will tackle the topics of: Postdoc research at BSC; Algorithms, Physics & Data Science; Scheduling, Interconnections and Simulations; Computer Sciences and Applications; Earth Sciences and Physics; and Life Sciences. The posters will be exhibited and presented during five poster sessions that will give the authors the opportunity to explain their research and results.

The keynote speaker Nuria Oliver will share with us her deep experience on scientific research in Telefonica R&D. She is the Scientific Director and founder of the User, Data and Media Intelligence research team in Telefonica Research, besides being an active researcher and has written over 100 scientific papers in international conferences, journals and book chapters. Her work has been widely recognized by the scientific community with over 8100 citations.

The training course will be an 8-hour course that will review the fundamentals of scientific writing and provide practice regarding these concepts. The lecturer is an experienced scientist, Josep Lluís Pelegrí, who has supervised 14 doctoral dissertations and published about 90 SCI papers.

This Book of Abstracts follows the order of the programme of the 2<sup>nd</sup> BSC International Doctoral Symposium, and all the information is arranged according to the order of participation of the students. Information about the participants and accepted extended abstracts are published after a rigorous review process and with permission of the authors.

We hope that this publication is a valuable contribution to the reflection and dissemination of how the usage of high performance computing technologies and resources are key-factors to promote the scientific and social development in many different areas.

BSC Education & Training team  
education@bsc.es







## WELCOME ADDRESS

I am delighted to welcome all the PhD students, Postdoc researchers, advisors, experts and attendees participating in the 2nd BSC International Doctoral Symposium.

This year edition has reached international relevancy and students from different countries and organizations will take part in the symposium. Nevertheless, the goal of the event continues to be providing a framework to share research results of the projects developed by PhD thesis that use High Performance Computing in some degree. Additionally, we will offer an 8-hours training on one skill that will be useful to all the participants as researchers and professionals.

The symposium was conceived in the framework of the Severo Ochoa Program at BSC, following the project aims regarding the talent development and knowledge sharing. Keeping that in mind, the symposium provides an interactive forum for PhD students considering both the ones just beginning their research and others who have developed their research activities during several years.

As a consequence, I highly appreciate the support provided by BSC and the Severo Ochoa Center of Excellence Programme that make possible to celebrate this event.

I must add that I am very grateful to the BSC directors for supporting the symposium, to the group leaders and to the worldwide advisors for encouraging the participation of the students in the event. Moreover, I wish to specially thank the keynote speaker Nuria Oliver and the invited lecturer Josep Lluís Pelegrí, for their willingness to, respectively, share with us their deep experience on scientific research in Telefonica R&D and lecture us on how to improve the scientific writing.

And last but not least, I would like to thank all PhD students and Postdoc researchers for their presentations and effort. I wish you all the best for your career and I really hope you enjoy this great opportunity to meet other colleagues and share your experiences.

Dr. Maria Ribera Sancho  
Manager of BSC Education & Training





# PROGRAMME

2nd International BSC Doctoral Symposium 2015 – 5th May		
Start time	Activity	Speaker/s
9.00h	Welcome and opening	Mateo Valero, BSC Director
9.15h	Keynote Speaker talk: Research in Telefónica R&D	Nuria Oliver, Scientific Director Telefónica R&D
10.10h	First Poster Session & coffee break <i>Heating Bulk Ions in DEMO with ICRF waves</i> Dani Gallart, BSC Computer Applications in Science & Engineering Dept. <i>Rayleigh wave ellipticity measurements in the Iberian Peninsula and Morocco</i> Clara Gómez-García, Spanish National Research Council <i>Implementation of Dynamic Aerosol-Radiation Interaction within the NMMB/BSC-CTM</i> Vincenzo Obiso, Earth Sciences Dept. <i>State-of-the-Art Climate Predictions for Energy Climate Services</i> Verónica Torralba, Institut Català de Ciències del Clima <i>Identification of novel type 2 diabetes susceptibility loci through wholegenome imputation using sequencing based reference panels into 13.201 cases and 59.656 controls</i> Sílvia Bonàs Guarch, BSC Life Sciences Dept. <i>Parallel programming issues and what the compiler can do to help</i> Sara Royuela, BSC Computer Sciences Dept.	
First Talk Session: Postdoc research at BSC (first part)		
10.40h	Methodology to predict scalability of parallel applications	Claudia Rosas, BSC Computer Sciences Dept.
11:00h	Mesh generation to simulate incompressible Navier-Stokes equations with wind turbines	Abel Gargallo-Peiró, BSC Computer Applications in Science & Engineering Dept.
11.20h	Measurements and computer simulations: whom do we trust for atmospheric composition?	Enza Di Tomaso, BSC Earth Sciences Dept.
11.40h	HPC and edge elements for geophysical electromagnetic problems: an overview	Octavio Castillo, BSC Computer Applications in Science & Engineering Dept.
12.00h	Second Poster Session & coffee break <i>Automatic Query Driven Data Modelling in Cassandra</i> Roger Hernandez, BSC Computer Sciences Dept. <i>Modelling the Contact Propagation of Nosocomial Infection in Emergency Departments</i> Cecilia Jaramillo, Universitat Autònoma de Barcelona <i>MACC: Mercurium ACcelerator Model</i> Guray Ozen, BSC Computer Sciences Dept. <i>Capturing the action of current antipsychotic drugs at G protein-coupled receptors (GPCRs) by means of Markov State Model analysis</i> Ismael Rodríguez-Espigares, Universitat Pompeu Fabra (UPF) - Hospital del Mar Medical Research Institute (IMIM) <i>A 3D-1D cardiac-vascular computational feedbacked model</i> Alfonso Santiago, BSC Computer Applications in Science & Engineering Dept.	



<b>2nd International BSC Doctoral Symposium 2015 – 5th May</b>			
<b>Start time</b>	<b>Activity</b>	<b>Speaker/s</b>	
<b>First Talk Session: Postdoc research at BSC (second part)</b>			
12.20h	From imaging to simulation: a framework applied to simulate the blood flowing the carotids	Ruth Arís, BSC Computer Applications in Science & Engineering Dept.	
12:40h	Optimizing a coarse-grained model for the recognition of protein-protein binding	Agustí Emperador, BSC Life Sciences Dept.	
13.00h	Lunch break		
15.00h	Training An Introduction to Scientific Writing starts; Lecturer Josep Lluís Pelegrí, ICM-CSIC		
17.00h	Coffee break		
17.15h	Training An Introduction to Scientific Writing continues; Lecturer Josep Lluís Pelegrí, ICM-CSIC		
19.00h	Adjourn		



2nd International BSC Doctoral Symposium 2015 – 6th May		
Start time	Activity	Speaker/s
9.00h	Opening of the second day	
Second Talk Session: Algorithms, Physics & Data Science		
9.10h	Study of Preconditioners based on Markov Chain Monte Carlo Methods	Oscar A. Esquivel (Postdoc), CONACYT-BSC Computer Sciences Dept.
9.30h	Supercomputing simulations for Beyond Standard Model theories at the TeV scale: non-standard gravity and electroweak interactions	Rafael Delgado, Universidad Complutense de Madrid
9.50h	Discovering most significant news using Network Science approach	Ilya Blokh, Perm State University
10.10h	Linompss – A Linear Algebra Library on OMPs	Sicong Zhuang, BSC Computer Sciences Dept.
10.30h	Third Poster Session & coffee break Evaluation of Modelling Systems in High Resolution to Assess the Air Pollutant Impacts on Human Health Sergio N. González (Postdoc), BSC Earth Sciences Dept. - CONACYT. <i>Binding Free Energy and Ligand Orientation Calculations using A Monte Carlo Method with Markov State Analysis</i> Daniel Lecina, BSCLife Sciences Dept. <i>Asynchronous parallel fluid solver and particle transport</i> Edgar Olivares, BSC Computer Applications in Science & Engineering Dept. <i>A parallel coupled algorithm for the solution of deformable two-body contact problem</i> Matias Rivero, BSC Computer Applications in Science & Engineering Dept. <i>Resource Management for Software Defined Data Centers for Heterogeneous Infrastructures</i> Marcelo Amaral, BSC Computer Sciences Dept. <i>A framework for multidimensional indexes on distributed and highly-available data stores</i> Cesare Cugnasco, BSC Computer Sciences Dept. <i>Mathematical Representation of the Hardware Round-Robin Scheduler Analytical Model for Single-ISA Heterogeneous Architectures</i> Daniel Nemirovsky, BSC Computer Sciences Dept.	
Third Talk Session: Scheduling, Interconnections and Simulations		
11.20h	Criticality-Aware Dynamic Task Scheduling for Heterogeneous Systems	Kallia Chronaki, BSC Computer Sciences Dept.
11.40h	Interconnect Energy Savings on MapReduce Clusters	Renan Fischer e Silva, BSC Computer Sciences Dept.
12.00h	Efficient Address Translation for Memory Intensive Workloads	Vasileios Karakostas, BSC Computer Sciences Dept.
12.20h	High Performance Computing Based Simulation for Healthcare Decision Support	Zhengchun Liu, Universitat Autònoma de Barcelona
12.40h	Hardware Scheduling Algorithms for Asymmetric Single-ISA CMPs	Nikola Markovic, BSC Computer Sciences Dept.
13.00h	Lunch break	
15.00h	Training An Introduction to Scientific Writing continues; Lecturer Josep Lluís Pelegrí, ICM-CSIC	
17.00h	Coffee break	
17.15h	Training An Introduction to Scientific Writing continues; Lecturer Josep Lluís Pelegrí, ICM-CSIC	
19.00h	End of the Training and Adjourn	



2nd International BSC Doctoral Symposium 2015 – 7th May			
Start time	Activity	Speaker/s	
9.00h	Opening of the third day		
<b>Fourth Talk Session: Life Sciences</b>			
9.10h	Macroscopic structures generated by microorganisms swimming in a fluid	Francisco Alarcón, Universitat de Barcelona	
9.30h	Assessment of scoring functions performance to re-rank docking decoys from FFT(Fast Fourier Transform) programs	Didier Barradas, BSC Life Sciences Dept.	
9.50h	Implementation of an Internal Coordinates Anisotropic Network Model in PELE	Victor A. Gil, BSC Life Sciences Dept.	
10.10h	Inter-residue interactions in membrane proteins	Eduardo Mayol, Universitat Autònoma de Barcelona	
10.30h	<p>Fourth Poster Session &amp; coffee break</p> <p><i>Galaxy Gears: Web Services integration into Galaxy workflows</i> Dmitry Repchevsky, BSCLife SciencesDept.</p> <p><i>Single amino acid mutation controls hole transfer dynamics in DNAmethyltransferase HhaI complexes</i> Marina Corbella, Universitat de Barcelona</p> <p><i>Catalytic surface radical in dye-decolorizing peroxidase</i> Marina Cañellas, BSCLife SciencesDept.</p> <p><i>Theoretical study on the activation mechanism of AMP-kinase by means of Molecular Dynamics Simulations</i> Carolina Estarellas (PostDoc), Universitat de Barcelona</p> <p><i>GWImp-COMPSs: An Integrated Framework for Large-scale Genome-wide Imputation and Association Testing</i> Marta Guindo, BSCLife SciencesDept.</p> <p><i>PMut2015: a web-based tool for predicting pathological mutations on proteins</i> Víctor López, BSCLife SciencesDept.</p> <p><i>Influence of Temperature on the Topological Features of Inner Cavities in Cytooglobin</i> Constantí Seira, Universitat de Barcelona</p> <p><i>Cation- <math>\pi</math>-cation interactions in proteins</i> Silvana De Souza Pinheiro, Universitat de Barcelona</p>		
<b>Fifth Talk Session: Computer Sciences and Applications</b>			
11.20h	A novel approach to reconstruct the plinian and co-ignimbrite phases of large eruptions - Campanian Ignimbrite	Alejandro Martí, BSC Computer Applications in Science & Engineering Dept.	
11.40h	Plasma Physics Code Contribution to the Mont-Blanc Project	Xavier Sáez, BSC Computer Applications in Science & Engineering Dept.	
12.00h	Folding: reporting instantaneous performance metrics and source-code references	Harald Servat, BSC Computer Sciences Dept.	
12.20h	Probabilistically Time-Analyzable Complex Processors in Hard Real-Time systems	Mladen Slijepcevic, BSC Computer Sciences Dept.	
12.40h	Improving the prefetching performance through code region profiling	Martí Torrents, UPC BarcelonaTech	
13.00h	Lunch break		



2nd International BSC Doctoral Symposium 2015 – 7th May			
Start time	Activity	Speaker/s	
<b>Sixth Talk Session: Earth Sciences and Physics</b>			
14.00h	Dynamics of double-polarity subduction: application to the Western Mediterranean	Mireia Peral, Institute of Earth Science Jaume Almera (CSIC)	
14.20h	Characterization of coal power plants plume dynamics under typical synoptic conditions over the Iberian Peninsula	Víctor Valverde, BSC Earth Sciences Dept.	
14.40h	Quantum dynamics study of the hydrogen molecule confined in single-walled carbon nanotubes	Manel Mondelo-Martell, Universitat de Barcelona	
15.00h	<b>Fifth Poster Session &amp; coffee break</b> <i>Hecuba: NoSql made easy</i> Guillem Alomar, BSC Computer Sciences Dept. <i>Scaling Irregular Array-type Reductions in OmpSs</i> Jan Ciesko, BSC Computer Sciences Dept. <i>DLP Acceleration on General Purpose Cores</i> Milovan Duric, BSC Computer Sciences Dept. <i>Enhancing Timing Analysis for COTS multicores for the Space industry: a software approach</i> Gabriel Fernandez, UPC BarcelonaTech <i>Methodology to select a I/O configuration (hardware resources and stack software) in cloud platform</i> Pilar Gómez, Universitat Autònoma de Barcelona <i>A Multiphysics implementation for conjugate heat transfer and compressible-low mach coupling</i> Miguel Zavala, BSC Computer Applications in Science & Engineering Dept. <i>Understanding Scientific Application's Performance</i> Oriol Tintó, Universitat Autònoma de Barcelona		
<b>Seventh Talk Session: Life Sciences</b>			
15.30h	Sugar Conformations that Enhance Cleavage of Glycosidic Bonds in Carbohydrate-Active Enzymes	Lluís Raich, Universitat de Barcelona	
15.50h	pyDock performance in 5th CAPRI edition: from docking and scoring to binding affinity predictions and other challenges	Chiara Pallara, BSC Life Sciences Dept.	
16.10h	Characterization of complex chromosomal rearrangements in cancer genomes	Marta Munar, BSC Life Sciences Dept.	
16.30h	Conclusions		
17.00h	End of the Doctoral Symposium		







## KEYNOTE SPEAKER

### Nuria Oliver

Telefónica R&D



Nuria Oliver is currently the Scientific Director and founder of the User, Data and Media Intelligence research team in Telefonica Research (Barcelona, Spain). She is responsible for the HCI, Mobile Computing, Big and Personal Data Mining, User Modeling and Multimedia Research Areas. Nuria received the BSc (honors) and MSc degrees in Electrical Engineering and Computer Science from the ETSIT at the Universidad Politecnica of Madrid (UPM), Spain, in 1992 and 1994 respectively. She received her PhD degree from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in June 2000. From July 2000 until November 2007, she was a researcher at Microsoft Research in Redmond, WA. At the end of 2007, she

returned to Spain to co-create the Research organization at Telefonica R&D by creating and leading her research team.

Her research interests include artificial intelligence, health monitoring, mobile computing, personal and big data analysis, statistical machine learning and data mining, smart environments, context awareness, multimedia data analysis, recommender systems, social network analysis, computational social sciences, and human computer interaction. She is currently working on the previous disciplines to build human-centric intelligent systems and improve the world with technology.

Nuria has written over 100 scientific papers in international conferences, journals and book chapters. Her work has been widely recognized by the scientific community with over 8100 citations. According to Google Scholar Nuria is the most cited female computer scientist in Spain. Nuria has over 40 patent applications and granted patents. She is also in the program committee and a reviewer of the top conferences in her research areas (IJCAI, IUI, UMAP, ACM Multimedia, ICMI-MLMI, SocialComp, Interaction, PervasiveHealth, MIR, LoCA, MMM, CVPR, Ubicomp, MobileHCI, ICCV, AAAI, etc...). She was program co-chair of IUI 2009 and of MIR 2010, general conference co-chair of UMAP 2011, industry-day co-chair of IJCAI 2011, track co-chair of ACM WWW 2013 and track founder and co-chair in ACM MM 2014, among others.

She regularly gives invited talks about her research both for scientific and the general public. She has been a keynote speaker at IEEE EUSIPCO 2011, European Wireless Conference 2014, ACM and IEEE MODELS 2014 and NTTTS 2015, among others. She has given two TEDx talks, a WIRED talk and is co-organizer of the first ever TEDxBCN event devoted to Education. She has spoken to the Spanish Senate.

She believes in the power of technology to empower and increase the quality of life of people. She has received a number of awards, including an ACM ICMI 2014 10-year Technical Impact Award (2014), a Rising Talent Award by the Women's Forum for the Economy and Society (2009), MIT's 'TR100 Young Innovators Award' (2004) and the First Spanish Award of EECS graduates (1994). She is a Senior Member of the ACM and IEEE.

Besides her scientific publications, she is very interested in making science available to the general public. She has been a technology writer for Tecno2000 magazine and 'El Pais' newspapers, among others. Her work has been featured on multiple newspapers, magazines, radio and TV stations both in Spain and the US. She has been featured in Glamour magazine as 'one of the top female directors in Spain' (2015), diari ARA, in EL PAIS Sunday magazine as one of a few 'female directors in technology' (2012), named <% Rising Talent by the Women's Forum for Economy & Society (October 2009),%> one of the '13 most influential young women in Spain' (MujerHoy Magazine, 2012), one of '100 leaders of the future' by Capital Magazine (May 2009) and one of the 'Generation XXI: 40 Spanish youngsters that will make news in the Third Millenium ' by EL PAIS (2000).

## LECTURER & TRAINING

### Josep Lluís Pelegrí

Institut de Ciències del Mar, CSIC



Josep Lluís has a PhD in Physical Oceanography and extensive experience working with the industry and as an undergraduate and graduate lecturer. He was the Dean for the School of Marine Sciences at the Universidad de Las Palmas de Gran Canaria and the Earth Sciences coordinator for the Spanish National Evaluation Agency. Since 2003 he works at the Institute of Marine Sciences (CSIC), currently as a Research Professor and member of CSIC's Natural Resources committee. He has advised 14 doctoral dissertations and published about 90 SCI papers.

#### Training “An Introduction to Scientific Writing”

To become a successful scientist, it is not enough to have a good theoretical background and to learn innovative technical skills. It is equally necessary to develop good writing skills so that you can efficiently communicate your ideas and results. Your articles and proposals need to be attractive and informative so that they get read and cited. This course will provide the essentials on scientific writing style, including the basic elements on how to plan, draft and write the final version of your paper.

### An Introduction to Scientific Writing

Prof. Josep L. Pelegrí, Institut de Ciències del Mar

5 and 6 May 2015, 15:00-19:00

#### General course description:

To become a successful scientist, it is not enough to have a good theoretical background and to learn innovative technical skills. It is equally necessary to develop good writing skills so that you can efficiently communicate your ideas and results. Your articles and proposals need to be attractive and informative so that they get read and cited. This course will provide the essentials on scientific writing style, including the basic elements on how to plan, draft and write the final version of your paper.

**Number of hours:** 8 hours lectures + 8 hours of home work

**Requirements:** Good English knowledge. Students are also required to bring their own laptop. Students are asked to write a short scientific essay (1-3 pages) and bring it to the first day of class. At the end of the first day, they will be asked to critically revise this essay, on their own or by groups, following the guidelines and ideas instructed during that day. Students will bring the revised essay on the second day, where the essays will be jointly revised.

#### Detailed program:

The first day will be dedicated to review the fundamentals of scientific writing. In the second day we will practice these concepts, students will team to analyse samples of their own work.

Part 1: Why becoming a professional writer. The basics of words and sentences. Tense choice: old and new information. Active versus passive voices. Topic, action/challenge and stress. From sentences to paragraphs and sections: the flowing message. Nested arcs. Internal coherence and esthetic consistency. From the first draft to the final manuscript.



Part 2: Targeting your audience. Laying out the elements of the story. Outline your paper: sections, figures and tables. Opening and challenge. Action and resolution.

### Recommended reading:

Gopem G, Swan J. 1990. The science of scientific writing. American Scientist.

<https://www.americanscientist.org/issues/pub/the-science-of-scientific-writing>

Marek, P. 2010. The basics of scientific writing in APA style.

[http://www.worthpublishers.com/Catalog/uploadedFiles/Content/Worth/Custom\\_Solutions/Psychology\\_ForeWords/Marek\\_Ch04\\_APASTyle\\_Color.pdf](http://www.worthpublishers.com/Catalog/uploadedFiles/Content/Worth/Custom_Solutions/Psychology_ForeWords/Marek_Ch04_APASTyle_Color.pdf)

Olson L., 2014. Guide to academic and scientific publication. How to get your writing published in scholarly journals. <http://www.proof-reading-service.com/guide/>

Peat J. 2002. Scientific writing. Easy when you know how. BMJ Books.

<http://image.sciencenet.cn/olddata/kexue.com.cn/upload/blog/file/2010/11/2010111022210817101.pdf>

Schimel J. 2012. Writing science – How to write papers that get cited and proposals that get funded. Oxford University Press, Oxford, 221 pp.

Strunk W, White EB. 2000. The elements of style. Fourth Edition, Longman, New York, 105 pp.

Williams JM. 1990. Style: Toward clarity and grace. The University of Chicago Press, Chicago, 208 pp.

### Online resources

- Grammar:

British Council: <http://learnenglish.britishcouncil.org/en/english-grammar>

Edu find: <http://www.edufind.com/english-grammar/english-grammar-guide/>

- Dictionary and thesaurus:

Reference web mail: [www.dictionary.com](http://www.dictionary.com), [www.thesaurus.com](http://www.thesaurus.com)

Oxford English dictionary: [www.oed.com](http://www.oed.com)

- Other:

The Writing Center, UNC at Chapel Hill: <http://writingcenter.unc.edu/handouts/>

Butte College TIP Sheets: <http://www.butte.edu/cas/tipsheets/>

Essay Info Writing Guide. [http://essayinfo.com/essays/narrative\\_essay.php](http://essayinfo.com/essays/narrative_essay.php)

## TALK PRESENTERS

### Claudia Rosas

Barcelona Supercomputing Center



Claudia Rosas got the B.S. degree in computer engineering in 2008 from Universidad Valle del Momboy (Venezuela). She got the M.Sc. in High-Performance Computing in 2009 and the Ph.D. in High-Performance Computing in 2012, both from Universitat Autònoma de Barcelona (Spain). Her research interests include high performance parallel applications, automatic performance analysis and dynamic tuning. In 2012, she joined as a Post-Doctoral Researcher in the Barcelona Supercomputing Center. Her daily tasks are related to the Intel-BSC Exascale Lab, being her main focus the prediction and extrapolation of the expected performance of parallel applications at Exascale Computing.

#### Talk “Methodology to predict scalability of parallel applications”

In the road to exascale computing, the inference of expected performance of parallel applications results in a complex task. Performance analysts need to identify the behaviour of the applications and to extrapolate it to non-existent machines. In this talk, I present a methodology based on collecting the essential knowledge about fundamental factors of parallel codes, and to analyse in detail the behaviour of the application at low core counts on current platforms. The result is a guide to generate the model that best predicts performance at very large scale. Obtained results from executions at low core counts showed expected parallel efficiencies with a low relative error.

### Abel Gargallo-Peiró

Barcelona Supercomputing Center



Since 2014 he is working as Postdoc researcher at BSC. After graduating in Mathematics in UPC (five-years degree) in 2010, he did a Master in Applied Mathematics and in 2014 he finished his PhD in Applied Mathematics, which was entitled “Generation of curved meshes for high-order unstructured methods” and was directed by J. Sarrate (UPC), and X. Roca (MIT). During the PhD thesis I developed two research stays at the Aeronautics and Astronautics Department from MIT (Cambridge, USA). My main research interests are mesh generation, curved mesh generation, high-order methods, mesh adaptivity, and mesh optimization.

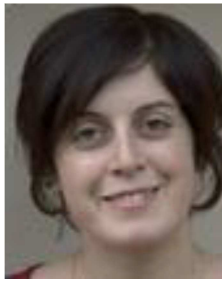
#### Talk “Mesh generation to simulate incompressible Navier-Stokes equations with wind turbines”

My talk proposes a new methodology to generate hybrid meshes to simulate incompressible Navier-Stokes equations in the framework of wind energy to estimate the production of a given wind turbine field. The target meshes must fulfil several requirements: boundary layer close to the topography, wind mesh conformably adapted to the disc with actuator disc, wake mesh size transition, wake mesh adaptation, radial mesh transition to conform the external mesh size.



## Enza Di Tomaso

Barcelona Supercomputing Center



Enza Di Tomaso has a degree in Physics from the University of Bologna in Italy and a PhD in Engineering Mathematics from the University of Bristol in UK. She has worked as lecturer at the University of Bristol before moving to the field of atmospheric science, initially working for the Italian Research Council on the retrieval of precipitation from satellite observations, and subsequently working at the European Centre for Medium-Range Weather Forecasts in Reading, UK, on exploiting satellite observations for Numerical Weather Prediction. Recently, at BSC, she has expanded her interest to atmospheric chemistry and in particular to aerosol data assimilation.

### Talk “Measurements and computer simulations: who do we trust for atmospheric composition?”

Computer simulations are able to provide a description of the atmosphere at unprecedented fine spatial and temporal scales, thanks to the most recent advances made in model developments and to the increasingly available computational resources. However, due to our limited understanding of physical processes in the atmosphere, their accuracy is not at the level of atmospheric measurements, which, on the other hand, suffer of a lower temporal resolution and sparser coverage and, by their nature, lack predictive skills. Given that limitations and uncertainties exist both in models and measurements, whom do we trust for the correct estimation of atmospheric variables? The Bayesian paradigm provides a probabilistic approach for optimally combining in a data assimilation framework the two pieces of information (model simulations and observations) weighted by their respective uncertainties. We report here on the developments on this field at the Earth Sciences Department of the Barcelona Supercomputing Center (BSC) for measurements and model simulations related to mineral dust. We resort to a data assimilation technique particularly suited to high-performance computing applications, and show that it improves significantly the characterisation of the atmospheric dust load.

## Ruth Arís

Barcelona Supercomputing Center



Ruth Arís was born in Barcelona. She holds an MSc in Physics from Universitat de Barcelona. She has worked as associate professor at Universitat de Barcelona and Universitat Politècnica de Catalunya.

Since 2007, Ruth Arís is working at the Computer Applications in Science and Engineering (CASE) Department of the Barcelona Supercomputing Center (BSC). As a PhD student, she developed a Large Scale Computational Model in the Ventricular Myocardium at the High Performance Computational Mechanics Group. The Cardiac Computational Mechanics tool involves Physical modelling, Mathematical algorithms, code development and optimization, all adapted to run efficiently in large-scale parallel computers. Alya System is the in-house HPCM tool.

His main lines are on Computational Mechanics and HPC, with a strong focus in Biomechanics.

Recently, Ruth Arís has been a Postdoctoral Fellow at Mount Sinai Hospital in New York. In the current project, she has extended her work in simulation to medical imaging, focusing on the simulation of the blood flow within the carotid arteries.

### Talk “From imaging to simulation: a framework applied to simulate the blood flow in the carotids”

During my stay at Mount Sinai in New York, I have developed a methodology to extract information from medical imaging and use it for hemodynamical simulation in arteries. Based on in-vivo magnetic resonance images (MRI), the velocity of the blood flow has been measured at different positions and times. Also, the anatomy of the vessel has been converted into a volume mesh suitable for numerical modelling. This data has been used to



Solve computationally the dynamics of the fluid inside the artery in healthy and pathologic cases. As an application, we have developed a computational model within the carotids. The next step in the pipeline will be to extend the simulation to fluid-structure interaction (FSI) to find the parameters in an atherosclerotic plaque that could lead to rupture.

### **Agustí Emperador**

**Barcelona Supercomputing Center**



He got a degree in Physics by the University of Barcelona in 1996 and became Doctor in Physics by the University of Barcelona in 2000. He also was teaching assistant, academic year 2000-2001 (Theory and problems lectures of Mechanics). Departament d'Estructura i Constituents de la Materia, Faculty of Physics, University of Barcelona. Agustí had a Postdoc position in the Department of Physics of the University of Trento (Italy), (October 2001 - June 2003) and a Postdoc position in the Department of Physics of the University of Oxford (UK), July (2003 - March 2004). After he got a Postdoc position in the Department of Physics and Nuclear Engineering of the Universitat Politècnica de Catalunya (April 2004 - March 2006) and became Ramon y Cajal researcher in the Institute for Research in Biomedicine in May 2006. In

December 2011 started working at BSC as research associate in the Institute for Research in Biomedicine.

#### **Talk “Optimizing a coarse-grained model for the recognition of protein-protein binding”**

Our team is optimizing a force-field to be used with our coarse-grained protein model for the recognition of protein-protein binding. We have found that, apart from ranking correctly the ligand-receptor conformations generated in a protein-protein docking algorithm, our model is able to distinguish binding (experimental structure) from nonbinding (false positive) conformations for many complexes. This suggests us that the model could have a good performance in complete cross-docking, a method aimed to recognize the possible binding between any two proteins that are unknown to interact.

### **Oscar A. Esquivel**

**CONACYT-Barcelona Supercomputing Center**



O. A. Esquivel-Flores received his Bachelor's degree in Applied Mathematics and Computing from Universidad Nacional Autónoma de México. M.S. degree in Computer Sciences from Universidad Autónoma Metropolitana, México and PhD degree in Computer Engineering from UNAM in 2013.

Actually he works on parallel and high performance computing as part of a postdoctoral position at the Barcelona Supercomputing Center as an international agreement with National Council of Science and Technology of México

#### **Talk “Study of Preconditioners based on Markov Chain Monte Carlo Methods”**

Nowadays, analysis and design of novel scalable methods and algorithms for fundamental linear algebra problems such as solving Systems of Linear Algebraic Equations with focus on large scale systems is a subject of study. This research focus on the study of novel mathematical methods and scalable algorithms for computationally intensive problems, such as Monte Carlo, and Hybrid Methods and Algorithms.

### **Rafael Delgado**

**Universidad Complutense de Madrid**

Rafael Delgado is PhD student in the Effective Theories in Modern Physics Group, at the Theoretical Physics Dept., Universidad Complutense de Madrid. His supervisors are Felipe J. Llanes-Estrada and Antonio Dobado. He is also working with J.A.R. Cembranos, S. Moretti, J.J. Sanz-Cillero, M.J. Herrero and Domènec Espriu.



His research field is particle physics phenomenology of Beyond Standard Model (BSM) theories. Specially, in the range of energies which can be reached at the LHC. He has considered two cases: effective theories implying a strongly interacting Electroweak Symmetry Breaking Sector (a non-standard Higgs sector), and Kaluza Klein models of gravity, involving extra dimensions. For both projects, He has needed computer time from the RES (Spanish Supercomputing Network), which was granted in the Tirant supercomputer.



**Talk “Supercomputing simulations for Beyond Standard Model theories at the TeV scale: non-standard gravity and electroweak interactions”**

The LHC experiment has given us two new pieces of information about the physics at the TeV scale. The first one is the existence of a new bosonic resonances at about 125 GeV which, with the available data, fits the properties of the Standard Model Higgs boson. The second one, a mass gap for the presence of new physics until 600-700 GeV. In this work, we are studying two models: a brane world model, and a strongly interacting 'Higgs' sector. Our aim is to guide experimental searches at the LHC and beyond by performing independent theory studies. The CPU time needed exceeds reasonable demands on our departmental 20-PC cluster and suggests employing a supercomputer facility at a modest level (1/4 M hours/year).

**Ilya Blokh**  
Perm State University



Born in 1990. After finishing mathematical school went to Perm State University to the Faculty of Mechanics and Mathematics. Finished in 2011 the bachelor program at department of mathematical maintenance of computer systems and finished in 2013 the master program in the same place. The theme of master project was “Genetic algorithm of route modelling by parameters for mobile tourist’s application”. Passing post-graduate study in PSU since 2014.

**Talk “Discovering most significant news using Network Science approach”**

The role of Internet mass media increased greatly in the recent years. Internet media attracts more and more audience coverage and thereby its role in forming of public opinion strengthens. Public organizations, political parties, companies as well as public figures, politicians and entrepreneurs are interested in any mentioning of them or of some specific facts concerning them in media including Internet media. They often trace a quantity of web-sites citing them, and different sources where some significant news could appear. Significant news finding could be also useful for common Internet uses since the amount of daily media publications is large and some news which are significant for someone could be missed. Thus significant (in some context) news identification is essential and actual problem.

**Sicong Zhuang**  
Barcelona Supercomputing Center



Sicong Zhuang graduated from the Master MIRI (Master in Innovation and Research in Informatics) specialized for HPC from UPC (Universitat Politècnica de Catalunya), Spain in 2014. During his master, his research focused on the process variation and soft error on register files and memory systems.

He started his collaboration with the BSC (Barcelona Supercomputing Center) during the last semester working on developing parallel sparse Cholesky factorization. He joined BSC Computer Science department as a resident student right after graduation and is currently a PhD candidate in UPC as well. His current research focuses on numerical linear algebra algorithms and parallel computing. He is also one of the developers of the Linompss.

**Talk “Linompss – A Linear Algebra Library on OMPs”**

LINOMPSS, a parallel linear algebra library built on top of OMPs. The research group I am part of not only





considers LINOMPSs as an environment for the development of linear algebra applications and benchmarks for the Computer Science Department. In our own department for example, we use LINOMPSs to experiment with mixed-precision and incomplete/inaccurate computations. With the expertise of the CASE and Earth Science Department at BSC, we envision LINOMPSs as a tool for cross-disciplinary collaboration and the development of industrial-strength engineering applications.

### **Kallia Chronaki**

**Barcelona Supercomputing Center**



Kallia Chronaki has a degree in Computer Science and a master in Computer Architecture, both from University of Crete, Greece. During her studies she also worked as a research assistant in the Computer Architecture and VLSI laboratory of the Institute of Computer Science – F.O.R.T.H., Greece. Currently she is working as a research engineer in Barcelona Supercomputing Center. Her research interests include high performance computing, parallel programming models, task scheduling and dependence analysis.

#### **Talk “Criticality-Aware Task Scheduling for Heterogeneous Systems”**

OmpSs is a powerful task-based programming model with dependency tracking and dynamic scheduling. In this talk we will describe the OmpSs approach on scheduling dependent tasks onto the asymmetric cores of a heterogeneous system. The proposed dynamic scheduling policy uses information discoverable at runtime and reduces total execution time. It first prioritizes the newly-created tasks at runtime according to the shape of the task dependency graph; it then detects the longest path of the dynamic task dependency graph, and finally it assigns critical tasks to fast cores and non-critical tasks to slow cores.

### **Renan Fischer**

**Barcelona Supercomputing Center**



Renan Fischer is PhD student in Computer Architecture at the UPC. He received both his MSc and BSc in Computer Science from the Federal University of Paraná, Brazil. His research interests include security and energy-efficiency on Local Area Networks and Wireless Networks.

#### **Talk “Interconnect Energy Savings on MapReduce Clusters”**

An important challenge of modern data centers is to reduce energy consumption, of which a substantial proportion is due to the network. Energy Efficient Ethernet (EEE) is a recent standard that aims to reduce network power consumption, but current practice is to disable it in production use, since it has a poorly understood impact on real world application performance. An important application framework commonly used in modern data centers is Apache Hadoop, which implements the MapReduce programming model. This work analyses the impact of EEE on MapReduce workloads, in terms of performance overheads and energy savings. We find that large energy savings are possible, but only if the links use packet coalescing. Packet coalescing must, however, be carefully configured in order to avoid an excessive performance loss in performance.

### **Vasileios Karakostas**

**Barcelona Supercomputing Center**



Vasilis Karakostas is a PhD student at Barcelona Supercomputing Center and Universitat Politècnica de Catalunya since March 2012. His thesis focuses broadly on computer architecture and more specifically in the hardware/software mechanisms for Virtual Memory. Vasilis received a MS in Computer Architecture, Networks and Systems from UPC in February 2012, and a Diploma in Electrical and Computer Engineering from the National Technical University of Athens in November 2009. dependence analysis.



**Talk “Efficient Address Translation for Memory Intensive Workloads”** “Virtual memory enhances programmer productivity by abstracting the available physical memory and improves system consolidation by isolating processes from each other [1]. However, the benefits of virtual memory are not free in terms of performance and energy.

## Zhengchun Liu

Universitat Autònoma de Barcelona



Zhengchun Liu is a PhD student of Computer Science at Universitat Autònoma de Barcelona (UAB). He received his master degree in Navigation Guidance and Control from the School of Astronautics, Northwest Polytechnic University (NPU) and a B.S. degree in Manufacturing Engineering of Flight Vehicle from School of Mechanical Engineering, NPU. His thesis in progressing is about simulating based decision support system for healthcare system management. More specifically, it is about simulating the Emergency Department by using agent based technique and makes it work as a part of decision support system. Up to now, he has finished the modelling work and written three full-reviewed conference papers (SIMUL 2014, published; ICCS2015, accepted and WSC2015 submitted). His current work is to propose an application framework for

knowledge discovery via simulating the individual behaviour. His research interests include modelling and simulation, agent-based simulation, decision support system and complex system.

### Talk “High Performance Computing Based Simulation for Healthcare Decision Support”

Due to the complexity and crucial role of an Emergency Dept. (ED) in the healthcare system. The ability to more accurately represent, simulate and predict performance of ED will be invaluable for decision makers to solve management problems.

One way to realize this requirement is by modelling and simulation. The objective of this research is to grasp the non-linear association between macro-level features and micro-level behaviour with the goal of better understanding the bottleneck of ED performance and provide ability to quantify such performance on defined condition. Agent-based modelling approach was used to model the healthcare staff, patient and physical resources in ED.

Instead of describe all the potential causes of this complex issue.

Rather, in this thesis, a layer-based application framework will be presented to discover knowledge of a complex system through simulating micro-level behaviours of its components to facilitate a systematic understanding of the aggregate behaviour.

## Nikola Markovic

Barcelona Supercomputing Center



Nikola Markovic is a PhD student at Universitat Politecnica de Catalunya and a researcher in the Computer Architecture for Parallel Paradigms (CAP) group at the Barcelona Supercomputing Center. His research focuses on hardware and software design for high-performance multithreaded computing on symmetric and asymmetric chip multiprocessors. Markovic has an MSc in computer architecture, networks, and systems from the Universitat Politecnica de Catalunya.

### Talk “Hardware Scheduling Algorithms for Asymmetric Single-ISA CMPs”

As thread level parallelism in applications has continued to expand, so has research in chip multi-core processors. Since more and more applications become multi-threaded we expect to find a growing number of threads executing on a machine.

Consequently, the operating system will require increasingly larger amounts of CPU time to schedule these threads efficiently. Instead of perpetuating the trend of performing more complex thread scheduling in the operating system, we propose a hardware thread scheduling mechanisms. This lightweight mechanisms help to identify multi-threaded application bottlenecks such as thread synchronization sections and complements the Fairness-aware Scheduler method.

## Francisco Alarcón

Universitat de Barcelona



Francisco Alarcón studied Physics at Universidad Michoacana de San Nicolás de Hidalgo (UMSNH) at México, he also did a Master degree on Physics at the University of San Luís Potosí (UASLP). During the master degree and afterwards Francisco worked in the Centro de Investigación en Polímeros (Grupo COMEX) which is the R&D department of the mexican industry of paints and coatings COMEX, during this time he developed and applied a computational model to simulate molecular systems at the mesoscale, in order to make relevant predictions about the structural conformation of electrically charged polymers in suspension. Nowadays Francisco Alarcón is finishing his last year of PhD at University of Barcelona (UB) under the advice of Prof. Ignacio Pagonabarraga

where they have developed and applied a theoretic-numeric scheme to study the collective behaviour of microorganisms. They have done systematic studies to understand how hydrodynamic signature affects to the emergency of collective motion.

### Talk “Macroscopic structures generated by microorganisms swimming in a fluid”

It is known that active particles induce emerging patterns as a result of their dynamic interactions, giving rise to amazing collective motions, such as swarming or clustering. Here we present a systematic numerical study of self-propelling particles; our main goal is to characterize the collective behaviour of suspensions of active particles as a result of the competition among their propulsion activity and the intensity of an attractive pair potential. Active particles are modelled using the squirmer model. Due to its hydrodynamic nature, we are able to classify the squirmer swimmer activity in terms of the stress it generates (referred to as pullers or pushers). We show that these active stresses play a central role in the emergence of collective motion. We have found that hydrodynamics drive the coherent swimming between swimmers while the swimmer direct interactions, modelled by a Lennard-Jones potential, contributes to the swimmers' cohesion. This competition gives rise to two different regimes where giant density fluctuations (GDF) emerge. These two regimes are differentiated by the suspension alignment; one regime has GDF in aligned suspensions whereas the other regime has GDF of suspensions with an isotropic orientated state. All the simulated squirmer suspensions shown in this study were characterized by a thorough analysis of global properties of the squirmer suspensions as well as a complementary cluster analysis.

## Didier Barradas

Barcelona Supercomputing Center



Didier Barradas got a BSc in Biochemistry by the Faculty Of Sciences, Autonomous University of Morelos (UAEMor); Mexico 2002-2007. He was Lab Technician. Center for research on infectious diseases, National Institute of Health, Mexico; 2007-2009 and got a Master Degree in Biochemical Sciences. Institute of Biotechnology National Autonomous University of Mexico (UNAM), Mexico; 2009-2011; since 2012 he is PhD Student.

### Talk “Assessment of scoring functions performance to rerank docking decoys from FFT (Fast Fourier Transform) programs”

To discriminate decoys from near native solutions, docking methods rely in Scoring functions (SF). Usually a docking method uses its own SF to rank accordingly and discriminate from the decoys those that are near native solutions. In this work we present a short analysis on the performance several SF in three different docking methods, showing that other SF are able to rank better than the in-built SF, remarking the possibility to combine them into a new SF or strategy and improve the overall success rate.

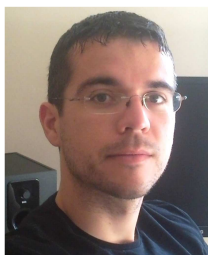
## Víctor Gil

Barcelona Supercomputing Center

Víctor Gil had his computer engineering diploma in 2007 from the Barcelona Faculty of Informatics (FIB/UPC). Due to his interest in the Life Science's field, he joined the Electronic and Atomic Protein Modeling group in the BSC while he was finishing his engineering final project. He left the group in order to enrol a Biomedical Engineering master one year later. He wrote his master thesis during an 8 months stage at the CRG research



center. After that he joined again BSC's Electronic and Atomic Protein Modeling group as a predoctoral researcher where he is currently involved in two main projects: First, the application of general clustering solutions to the analysis of conformational search software results; second, the improvement of the in-house software PELE. This includes enhancing its code by a complete rewrite to C++, to boost its performance by applying HPC techniques and to algorithmically improve it by implementing innovative ways to perform the conformational search.



### **Talk “Implementation of an Internal Coordinates Anisotropic Network Model in PELE”**

The use of computational methods to elucidate the ligand-protein binding mechanisms is of utmost importance for the pharmaceutical industry. PELE (Protein Energy Landscape Exploration) software, has proved to have good predictive power. We want to further improve it by changing its conformational sampling step.

## **Eduardo Mayol**

**Universitat Autònoma de Barcelona**



He got a degree in Chemistry at Universitat de Barcelona (2004-2012) and a degree in Biochemistry at Universitat de Barcelona (2009-2012). He also got a MSc in Bioinformatics at Universitat Autònoma de Barcelona (2012-2013). In 2013, after the MSc practicum at Leonardo Pardo's group, he could start the PhD at the same group, directed by Arnau Cordoní and Leonardo Pardo in the program of Biochemistry, Molecular Biology and Biomedicine with the title: Development of Bioinformatics tools for the study membrane proteins. During the degrees he has carried out many internships including one at the BSC (July-October 2011) directed by Ramón Goñi. He has also joined a Collaborative

Fellowship at the “Centres Científics i Tecnològics” of the Universitat the Barcelona, at the Confocal Microscopy Department between September 2011 and September 2012.

### **Talk “Inter-residue interactions in membrane proteins”**

Knowing the precise 3D-structure of a protein is crucial to understand its functional mechanism at the molecular level and to develop new pharmacological agents to targeting it. Nowadays only a few hundred integral membrane protein structures have been solved at high resolution due to the associated technical difficulties. In the present study we aim to characterize the main interactions in alpha and beta membrane proteins that are responsible of the maintenance of the overall structure. With this purpose, two non-redundant databases of alpha and beta transmembrane segments were constructed and analysed. The interactions that stabilize the structure of alpha and beta membrane proteins were quantified. The results reveal important differences in inter-residues interactions between alpha and beta membrane proteins. This novel structural information may be useful in predicting 3D models of proteins lacking structural information or in refining initial models of alpha and beta membrane proteins.

## **Alejandro Martí**

**Barcelona Supercomputing Center**



Barcelona Supercomputing Center (BSC-CNS), and a PhD candidate at the Polytechnic University of Catalonia (UPC). His research aims at quantifying the feedback effects of volcanic ash clouds and aerosols emitted during large-magnitude explosive eruptions on regional meteorology using a state-of-the-art chemical weather prediction model (NMMB/BSC-CTM). His research is sponsored by the NEMOH-ITN FP7 Network, which aims to develop and form the next generation of European volcanologists.

Before joining the BSC, Mr. Martí worked as climate systems researcher at the UK MetOffice/University of Exeter where he researched the impact of changes in atmospheric composition and land management on terrestrial ecosystems. Previously, he had a successful career in United States as a geoengineer, where he spent almost 10 years developing computer applications for



land-use management for a US-Governmental planning agency. Finally, he served as an Assistant Professor at different American Universities teaching courses in GIS and Remote Sensing.

He holds a Bachelor degree in Environmental Sciences, a MSc in Environmental Engineering and a MSc Geographic Information Systems and Remote Sensing from Rutgers and Princeton University.

### **Talk “A novel approach to reconstruct the plinian and co-ignimbrite phases of large eruptions - Campanian Ignimbrite”**

Reconstructing the volume and tephra dispersal from volcanic super-eruptions is necessary to assess the widespread impact of these massive events on climate, ecosystems and humans. Recent studies have demonstrated that volcanic ash transport and dispersion models are unrivaled in accurately constraining the volume of material ejected and provide further insight about the eruption dynamics during these gigantic events. However, the conventional simplified characterization of caldera-forming supereruptions as a single-phase event can lead to inaccurate estimations of the eruption dynamics and its impacts. Here, we apply a novel computational inversion method to reconstruct, for the first time, the two phases of the largest eruption of the last 200 ky in Europe, the Campanian Ignimbrite (CI) super-eruption.

Additionally, we discuss the eruption’s contribution to the Middle to Upper Palaeolithic transition by evaluating its environmental and climate implications.

#### **Xavier Sáez**

##### **Barcelona Supercomputing Center**



Xavier Sáez received the BSc degree in computer engineering from the Universitat Politècnica de Catalunya (UPC), Spain. He works at Barcelona Supercomputing Center since 2006 in the Computer Applications in Science & Engineering Dept. (CASE), where he is finishing the PhD degree in the plasma physics group. Xavier from March 2015 onwards is member of the EUROfusion High Level Support Team, which provides support to scientists in the EUROfusion consortium for the development and optimization of their codes. He has participated in several European projects, such as PRACE, EUFORIA and Mont-blanc, as well as collaborated with IBM and AIRBUS.

### **Talk “Plasma Physics Code Contribution to the Mont-Blanc Project”**

The talk will explain the development of strategies for adapting a particle-in-cell code to heterogeneous computer architectures and, in particular, to an ARM-based prototype of the Mont-Blanc project using OmpSs programming model and the OpenMP and OpenCL languages. Dependency graph; it then detects the longest path of the dynamic task dependency graph, and finally it assigns critical tasks to fast cores and non-critical tasks to slow cores.

#### **Harald Servat**

##### **Barcelona Supercomputing Center**



Harald earned his B.Sc. in Computer Engineering in 2003 and then he started working on the Performance analysis tools from the CEPBA-IBM Research Institute. He’s the responsible for the instrumentation part of these tools. In 2007, he started his PhD studies at partial time under the advices of Prof. Jesús Labarta, and he expects to defense his thesis during 2015.

### **Talk “Folding: reporting instantaneous performance metrics and source-code references”**

Despite supercomputers deliver huge computational power, applications only reach a fraction of it. There are several factors limiting the application performance, and one of the most important is the single processor efficiency because it ultimately dictates the overall achieved performance. We present the folding mechanism, a process that combines measurements captured through minimal instrumentation and coarse-grain sampling ensuring low time dilation (less than 5%). The mechanism reports instantaneous performance and source-code references for optimized binaries accurately by taking advantage of the repetitiveness of many applications,



especially in HPC. The mechanism enables the exploration of the application performance and guides the analyst to source-code modifications.

### Mladen Slijepcevic

Barcelona Supercomputing Center



Mladen Slijepcevic is a PhD student in the CAOS group at Barcelona Supercomputing Center (BSC). Mladen joined BSC in December 2012. He holds a MSc and BSc in Electrical Engineering and Computer Science from University of Belgrade, Serbia. Mladen participates in FP7 PROXIMA project. His main focus is hardware design for real-time embedded systems.

#### Talk “Probabilistically Time-Analysable Complex Processors in Hard Real-Time systems”

Critical Real-Time Embedded Systems (CRTES) feature performance-demanding functionality. High-performance hardware and complex software can provide such functionality, but the use of aggressive technology challenges time-predictability. Our work focuses on the investigation and development of hardware mechanisms to control inter-task interferences in shared time-randomized caches and manycore network-on-chip designs meeting the requirements of Probabilistic Timing Analysis (PTA).

### Martí Torrents

Universitat Politècnica de Catalunya BarcelonaTech



Martí Torrents studied all his university degrees in the Polytechnic University of Catalonia (UPC). He started studying computer science in 2003 and coursed a 3 years technical enginery in computer science. After that, he spent two years studying computer science engineering.

During these two years he got specialized in High Performance Networks and Operating Systems. In 2008 he spent one extra year doing a Master in Information of Technology where he was specialized in computer architecture. After that, he worked as an intern in the Intel Barcelona Research Center (IBRC) with Antonio González, and combined the internship with PhD courses for 1 year and a half. He currently combines a Researcher

position at UPC pursuing a PhD directed by Carlos Molina and Raul Martínez and supervised by Antonio González with a part-time job in the European Project CLERECO-IBM and AIRBUS.

#### Talk “Improving the prefetching performance through code region profiling”

We propose a new technique to improve the performance of hardware data prefetching. This technique is based on detecting periods of time and regions of code where the prefetcher is not working properly, thus not providing any speedup or even producing slowdown. Once these periods of time and regions of code are detected, the prefetcher may be switched off and later on, switched on. To efficiently implement such mechanism, we identify three orthogonal issues that must be addressed: the granularity of the code region, when the prefetcher is switched on, and when the prefetcher is switched off.

### Mireia Peral

Institute of Earth Science Jaume Almera (CSIC)



Mireia's research interests lie on modelling geodynamic processes. She obtained the Bachelor degree in Physics and a Master in reservoirs Geology and Geophysics in 2013 at the University of Barcelona. The master's final project was focused on the use of gravity gradients from GOCE satellite data in the interpretation of major crustal and lithospheric structures. Nowadays, she is completing her educational background with a PhD thesis at the Institute of Earth Science Jaume Almera (CSIC). Her research project, motivated by a recent model evolution of the Western Mediterranean region, requires



among others the application of principles of study of plate tectonics, fluid dynamics, rheology of the Earth, subduction processes and numerical and analogue modelling.

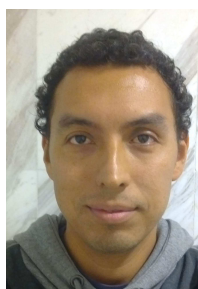
### **Talk “Dynamics of double-polarity subduction process: application to the Western Mediterranean”**

The Western Mediterranean tectonic setting and history is studied by means of three-dimensional numerical models. Goals are to understand its dynamics and to test the feasibility of a double-polarity subduction process that could have a key importance in the complex setup of this area.

The physical models, the numerical techniques involved, and some results in two and three dimensions are presented in this talk.

### **Víctor Valverde**

**Barcelona Supercomputing Center**



Víctor is an environmental scientist interested in the relationship between anthropogenic activities and their impact on ecosystems and human health. He holds a Bachelor in Environmental Biology from the Alicante University. He pursued his studies in Madrid where he studied an Applied Meteorology Master at the Spanish MetOffice. Moreover, he gained research experience at the Ecology Department of the Complutense University, where he conducted specialized studies on Transport Processes in nature and Ecological Modelling. After working for five years as an environmental consultant in the private sector he enrolled as doctoral student at the Earth Sciences Department of the BSC in 2012. His PhD aims at characterizing air pollution dynamics in Spain under different synoptic meteorological conditions by means of air quality modelling.

### **Talk “Characterization of coal power plants plume dynamics under typical synoptic conditions over the Iberian Peninsula”**

The talk will show the pollution plumes of seven Spanish coal power plants under the most typical meteorological conditions that affect the Iberian Peninsula at synoptic scale. The aim is to understand how meteorology modulates the plume dynamics (length, altitude, orientation) and their contribution to NO<sub>2</sub> and SO<sub>2</sub> surface concentration. Using the BSC-ES operational air quality forecasting system (CALIOPE-AQFS), the behaviour of the plumes is analysed for a representative day of each of the six most common synoptic situations. The results show that the plumes from Atlantic facilities are mainly driven by synoptic conditions whereas for power plants located over the Mediterranean and on mountainous regions, mesoscale dynamics dominate. Moreover, when the injection of the pollutants is done within the planetary boundary layer there is an increase in the NO<sub>2</sub> and SO<sub>2</sub> surface concentrations close to (<15-20 km) the sources.

### **Manel Mondelo-Martell**

**Universitat de Barcelona**



He started his Bachelor Degree in Chemistry at the University of Barcelona in 2008 with a strong interest for theoretical and computational chemistry. In my fourth year he earned a collaboration internship in the Physical Chemistry Department, under the supervision of Dr Huarte-Larrañaga, and made a first introductory course on computational chemistry. This experience suited him, and therefore he enrolled in a Master course on Theoretical and Computational Chemistry offered by URV, together with UB, UAB and UdG; again under the supervision of Dr Huarte-Larrañaga. Finally he started the PhD studies in the Nanoscience program at University of Barcelona in 2013 following the topic of my Master thesis, i.e. the effects of quantum confinement in the

dynamics and structure of light molecules. He received scholarships to make my PhD by the UB, the Catalan and the Spanish Governments, accepting finally the latter.

### **Talk “Quantum dynamics study of the hydrogen molecule confined in single-walled carbon nanotubes”**

We developed a full-dimensional study of the dynamics of a hydrogen molecule confined in a narrow Single-



walled Carbon Nanotube (SWCNT) is performed by using the Multiconfigurational Time-dependent Hartree approach. New insights on the coupling between the different degrees of freedom of the molecule during the diffusion along the nanotube are found and discussed.

### Lluís Raich

Universitat de Barcelona



Lluís Raich is a PhD student at the University of Barcelona (UB, Departament de Química Orgànica).

He did a B.Sc. in Chemistry at the same university and afterwards he obtained an interuniversity M.Sc. in Theoretical and Computational Chemistry by Universitat Rovira i Virgili, Universitat de Girona, Universitat Autònoma de Barcelona and Universitat de Barcelona. His research is focused on carbohydrate-active enzymes and in particular how to convert glycoside hydrolases (enzymes that break glycosidic bonds) into transglycosidases (enzymes that form glycosidic bonds).

#### Talk “Sugar Conformations that Enhance Cleavage of Glycosidic Bonds in Carbohydrate-Active Enzymes”

Carbohydrate-active enzymes select particular conformations of one sugar of their carbohydrate substrates to enhance catalysis. The evolution of this conformation during catalysis, known as the catalytic itinerary, is of fundamental interest in glycobiology. Here we calculate the free energy landscape of  $\beta$ -xylose to predict the possible catalytic itineraries of  $\beta$ -xylan enzymes. Our results discard one of the three catalytic itineraries proposed on the basis of X-ray structures of mutant enzymes. Additional classical and quantum mechanics/molecular mechanics (QM/MM) calculations indicate that complexes obtained from modified enzymes do not necessarily represent the structures that take place in the reaction coordinate.

### Chiara Pallara

Barcelona Supercomputing Center



Chiara Pallara received her degree in Pharmaceutical Chemistry, at the University of Bologna in 2009. In 2010, she joined the Protein Interactions and Docking Group, led by Juan Fernandez-Recio, as PhD student.

During her thesis she gained experience in theoretical chemistry and structural biology, especially in the fields of protein-protein docking and molecular dynamics methods.

Her main research line is currently focused on the development and optimization of computational protocols to integrate conformational flexibility into docking simulations for the characterization of protein-protein interactions. In addition she contributed to several multidisciplinary research projects involving the application of standard molecular dynamics simulations on different protein systems to understand the mechanistic effects of specific

mutations, ligands binding or proteins association.

Together with the other PID group members she participated in the last CAPRI edition, a world-wide community experiment that aims to evaluate the current state of protein-protein docking prediction. Moreover, thanks to the prestigious Severo Ochoa mobility grant offered by BSC-CNS, during last year she performed a short-stay in the Biomolecular Modeling group led by Prof. Gervasio in the Department of Chemistry of UCL, where she gained experience in enhanced molecular dynamics methods to sample rare events in bio-molecular systems.

#### Talk “pyDock performance in 5th CAPRI edition: from docking and scoring to binding affinity predictions and other challenges”

In order to test pyDock, a docking scoring algorithm developed in our group, the PID (Protein Interaction and Docking) group of the BSC Life Science Department, we have participated in all the 15 targets (T46 to T58) of the 5th CAPRI edition (2010-2012). Our automated protocol confirmed to be highly successful to provide correct models in easy-to-medium difficulty protein-protein docking cases placing among the Top5 ranked groups out of more than 60 participants.



**Marta Munar****Barcelona Supercomputing Center**

Marta Munar is PhD student at the Computational Genomics Group in the Barcelona Supercomputing Center, under the supervision of Dr. David Torrents. She studied a BSc in Biotechnology at the University of Barcelona and a MSc in Bioinformatics at Brussels Univ. Furthermore, she did several stays where she learned a lot, either in a personal and professional way. In Germany she had a first contact with the field of Bioinformatics in Prof. Thomas Dandekar group, where she worked in the assembly and annotation of bacterial genomes and later in Brussels, with Dr. Jeroen Raes where she studied the antibiotic resistance genes in metagenomes.

She is currently working on deciphering the mechanism of chromosomal rearrangements in Cancer. She has been working in Chronic Lymphocytic Leucemia (CLL), in collaboration with Hospital Clinic de Barcelona and Universidad de Oviedo. Nowadays she is also working in the study of chromosomal rearrangements in more than 2000 tumors of different types of cancers in the ICGC PanCancer project.

**Talk “Characterization of complex chromosomal rearrangements in cancer genomes”**

## POSTER PRESENTERS

### Dani Gallart

Barcelona Supercomputing Center



Dani Gallart studied theoretical physics at Universitat de Barcelona (UB), he was granted with Catalunya-La Pedrera grant when he enrolled the master in nuclear engineering at ETSEIB (UPC) where he became a master student for the fusion group at CASE Dept. Presently, he is a PhD student at CASE Dept. in BSC under the supervision of ICREA Prof. Mervi Mantsinen. He is working in the nuclear fusion group and his main goal is the modelling of Ion Cyclotron Resonance Frequency Heating (ICRF). His MSc thesis is an analysis of ICRF at DEMO, the first fusion reactor that is intended to be commercial, i.e. connected to the grid.

#### Poster “Heating Bulk Ions in DEMO with ICRF waves”

Ion cyclotron resonance frequency heating (ICRF) is one of the auxiliary heating schemes presently envisaged for ITER and DEMO. In this poster we analyse the potential of ICRF waves to heat the fuel ions in DEMO. Our analysis is carried out for the EU DEMO design ( $B = 6.8$  T,  $I = 18.6$  MA,  $R = 9.25$  m,  $a = 2.64$  m) optimized for a maximum pulse length of 2.3 hrs using the ICRF modelling codes PION and TORIC. We focus on second harmonic heating for tritium and fundamental minority heating of  $^3\text{He}$  with a few percent of  $^3\text{He}$  in a 50%:50% D-T plasma. The dependence of the ICRF characteristics and the ICRF-accelerated ions on the ICRF and plasma parameters is investigated, giving special attention to the DEMO design point at a plasma temperature of 30 keV and an electron density of  $1.2 \cdot 10^{23}\text{m}^{-3}$ .

### Clara Gómez-García

Institute of Earth Science Jaume Almera (CSIC)



Clara is in the first year of her PhD at Institute of Earth Sciences “Jaume Almera” and it deals about Passive Seismology and related with Interferometric methods for imaging the shallow subsurface. She studied Physics degree at Universidad Complutense de Madrid and a two-year MSc degree in Geophysics at the same university. Recently, she has submitted an article as first author related with my MSv Thesis titled “Rock magnetic characterization of the mine tailings in Portman Bay (Murcia, Spain) and its contribution to the understanding of the bay infilling process” to the Journal of Applied Geophysics.

#### Poster “Rayleigh wave ellipticity measurements in the Iberian Peninsula and Morocco”

We combine Rayleigh wave ellipticity, or H/V (horizontal to vertical) amplitude ratios, measurements obtained using teleseismic earthquake recordings and ambient noise cross correlations to provide improved constraints on the crustal models across the Iberian Peninsula and Morocco. To obtain the H/V ratios, we use more than 250 shallow ( $h < 40$  km) teleseismic events with magnitudes  $M_w > 6.0$  recorded at more than 450 seismic stations. We also use all the multicomponent ambient noise cross correlations computed for each station pair for the time period 2010 to 2012. Periods between 20 and 100 seconds are investigated. We observe a good agreement between the uppermost geological features of the crust and the obtained Rayleigh H/V ratios, with low values in major mountain ranges and high ratios in sedimentary basins. Combination of Rayleigh H/V ratio measurements from both earthquakes and ambient noise data with phase velocities and other types of seismic data will help to better constrain the Earth’s structure at different crustal levels.

**Vincenzo Obiso**

Barcelona Supercomputing Center



Vincenzo has a BSc in Physics and Astrophysics and a Master's Degree in Theoretical Physics from the University of Rome "Sapienza". After graduated he worked for 6 months as trainee at the company of aerospace and satellite technology Thales Alenia Space (Turin, Italy). He started his PhD at the Earth Sciences Department of the Barcelona Supercomputing Center (Environmental Engineering, UPC) in 2013. The main objective of his research Plan is the evaluation of the impact of the Aerosol-Radiation Interaction on the short- and mid-term meteorological processes using the fully on-line coupled meteorology-chemistry model NMMB/BSC-CTM, after the implementation within it of a dynamic coupling between the aerosol module and the meteorological core.

**Poster "Implementation of Dynamic Aerosol-Radiation Interaction within the NMMB/BSC-CTM"**

I will present some results of my current work related with the implementation of the dynamic aerosol-radiation interaction within the NMMB/BSCCTM, an on-line coupled meteorology-chemistry model under development at the Earth Sciences department of the BSC. In particular, I will present some results of the experiment that I am carrying out during my stay at the department of Environmental Physics of the University of Heidelberg. This experiment concerns the calculation of aerosol single-particle optical properties with two different optical codes under different assumptions on microphysical and chemical properties of the particles. The goal is to understand how much different shapes and chemical compositions of the particles affect the optical properties averaged over an ensemble of different sizes and obviously to understand how the skills and the limitations of the two optical codes are.

**Verónica Torralba**

Institut Català de Ciències del Clima



Verónica studied Physics at Complutense University of Madrid (UCM), where she also achieved a Master in Meteorology and Geophysics. She joined the Climate Forecasting Unit (CFU) in IC3 in October 2013, where she was involved in the development and communication of climate services for energy. She is currently a PhD student working on different European Projects. My work focuses on the seasonal wind speed and wind power forecasting.

**Poster "State-of-the-Art Climate Predictions for Energy Climate Services"**

Seasonal predictions of 10-m wind speed can be used by the wind energy sector in a number of decision making processes. Two different techniques of post-processing are applied in order to correct the unavoidable systematic errors present in all forecast systems. Besides an assessment of the impact of these corrections on the quality of the probabilistic forecast system is provided.

**Sílvia Bonàs**

Barcelona Supercomputing Center

Sílvia Bonàs holds a Degree in Biotechnology (Universitat Autònoma de Barcelona) and MSc in Bioinformatics for Health Sciences (Universitat Pompeu Fabra).



The focus of interest of her thesis is the study of human genetic variability in order to achieve a better understanding of the aetiology of complex diseases such as Type 2 diabetes or asthma. Our main goal is exploiting tools such as genotype imputation or systems biology approaches in order to enhance the results from genome-wide association studies within the context of the description of the genetic factors related to complex traits.

**Poster "Identification of novel type 2 diabetes susceptibility loci through wholegenome imputation using sequencing based reference panels into 13.201 cases and 59.656 controls"**

## Sara Royuela

Barcelona Supercomputing Center



Sara Royuela is a PhD student at BSC. She's been working with the BSC Programming Models group for the past four and a half years. In the meantime, she's also been collaborating with the Center for Applied Scientific Computing at Lawrence Livermore National Laboratory, California. She is interested in advanced compiler techniques for code analysis and optimization. Specifically, she is researching in analysis techniques for enhancing users experience with the OpenMP and OmpSs programming models, as well as for enlightening runtime libraries by pushing more responsibilities into the compiler instead of letting the runtime library to do all the work. Sara is also interested in new education techniques and she wants to become a teacher someday.

### Poster "Parallel programming issues and what the compiler can do to help"

Twenty-first century parallel programming models are becoming real complex due to the diversity of architectures they need to target (Multi- and Many-cores, GPUs, FPGAs, etc.). What if we could use one programming model to rule them all, one programming model to find them, one programming model to bring them all and in the darkness bind them, in the land of MareNostrum where the Applications lie. OmpSs programming model is an attempt to do so, by means of compiler directives.

Compilers are essential tools to exploit applications and the architectures the run on. In this sense, compiler analysis and optimization techniques have been widely studied, in order to produce better performing and less consuming codes.

## Roger Hernandez

Barcelona Supercomputing Center

Roger studied Computer Science at Universitat Politècnica de Catalunya and is currently working in the Severo Ochoa project in the Barcelona Supercomputing Center, researching into big data and supporting the other departments with our findings.

### Poster "Automatic Query Driven Data Modelling in Cassandra"

Non-relational databases have recently been the preferred choice when it comes to dealing with Big Data challenges, but their performance is very sensitive to the chosen data organisations. We have seen differences of over 70 times in response time for the same query on different models. This brings users the need to be fully conscious of the queries they intend to serve in order to design their data model. The common practice then, is to replicate data into different models designed to fit different query requirements. In this scenario, the user is in charge of the code implementation required to keep consistency between the different data replicas. Manually replicating data in such high layers of the database results in a lot of squandered storage due to the underlying

system replication mechanisms that are formerly designed for availability and reliability ends. We propose and design a mechanism and a prototype to provide users with transparent management, where queries are matched with a well-performing model option. Additionally, we propose to do so by transforming the replication mechanism into a heterogeneous replication one, in order to avoid squandering disk space while keeping the availability and reliability features. The result is a system where, regardless of the query or model the user specifies, response time will always be that of an affine query.



## Cecilia Jaramillo

Universitat Autònoma de Barcelona



Cecilia is PhD student of High Performance Computing at Universitat Autònoma de Barcelona (UAB). She received my master degree from the Computer Architecture and Operating Systems (CAOS) at UAB and a BSc degree in Computer Science from Universidad Técnica Particular de Loja (UTPL) from Ecuador. My research interests include modelling and simulation, agent-based models, and parallel computer architectures.

### Poster “Modelling the Contact Propagation of Nosocomial Infection in Emergency Departments”

The nosocomial infection is a special kind of infection that is caused by microorganisms acquired inside a hospital. In the daily care process of an emergency department, the interactions between patients and sanitary staff create the environment for the transmission of such microorganisms. Rates of morbidity and mortality due to nosocomial infections are important indicators of the quality of hospital work. In this research, we use Agent Based Modeling and Simulation techniques to build a model of Methicillin-resistant Staphylococcus Aureus propagation based on an Emergency Department Simulator which has been tested and validated previously. The model obtained will allow us to build a contact propagation simulator that enables the construction of virtual environments with the aim of analysing how the prevention policies affect the rate of propagation of nosocomial infection.

## Guray Ozen

Barcelona Supercomputing Center



Guray Ozen is PhD student at Barcelona Supercomputing Center, since 2014. His research is to explore the concepts of compiling, optimizing and code generating for hardware accelerators. The key contribution is MACC, directive based emergent programming model compiler for GPUs. He received my BSc degree from the Department of Computer Engineering at Dokuz Eylul University in Turkey. In 2014 and his MSc degree from Universitat Politècnica de Catalunya BarcelonaTech with High-Performance Computing title.

### Poster “MACC: Mercurium ACCelerator Compiler”

GPU Offloading is emergent programming model. OpenMP includes in its latest 4.0 specification the accelerator model. In this paper we present a newly implementation of this specification while generating "native" GPU code in the OmpSs programming model developed at the Barcelona Supercomputing Center. Focused on targeting NVIDIA GPUs, our work is based on an OpenMP 4.0 implementation in the Mercurium source-to-source compiler infrastructure referred MACC (Mercurium ACCelerator compiler).

## Ismael Rodríguez-Espigares

Barcelona Supercomputing Center



Ismael Rodríguez-Espigares is a PhD in Biomedicine student at Universitat Pompeu Fabra (UPF). He received a M.Sc. degree on Bioinformatics from UPF in 2014 and a Biotechnology degree from Universitat Autònoma de Barcelona in 2012.

He works at Jana Selent's subgroup at Pharmacoinformatics group from the Research Programme on Biomedical Informatics (GRIB). His research group studies the mechanism of action of drugs that target to G-protein coupled receptors (GPCRs) in order to propose new drug candidates for central nervous system diseases. Currently,



he is focused on the development of a pipeline to analyse GPCR micro-domains dynamics through Markov State Models for the development of new antipsychotic drugs for schizophrenia treatment.

Ismael is one of the authors of MEMBPLUGIN, a free available Visual Molecular Dynamics (VMD) extension for the analysis of membrane bilayer properties in molecular dynamics simulations.

### **Poster “Capturing the action of current antipsychotic drugs at G protein-coupled receptors (GPCRs) by means of Markov State Model analysis”**

The aim of this project is to decipher the molecular mechanism of current antipsychotic drugs targeting G-protein coupled receptors (GPCR) considering the novel concept of a coupled allosteric network being responsible of signal propagation. For this purpose, in a first step, we simulated the antipsychotic action under physiological conditions, and in a second step, we applied Markov State Model (MSM) analysis in order to extract metastable states that are relevant for the action of antipsychotic drugs. Exploiting the structural insights obtained, the ultimate aim is to discover original pharmacological interventions and suggest new lead structures for antipsychotic therapy.

## **Alfonso Santiago**

### **Barcelona Supercomputing Center**



He was born in Bahía Blanca, in “La Patagonia Argentina”, he got his degree in Bioengineering the 21th of September of 2014, with the presentation of his thesis entitled “Modelización de la unidad microeyectora para nebulizadores de alta eficiencia por membrana perforada vibrante” (Modelization of the micro-ejecting unit for high efficiency nebulizers by vibrating mesh technology) with a grade of 10 over 10. He moved to Barcelona six months after getting my degree to work in his Ph.D. in the Barcelona Supercomputing Center. During this research he keeps working with computational modelling, but now in the biomechanical area rather than industrial applications. He is involved in the project of the computational heart, coupling a model

of the arterial network in an international cooperation project. He prefers working in group to learn from his colleagues.

### **Poster “A 3D-1D cardiac-vascular computational feedbacked model”**

This work presents a first version of an HPC-optimized cardiovascular coupled model using two already developed models. The two modelling tools are Alya, the BSC (Barcelona, Spain) in-house tool for biomechanical simulations; and ADAN55, the LNCC (Petropolis, Brasil) model of the arterial blood flow circulation.

Alya is a multiphysics, HPC code that allows modelling the heartbeat problem using the Finite Element Method (FEM). The ADAN55 is a reduced 55-artery version of the Anatomically Detailed Arterial Network (ADAN) model, which takes into account more than 1500 arteries from an average human body. Both codes are coupled by a black-box decomposition approach. At each time step the Fluid Structure Interaction (FSI) problem is solved in the ventricle, computing the output flow through the aorta. This flow is passed to the arterial network model that solves the equations in the ADAN55 and, after that, the pressure response is given back to Alya.

In previous generations of the involved models, each one had predefined boundary conditions: for Alya a pressure value was imposed on the endocardium, and a flow contour was imposed in the aortic root of ADAN55 model. Attaching an arterial network model to the Alya computational heart allows having a more physiological response to the blood ejection, as the systemic impedance is a complex result of the entire 1D arterial model. In the same way, using the heart model output flow connected to the ADAN55 input, offers the possibility to obtain a physiological stimuli that varies with changes in the arterial system.

A cardiovascular simulator will allow bioengineers and medical researchers to study the response of the full model to changes in any of the components, extending the predictive and descriptive capabilities of the standalone versions of the models, and providing further insight in the cardiac-systemic interactions.



## Sergio Natan González

Barcelona Supercomputing Center - CONACYT



Sergio is Chemical engineer (1991) at Universidad Veracruzana (UV) in Poza Rica, Veracruz, Mexico; specialist in quality control management, MSc in computational science and MSc in environmental science at UV; PhD in environmental management for development (2012) from Universidad Popular Autónoma de Veracruz (UPAV); Head of the Chemical science Faculty between 2002 – 2006; Project Manager at UV in 2007 – 2009; full professor and academic technician since 1994 in bachelor's degree in chemical engineering, environmental engineering and petroleum engineering; teacher at post grade at UV in the MSc in environmental management for sustainability in the Chemical Science Faculty in Poza Rica, Veracruz, Mexico between 2012-2013, and

Universidad Interamericana para el desarrollo (UNID) at the MSc in Information Technologies in networks, the MSc in Information Technologies for education and MSc in Quality management; post grade teacher in UPAV in Poza Rica, Veracruz, México in 2013. supervisor of thesis and receptional works at bachelor's degree and MSc degree at UV and UNID; author, co-author and collaborator of articles in the environmental field; participant in several international congress in Mexico and La Habana, Cuba, author and co-author of two books, two chapter books and diffusion material on iTunes related to air pollution, water quality index, ecology construction and sustainability; nowadays in a postdoctoral fellowship by CONACYT Mexico in the BSC – CNS Earth Sciences Department.

### Poster “Evaluation of Modelling Systems in High Resolution to Assess the Air Pollutant Impacts on Human Health”

Nowadays the modelling of systems in high resolution is being used for air quality and other forecasting applications, where a spatial area is related with different interrelated variables that could be displayed on a map. This area is usually represented by global domains (hundreds to thousands of square km); when smaller regions need to be represented, a high resolution modelling system can be used, these systems goes from one square km to dozen of square km, health is one of these issues where this kind of resolution can be used. In Europe, Asia, North America, South America and other countries, health problems related with the air pollution and climate change is a concern for individuals and world organizations like the WHO; today studies show the relation between morbidity and mortality rates, air pollution and effects on human health; these modelling systems in high resolution help us to simulate scenarios and propose solutions to this problematic. So the objective of this work is to evaluate the system performance WRF – CMAQ and CALIOPE on high resolution (4 km x 4 km) to determine air pollutant impacts of PM10, PM2.5, Ozone, NO2 and SO2 on population, using BenMAP for assess impact on health. The methodology suggested is the time series analysis of two years of hospital admissions, morbidity and mortality rates and the air quality forecasting of the cities selected, previously modelled in WRF, CMAQ and CALIOPE; after that, the Response Functions (DRF/ERF) to determine the impacts on health and the BenMAP software will be used. It is expecting find the scenarios that could decrease the mortality and morbidity rates in diseases like lung cancer, chronic respiratory obstructive disease, asthma, and the acute respiratory diseases in adults and children under ten years old.

## Daniel Lecina

Barcelona Supercomputing Center



Daniel studied physics in the Universitat de Barcelona. Afterwards he obtained the master's degree in computational and applied physics in a joint collaboration of the Universitat Politècnica de Catalunya and the Universitat de Barcelona. He is currently enrolled in the physics PhD program, under the supervision of Victor Guallar.

### Poster “Binding Free Energy and Ligand Orientation Calculations using A Monte Carlo Method with Markov State Analysis”

Computing binding free energies has great implications in drug design. Using PELE technique, it has been shown that one can get quick and accurate estimations by



means of a Markov state model<sup>3</sup>. We improved our methodology to compute faster binding relative free energy differences, mainly by analysis reducing the sampled region. This possibility opens a way in all-atom drug lead optimization by efficiently scoring a list of potential candidates in terms of binding affinities (approximately in 24hours), while still modelling accurately the protein-drug induced fit. Furthermore, we added information of the ligand orientation allowing us to obtain a better insight of the entrance mechanism. First, we show benchmark results - a series of benzamidine-like inhibitors in trypsin. Then, we apply our method to a more realistic scenario: the binding to a glucocorticoid receptor, and we show the performance for a new benchmark with a larger range of binding free energies (~14 kcal/mol). Simulations are obtained with our new in-house code PELE++.

## Edgar Olivares

### Barcelona Supercomputing Center

Edgar is a PhD student in Barcelona Supercomputer Center working in the department of Computer applications in Science and Engineering. He works on particles transportation in a fluid applied to high performance computational. He mainly applies it to respiratory system, where the insertion of particle find several medicine interests, such as nasal drug's delivery or corrosive particles inhalation as illustrative examples. He is from Barcelona and grew up in a outskirts town called Alella. He holds a BSc in Physics from Universitat de Barcelona (2010) and a MSc in Computational Physics in Universitat Politecnica de Catalunya (2013).

### Poster "Asynchronous parallel fluid solver and particle transport"

This work proposes a parallel numerical strategy to solve fluid dynamics problems and transport Lagrangian particles. There exist two main approaches to address the coupling. The first one consists in transporting the particles in post process, once the solution have been written on disk. This solution is not suitable for simulations with very large meshes and lots of time steps. In our case, in particular, we are simulating respiratory system. From nose to lungs with 5th order bronchial generation, this implies working with a mesh with 500 millions of elements. In this case, we need to write 100Gb every time step, we easily simulate 500 time steps. The amount of data we need to write in the disk makes us opt for the second option. The second option consists in solving the particle transport just after the Navier-Stokes solution at each time step. The most common strategy in the literature uses the same code and resources to solve the fluid equations and to transport the particles.

In this work, we follow a multi-code approach which enables to solve the fluid and particle asynchronously. In addition, the particles are parallelized in a hybrid mode, using both MPI and OpenMP, together with a dynamic load balance library on this second level of parallelism.

## Matías Rivero

### Barcelona Supercomputing Center

Matías Rivero is a PhD student at the Computer Applications in Science and Engineering (CASE) Department of the Barcelona Supercomputing Center - Centro Nacional de Supercomputación de España

(BSC-CNS), under the supervision of Mariano Vázquez. His research is focused on HPC-based Computational Contact and Impact Mechanics, Fluid Structure Interaction and Coupled Problems. He holds a bachelor degree in Nuclear Engineering from Instituto Balseiro and a master degree in Control and Numerical Simulation from Universidad de Buenos Aires, Argentina. Before joining the BSC, Matías has worked as R&D engineer in the Atucha II Nuclear Power Plant project for 4 years, developing a coupled calculation suite for operational transient simulations.



### Poster "Implementation of a coupled algorithm for the solution of deformable two-body contact problem"

The poster presents a parallel iterative method for numerical solving frictionless contact problem for two elastic bodies. Each iterative step consists of a Dirichlet problem for the one body and a Neumann problem for the other in order to enforce the contact boundary conditions.





## Marcelo Amaral

Barcelona Supercomputing Center



Marcelo Amaral is currently enrolled in a PhD program at UPC and working in BSC with resource management (scheduling and provisioning) in cluster environment. This actual project is sponsored by IBM T. J. Watson Research. He concluded in the early of 2014 a Master degree about availability analysis for energy-aware network management, which project was sponsored by Ericsson Research Sweden. Before that, he worked with policy driven network management system; cache management in Cloud Computing environment; and IPTV over P2P networks.

### Poster “Resource Management for Software Defined Data Centers for Heterogeneous Infrastructures”

Software Defined Data Center (SDDC) provides more resource management flexibility since everything is defined as a software, including the network as Software Defined Network (SDN). Typically, cloud providers overlook the network, which is configured in static way. SDN can help to meet applications goals with dynamic network configuration and provide best-efforts for QoS. Additionally, SDDC might benefit by instead of be composed by heavy Virtual Machines, use light-weight OS Containers. Despite the advantages of SDDC and OS Containers, it brings more complexity for resource provisioning. The goal of this project is to optimize the management of container based workloads deployed on Software-defined Data Centers enabled with heterogeneous network fabrics through the use of network-aware placement algorithms that are driven by performance models.

## Cesare Cugnasco

Barcelona Supercomputing Center



Cesare Cugnasco, born the 26th of March 1988 in Turin (Italy), is a Computer Engineer currently employed at the Barcelona Supercomputing Center as a support engineer for Big Data platforms in the Autonomic Systems and eBusiness Platform group. The focus of his work has been studying, and enhancing the usage of non-relational data stores in order to support scientific applications. In the context of the Severo Ochoa project, he is carrying out collaborations with the CASE and the LIFE Departments of BSC.

Cesare obtained a bachelor's degree in Computer Engineering at the Politecnico of Torino in 2011 and a Master in Software Engineering in 2014 with a thesis developed at the BSC thanks to a grant from the ERASMUS project.

He is currently working on his Ph.D. from the Department of Computer Architecture of the Universitat Politècnica de Catalunya focusing on multidimensional indexes on key-value data stores.

### Poster “A framework for multidimensional indexes on distributed and highly-available data stores”

No-relational databases are nowadays a common solution when dealing with huge data set and massive query work load. These system have been redesigned from scratch in order to achieve scalability and availability at the

cost of providing only a reduced set of low-level functionality, thus forcing the client application to implement complex logic. As a solution, our research group developed Hecuba, a set of tools and interfaces, which aims to facilitate programmers with an efficient and easy interaction with non-relational technologies.

This poster presents a part of Hecuba related to a specific missing feature: multidimensional indexing. Our work focuses on the design of architectures and the algorithms for providing multidimensional indexing on a distributed database without compromising scalability and availability.

## Daniel Nemirovsky

Barcelona Supercomputing Center

He received a BSc in Electrical Engineering and Political Science from the University of Michigan (Ann Arbor, MI) in 2008 and a MSc in Computer Architecture and Networking Systems from the Polytechnic University of



Catalonia (UPC) in 2013. Currently he is PhD student at the UPC and a researcher with the Barcelona Supercomputing Center with interests in heterogeneous systems, dynamic hardware supported scheduling mechanisms, analytical models, and non-conventional architectures.



**Poster “Mathematical Representation of the Hardware Round-Robin Scheduler Analytical Model for Single-ISA Heterogeneous Architectures”**

Introduction Over the past few years, hardware chip manufacturers have been shifting strategies to overcome the well-known Power Wall which imposes the practical limitations to increasing CPU performance by rising the processor’s running frequency. The most recently applied schemes to exploit high single and multi-threaded performance at lower energy costs have been to increase the number of cores per chip as well as their diversity

leading to what is commonly referred to as heterogeneous multi or many core architectures. In spite of the visible gains the architectures offer, heterogeneity poses significant challenges to the OS, one of the most critical being thread scheduling.

**Dmitry Repchevsky**

Barcelona Supercomputing Center



Holding the engineering degree in Automated Systems of Information Processing and Control (Computer Science) from St. Petersburg’s Electrotechnical University, he started his career in RUBIN State Research Institute (St. Petersburg). Participation in European Visiting Scientists Program brought me to the Instituto Tecnológico de Aragon (ITA) where he participated in IBCoBN project. Since then he participated in "HAS Video" and "IST@HOME" European e-health research projects as a lead developer and a responsible for the work package accordingly.

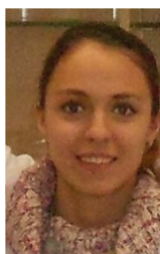
Since 2007 he is part of the Computational Bioinformatics node of Life Science department in the Barcelona Supercomputing Center. He holds a MSc Bioinformatics degree from Pompeu Fabra University (UPF) and at the moment he is PhD candidate in the University of Barcelona

**Poster “Galaxy Gears: Web Services integration into Galaxy workflows”**

Modern computational biology analyses are usually comprised of different tasks and involve many software tools. Such multi-step operations are usually denoted as workflows. Growing complexity of biological pipelines led to creation of several platforms that facilitate workflows creation, management and execution. Galaxy and Taverna are indeed the most remarkable workflow systems in bioinformatics. While Taverna is essentially oriented to Web services, Galaxy targets local task execution. Web services integration into Galaxy is not a trivial task and requires many efforts from workflow developers. Galaxy Gears provides seamless integration of Web services into Galaxy pipelines thus extending the number of tools that can participate in complex computational pipelines.

**Marina Corbella**

Universitat de Barcelona



After obtaining her degree in Chemistry in the Universitat Autònoma de Barcelona she moved to the Faculty of Pharmacy of the Universitat de Barcelona, where she performed a MSc thesis in the synthesis of compounds as potential antitumor agents. Now, she is performing her PhD in Computational Biology and Drug Design group under the supervision of Dr. Carles Curutchet. Her research topic is charge transfer in biological systems such as DNA and phycobiliproteins (antenna proteins involved in light-harvesting processes).

**Poster “Single amino acid mutation controls hole transfer dynamics in DNAmethyltransferase HhaI complexes”**

Different mutagenic effects are generated by DNA oxidation that implies the formation of radical cation states (so-called holes) on purine nucleobases in the  $\pi$  stack. The interaction of DNA with proteins may protect DNA from the oxidative damage owing to hole transfer (HT) from the stack to aromatic amino acid residues. The HT



dynamics in such system is still poorly understood. Here, we report a computational study of HT in DNA complexes with methyltransferase HhaI and its mutant Q237W, which were experimentally investigated in the Barton group. We employ a combined approach based on molecular dynamics simulations and quantum mechanical calculations to estimate the rate for all individual steps involved in the HT pathways; finally the overall HT kinetics are explored using the Monte-Carlo method. Our results indicate that the HT characteristics are strongly affected by structural deformations of DNA upon its binding to the protein. In the wild-type enzyme complex, a Gln residue inserted in the DNA  $\pi$ -stack is shown to destabilize the radical cation states of neighbouring guanines and thereby inhibits the long-range HT in line with experimental findings. In contrast, the HT is estimated to be quite fast in a complex of the Q237W mutant where Trp237 stabilizes hole states on the adjacent G bases and enhances the electronic coupling of these sites. An alternative HT pathway that implies the formation of a Trp<sup>+</sup> radical is predicted to be less efficient. Our study provides a consistent molecular picture on how long-range HT in DNA protein complexes is controlled by amino acids closely interacting with the  $\pi$  stack.

### Marina Cañellas

Barcelona Supercomputing Center



Marina performed a Bachelor degree in Biotechnology in “Universitat Rovira i Virgili”, Tarragona. While she was doing it, she discovered the protein modelling world, and she decided to do a master about it. Thus, she performed a master in Bioinformatics for genomics and drug design in “Universitat Autònoma de Barcelona”, having the chance to accomplish the master practices in Almirall pharmaceutical company and to perform a really interesting study: “Protein-protein interactions: methodology for peptide and small molecule inhibitor design”.

Now, she is doing my PhD in BSC (Barcelona Supercomputing Center), Barcelona, and she is having a really great time. Her research project is about doing computational rational design of hemeperoxidases and peroxygenases in order to use them for industrial and technological applications, using methodologies so as Protein Energy Landscape Exploration (PELE), Molecular Dynamics and Quantum Mechanics/Molecular Mechanics (QM/MM) simulations.

#### Poster “Catalytic surface radical in dye-decolorizing peroxidase”

Dye-decolorizing peroxidase (DyP) of *Auricularia auricula-judae* has been expressed in *E. Coli*. The crystal structure of DyP shows a buried haem cofactor, and surface tryptophan and tyrosine residues potentially involved in long-range electron transfer from bulky dyes. Simulations using PELE software provided several binding-energy optima for the anthraquinone-type Reactive Blue 19 (RB19) near the above aromatic residues, and the haem access-channel. Subsequent QM/MM calculations showed a higher tendency of Trp-377 than other exposed haem-neighbour residues to harbour a catalytic protein radical, and identified the electron transfer pathway. Computational calculations, coupled with experimental studies performed by our collaborators, enable us to elucidate surface exposed residues responsible for the oxidation of RB-19 dye in AauDyP peroxidase.

### Carolina Estarellas

Barcelona Supercomputing Center



Carolina Estarellas obtained her degree in Chemistry in the University of the Balearic Islands (UIB) on 2008. A year later she completed a Master of Chemical Science and Technology and on 2012 obtained her PhD in the area of computational organic chemistry under the supervision of Profs. Pere M. Deyà and Antonio Frontera. She did a postdoctoral stay with Prof. Jiri Sponer at Masaryk University from October 2012 to October 2013 and then moved to Barcelona where she started to work in the group of Prof. F. Javier Luque as an associate professor of the Physical Chemistry department in the Food Science Campus of the University of Barcelona and became a Juan de la Cierva postdoctoral fellow on January 2015. Her research develops in two different

lines, one focused on reactivity studies using Quantum Mechanics calculations and another on the study of biological systems by means of Molecular Dynamics Simulations.



### Poster “Theoretical study of the activation mechanism of AMP-kinase by means of Molecular Dynamics Simulations”

Mammalian AMP-activated protein kinase (AMPK) is a Ser108/Thr132 protein kinase with a key role as sensor in the cellular energy homeostasis. It is a heterotrimeric enzyme complex composed by a catalytic  $\alpha$ -subunit and two regulatory subunits known as  $\beta$  and  $\gamma$ . It is regulated by several mechanisms, including indirect activators such as metformin, rosiglitazone and resveratrol, and direct activators, such as compound A-769662. Some of these activators show a clear specificity for a particular AMPK subunit, which provide a possible way to design new compounds that could activate AMPK in specific tissues, minimizing the putative negative effects of AMPK activation at whole body level.

### Marta Guindo

Barcelona Supercomputing Center



She studied both biologist and biochemistry at Universidad de Navarra and obtained her master's degree in Biomedical Research at Universitat Pompeu Fabra in Barcelona in 2013. After been working on the wet lab for several years, she followed her interests of answering biological questions from a holistic and quantitative point of view. Since February 2014, she is a PhD student at the Computational Genomics group led by David Torrents under Josep M. Mercader and David Torrents supervision at the Barcelona Supercomputing Center (BSC). Her project is focused on the genomics of complex diseases aiming to identify new associated variants and comorbidities through advanced imputation, statistical and systems biology techniques. Thereupon, up to 69

available GWAS datasets comprising 44 different diseases and more than 400.000 samples will be analysed, which expect will provide the identification of key pathways and processes involved in the aetiology of complex diseases.

### Poster “GWImp-COMPSs: An Integrated Framework for Large-scale Genome-wide Imputation and Association Testing”

Current Genome-wide association studies (GWAS) strategies are based on complex, fragmented and environment specific workflows that require the combination of tools with the continuous intervention of expert users. As a solution to these limitations, we developed GWImp-COMPSs, an integrated tool that performs phasing and genotype imputation together with association testing in a single execution and without user intervention.

### Víctor López

Barcelona Supercomputing Center

He studied Mathematics and Computer Science in Universitat Politècnica de Catalunya with a CFIS (Center for Interdisciplinary Studies) fellowship and got my degrees in 2013 and 2014. He worked as a software engineer in the UPC spin-off Talaia Networks since February 2013 to August 2014. On September 2014 he joined the BSC Life Sciences department starting my PhD in Bioinformatics with a La Caixa-Severo Ochoa fellowship, Identification of key pathways and processes involved in the aetiology of complex diseases.

### Poster “PMut2015: a web-based tool for predicting pathological mutations on proteins”

Assessing the impact of amino acid mutations in human health is an important challenge in biomedical research. As sequencing technologies are more available, and more individual genomes become accessible, the number of identified variants has dramatically increased. PMut, released back in 2005, has been one of the popular predictors in this field. PMut was a neural-network based classifier using sequence data to provide a pathology

score for point mutation in proteins. We present here a new, revised, and much more powerful version of the predictor. PMut2015 introduces the use of state-of-the-art machine learning algorithms and an updated training



set based on SwissVar. It achieves an accuracy of 80%, performs 60% better than random. PMut2015 includes a fully featured training and validation engine that can be optimized to generate predictors adapted to user specific training sets. The engine is implemented in python using MongoDB engine for data management. It has been adapted to run at the HPC level to cover large scale annotation projects.

## Constantí Seira

### Barcelona Supercomputing Center



He got the BSc in Pharmacy (University of Barcelona) between 2007 and 2012, after that he obtained the Master in Research, selection and development of drugs (Pharmacy's Faculty of Barcelona) among 2012 and 2013, in autumn 2013 he started my PhD in Computational Chemistry under the supervision of Dr's Luque and Bidon-Chanal. During my BSc he made practical work entitled "Synthesis of Lycopodine & Robinson aza Michael ring size analogs" at the Organic Chemistry Department of the Faculty of Pharmacy under the supervision of Dr's Bonjoch and Bradshaw, and pending the Master he completed my final project at chemistry department (also in Faculty of

Pharmacy, University of Barcelona), entitled "molecular modelling of cdk4 inhibitors" in the Dr Luque's group.

### Poster "Influence of Temperature on the Topological Features of Inner Cavities in Cytoglobin"

Cytoglobin (Cygb) is a novel member of the globin family in man, but there is no clear evidence about its biological function. Cygb exhibits a highly complex ligand rebinding kinetics, which agrees with the structural plasticity of the inner cavities and tunnels found in the protein matrix. In this work we have examined the effect of temperature on the topological features of Cygb. To this end, the structural and dynamical properties of human Cygb are compared with those determined for the Antarctic fish *Chaenocephalus aceratus*. The results support a distinct temperature-dependence of the topological features in the two proteins, suggesting different adaptations to cold and warm environments.

## Silvana De Souza Pinheiro

### Universitat de Barcelona

Scholarship student at CNPQ – Conselho Nacional de Desenvolvimento Científico e Tecnológico. Postdoc student at University of Barcelona (UB) in the Computational Biology Drug Design Group (CBDDG), where she works with Electronic Energy Transfer (EET) in model systems. Master in Genetics and Molecular Biology at University of Pará (UFPA/BRA). She worked with methods QM/MM and molecular dynamics to study the catalytic mechanism, in the Drug Planning and Development Laboratory (LPDF). Specialization has applicability of new technologies as teaching tools for science education at University of Pará (UFPA). Graduated in Full Degree in Chemistry -University of Pará (UEPA) and Pharmacy at University Center of Pará (CESUPA).



### Poster "Cation- $\pi$ -cation interactions in structural biology"

Biological structures are stabilized by a variety of noncovalent interactions, such as hydrogen bonds,  $\pi$ -stacking, salt bridges or hydrophobic interactions. Besides hydrogen bonds and  $\pi$ -stacking, cation- $\pi$  interactions between aromatic rings and positively charged groups have emerged as one of the most important interactions in structural biology. Although the role and energetic characteristics of these interactions is well established, a special case involving three molecular species, termed cation- $\pi$ -cation

interaction, is still poorly understood. In this contribution, we aim at advancing in the understanding of cation- $\pi$ -cation interactions and their role in the structure and stability of biosystems via two complementary approaches. The first one consists of a statistical study of the occurrence, composition and geometry of cation- $\pi$ -cation interactions identified on a non-redundant set of protein structures from the PDB (Protein Data Bank), which demonstrates that cation- $\pi$ -cation interactions are indeed common in proteins. We also analyse the degree of conservation of the



interactions by inspection of similar sequences obtained through the sequence alignment tool BLAST. The second part of the study consists of an energetic analysis of the most relevant interactions at the SCS-MP2/CBS level of theory, as well as an energy decomposition analysis for representative cases performed at the SAPT2 level. Besides the well-known deficiency of standard additive force fields to describe the relevant polarization contribution to cation- $\pi$  interactions, our results indicate that non-additive three-body contributions are significant in this case, implying that cation- $\pi$ -cation interactions constitute an even more challenging case expected to be dramatically misrepresented by standard additive force fields. In vacuum, the interactions identified are strongly repulsive. Further work is being carried out in order to understand the strength of cation- $\pi$ -cation interactions in a protein environment, where the cation-cation repulsion will be strongly screened.

## Guillem Alomar

Barcelona Supercomputing Center



Guillem Alomar, born the 13th of March 1989 in Palma de Mallorca (Spain) is a Computer Engineer currently employed in the Barcelona Supercomputing Center as junior developer in the Autonomic Systems and eBusiness Platform group. Guillem obtained a bachelor degree in Computer Science at the Universitat Pompeu Fabra in 2012 and in 2013 a master degree in High Performance Computing, Information Theory and Security (specialization in H.P.C.) at the Universitat Autònoma de Barcelona. He is currently working on the design and implementation of a tool that aims to facilitate programmers the development of parallel applications, so that a distributed DDBB can be efficiently used in the background in a transparent way.

### Poster “Hecuba: NoSql made easy”

Non-relational databases are nowadays a common solution when dealing with huge data set and massive query work load. These systems have been redesigned from scratch in order to achieve scalability and availability at the cost of providing only a reduce set of low-level functionality, thus forcing the client application to take care of complex logics. As a solution, our research group developed Hecuba, a set of tools and interfaces, which aims to facilitate programmers with an efficient and easy interaction with non-relational technologies.

## Jan Ciesko

Barcelona Supercomputing Center

He is working in the parallel programming models group at the Barcelona Supercomputing Center. His research is focused on programming model support for an algorithmic pattern that is known as reductions. Reductions represent non-atomic memory updates and occur in many scientific applications. Their concurrent execution on modern systems is not trivial and requires careful handling to achieve scalability and correctness. Prior to my research position he studied computer science at the Friedrich-Alexander University in Germany and initiated several mobile app and business ventures including Numerical Monkeys and Realizr.net.



### Poster “Scaling Irregular Array-type Reductions in OmpSs”

Array-type reductions represent a frequently occurring algorithmic pattern in many scientific applications. A special case occurs if array elements are accessed in a non-linear, often random manner, which makes their concurrent and scalable execution difficult. In this work we present a new approach that consists of language- and runtime support to facilitate programming and delivers high scalability on modern shared-memory systems for such irregular array-type reductions. A reference implementation in OmpSs, a task-parallel programming model, shows promising results with speed-ups up to 15x on the Intel Xeon processor.



## Milovan Duric

### Barcelona Supercomputing Center

He is in his last year as PhD student at the Department of Computer Architecture, UPC - BarcelonaTech, and research student at Barcelona Supercomputing Center. Before coming to Barcelona, he received the BSc (in

2008) and MSc (in 2010) from the Department of Computer Engineering and Computer Science, School of Electrical Engineering, University of Belgrade, Serbia.

His research interests are computer architecture and multicore system co-design (hardware acceleration and heterogeneous system processors, SIMD extensions and vector accelerators, coarse-grained reconfigurable compute accelerators).



### Poster “DLP Acceleration on General Purpose Cores”

High-performance and power-efficient multimedia computing drives the design of modern and increasingly utilized mobile devices. State-of-the-art low power processors already utilize chip multiprocessors (CMP) that add dedicated DLP accelerators for emerging multimedia applications and 3D games. Such heterogeneous processors deliver desired performance and efficiency at the cost of extra hardware specialized accelerators. In this paper, we propose dynamically-tuned vector execution (DVX) by morphing one or more available cores in a CMP into a DLP accelerator. DVX improves performance and power efficiency of the CMP, without additional costs for dedicated accelerators.

## Gabriel Fernandez

### Universitat Politècnica de Catalunya BarcelonaTech



He obtained his degree in Computer Science Engineering by Las Palmas de Gran Canaria's college in 2012, and then a Master in Research and Investigation on Informatics by the Universitat Politècnica de Catalunya in 2014. Since 2012 he has worked as student in the group CAOS from BSC under the supervision of Francisco J. Cazorla to pursue my PhD.

### Poster “Enhancing Timing Analysis for COTS multicores for the Space industry: a software approach”

Critical Real Time Embedded Systems (CRTES) are growing in complexity and performance requirements, making the use of multicores very appealing for the industry. Multicores provide better performance per watt, simplify the processor design and reduce size, weight and power-related costs. The drawback of this kind of processors is the increase of the complexity in the timing analysis given their shared resources. Our objective is to provide software techniques and methodologies to study and deal with on-chip contention in COTS multicores.

## Pilar Gómez

### Universitat Autònoma de Barcelona



She studied Computer Science (5 years degree) at Universitat Autònoma de Barcelona (UAB). In 2004, She received my diploma in Computer Science from UAB. She worked during some years at the private companies. At the same time, She was working as an associated professor at UAB teaching Computer Structure practices using Assembly language Motorola HC(S)12, Intel x86 and supervising many final projects. Now, she is a research Fellow, doing a PhD at UAB. The research subject is about Input/Output (I/O) system in High Performance Computing (HPC). She works in the traditional cluster platform and cloud platform. She has published in The Journal of Computer Science and Technology (JCS&T) an article where it has described the methodology to select a configuration in cloud platform.



**Poster “Methodology to select a I/O configuration (hardware resources and stack software) in cloud platform”**

Currently, there is an increasing interest about the cloud platform by the High Performance Computing (HPC) community, and the Parallel Input/Output (I/O) for High Performance Systems is not an exception. However, in cloud platform increase the number of parameters that user can select to I/O system. For this reason, we propose a methodology to help the user to select a configuration.

**Miguel Zavala**

**Barcelona Supercomputing Center**

He studied a B.S., Engineering Physics in the Universidad Autónoma de Yucatán (México), a MSc in Physics in the Universidad Nacional Autónoma de México and currently he is developing is PhD in the Universitat Politècnica de Catalunya BarcelonaTech.

**Poster “A Multiphysics implementation for conjugate heat transfer and compressible-low mach coupling”**

A parallel procedure for modelling multiphysics problems using the Parallel Location and Exchange library is presented. The methodology is applied to applications of industrial relevance: conjugate heat-transfer and coupling of low Mach number with compressible flow.

**Oriol Tintó**

**Universitat Autònoma de Barcelona**

He was born in Girona at late 80's and grow up in a small town in Catalonia's countryside. He studied physics at the Universitat Autònoma de Barcelona. When he finished his bachelor, his interest in learning how to apply his knowledge in maths and physics into a wide range of fields lead him to study the Master in Modelling for Science and Engineering. There was where he got in touch with High Performance Computing and the potential that it has in almost every field of science fascinated him. After this, he worked at the Institut Català de Ciències del Clima in Barcelona by studding HPC aspects of the ocean model used for Climate Studies. From the interesting work that we started during my stay in IC3 emerged the possibility to start a PhD. My interests are science and HPC, but also sports, music, technology and many others.

**Poster “Understanding Scientific Application's Performance”**

Improve Earth System model's scalability in High Performance Computing infrastructure is crucial to achieve efficiency in upcoming supercomputers. We study how to increase the scalability of the widely used Oceanic Model of the Nucleus for European Modelling of the Ocean (NEMO).





# EXTENDED ABSTRACTS

# Methodology to predict scalability of parallel applications

Claudia Rosas<sup>1</sup>, Judit Giménez<sup>1,2</sup> and Jesús Labarta<sup>1,2</sup>  
Barcelona Supercomputing Center<sup>1</sup> and Technical University of Catalonia<sup>2</sup>  
crosas@bsc.es

**Abstract-** In the road to exascale computing, the inference of expected performance of parallel applications results in a complex task. Performance analysts need to identify the behavior of the applications and to extrapolate it to nonexistent machines. In this work, we present a methodology based on collecting the essential knowledge about fundamental factors of parallel codes, and to analyze in detail the behavior of the application at low core counts on current platforms. The result is a guide to generate the model that best predicts performance at very large scale. Obtained results from executions at low core counts showed expected parallel efficiencies with a low relative error.

**Keywords:** parallel efficiency, analysis and prediction, exascale computing.

## LIST OF PUBLICATIONS

- Rosas, C.; Giménez, J. and Labarta, J. "Scalability prediction from fundamental performance factors." *Supercomputing Frontiers and Innovations. An Intl. Journal.* vol. 1. Num 2. pp. 4-19, Oct 2014.

## I. INTRODUCTION

Inferring the scaling capacity of parallel codes is getting harder than ever [1]; applications need to be modeled for restricted or nonexistent systems.

Our goal is to use *efficiency principles* of parallelism in the code to infer its potential performance. We consider as fundamental factors, components that represent essential features of the program and their evolution may be related to primary models of parallelism such as Amdahl's Law.

Our method relies on the detailed preprocessing of available traces to determine appropriate sections and clean noise. We follow this approach because: (i) having few points to fit, the blind use of functions with many parameters would lead to undetermined systems with many possible solutions; and (ii) low core count runs provide enough information on the fundamental behavior of the code.

The main contributions of this work are:

- The capability to identify interesting regions within the execution of parallel applications;
- The collection of fundamental factors, such as: Load Balance, Serialization and Transfer, to infer their evolution in the application;
- A general model to extrapolate parallel efficiency based on Amdahl's Law.

## II. METHODOLOGY

### A. Identify Structure

Our method begins with a trace, obtained from executing an instrumented parallel code on a target machine. We visualize the trace to generate clean cuts of the representative phases from the temporal structure of the execution. A main phase suggests regions of computation and/or communications that

may show different behaviors or that are independent between them.

### B. Phase Performance Analysis

We detail the performance of the application based on observations from collected measurements. Data measured from the trace of each phase is put together into a multiplicative analytical model, as shown in (1).

$$\eta_{||} = LB * Ser * Trf. \quad (1)$$

The model decomposes the parallel efficiency metric ( $\eta_{||}$ ) as a product of factors with normalized values between 0 (very bad) and 1 (perfect) [2]. The factors correspond to fundamental components of parallel coded and are load balance ( $LB$ ), serialization ( $Ser$ ) and transfer ( $Trf$ )<sup>1</sup>.

### C. Scalability Prediction

By independently extrapolating the individual components of this model we can observe how relevant they are to the overall performance of the parallel code. The basic default model is the Amdahl's law formulation. This is a first approach to describe the effect of non-parallel regions, where inefficiencies are caused by an activity that cannot be executed concurrently. Other possible patterns of concurrency correspond to pipelined computations or suggest a constant value when there are no changes in efficiency when scaling.

## III. EXPERIMENTAL EVALUATION

We use three applications from the CORAL suite: HACC [3], Nekbone [4] and AMG2013 [5]; and the CFD application AVBP [6]. CORAL applications were executed in MareNostrum III, and AVBP was executed in Juropa. The machines operated in normal production using fully populated nodes. Executions are summarized in Tab. I.

TABLE I  
APPLICATIONS USED FOR THE EXPERIMENTS

Strong Scaling	Ranks	Weak Scaling	Ranks
HACC	16, 32, 64, 128, 256, 512, 1024, 2048, 4096	AMG2013	32, 64, 96, 128, 192, 256, 384
Nekbone	2, 4, 8, 16, 32, 64, 128, 256, 512	AVBP	16, 32, 64, 96, 128, 192, 256, 520, 768, 1024, 1040, 1280, 1536

### A. Identify Structure

From a manual process, we obtain clean cuts of interesting phases within the execution; for example, one iteration of HACC shows a large computationally intensive phase of around 250 sec (left side of Fig.1), with low communications

<sup>1</sup>For more details, please refer to [3].

and some imbalances at the end. It also has a small communication phase (4 sec, at the right of Fig. 1).

### B. Phase Performance Analysis

With the automatic framework, we extract the main performance metrics for each phase. Because of the space, we show here only the results for HACC. In this application, the evolution of the three factors for the computational phase (left side in Fig. 2) describes a highly efficient region with almost none unbalance or contention, what suggest that it scales well. In the communication phase, it can be seen how Transfer and Serialization are dominant factors in the overall performance lost.

### C. Scalability Prediction

- **Model fitting:** Amdahl's Law reflects the presence of contention when increasing parallelism. In consequence, we fit the fundamental factors of most of the applications following the Amdahl model by default. Inside phase 2 of HACC (Fig. 3 right), processors seem to perform computations in stages we adapted the fitting model to a pipeline behavior.

- **Validation of results:** Our method extrapolates expected efficiencies for a specific machine from traces obtained from it. The error of prediction for each factor is summarized in Tab. II, it shows how optimistic (+) or pessimistic is our prediction (-). We used executions of HACC from 16 to 256 ranks to predict the efficiency for the next values, and compared the real measurements with the expected efficiencies.

- **Projection for large core counts:** our framework extrapolates the expected total parallel efficiency for up to  $10^6$  cores. Phase 1 of HACC shows in Fig. 4 (left) a constant behavior for serialization and transfer, and a soft degradation of load balance. For phase 2, we expect that at 2k cores the parallel efficiency is below 0.4, thus suggesting that main problem will be communication contention.

machines. Scalability projections showed initial insights of expected parallel efficiency and obtained results seem promising to provide our methodology as a tool to infer the scalability of parallel codes.

### ACKNOWLEDGMENT

This work has been done thanks to the Intel-BSC Exascale Lab and the DEEP Project.

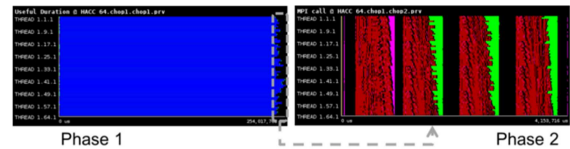


Fig. 1. Phases within one iteration of HACC code.

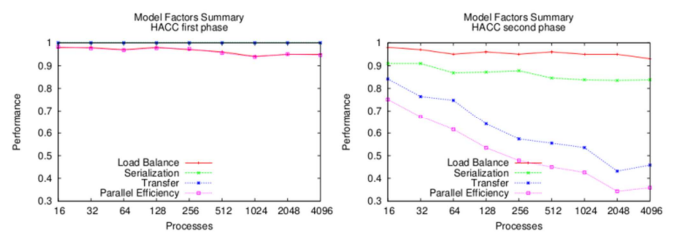


Fig. 2. Fundamental factors phase 1 and 2 for HACC

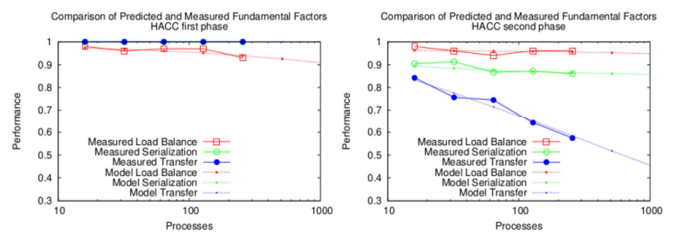


Fig. 3. Fitting phase 1 and 2 for HACC

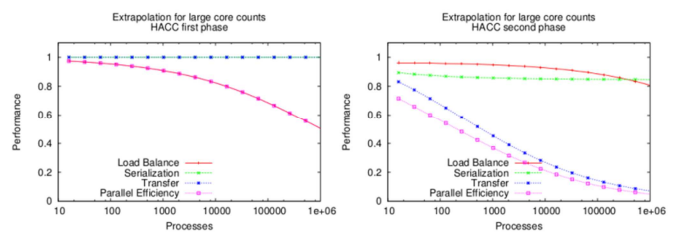


Fig. 4. Extrapolation phase 1 and 2 for HACC

TABLE II

PREDICTED EFFICIENCY AND RELATIVE ERROR FOR LARGER CORE COUNTS HACC (PHASE 2), EXTRAPOLATED FROM RUNS USING 16 TO 256 CORES

Ranks	Load Balance	Serialization	Transfer	Parallel Efficiency
512	0.952 (-0.82%)	0.860 (+1.81%)	0.517 (-6.78%)	0.424 (-5.61%)
1024	0.948 (-0.17%)	0.857 (+2.34%)	0.452 (-15.47%)	0.368 (-13.64%)
2048	0.943 (-0.68%)	0.855 (+2.45%)	0.391 (-9.39%)	0.315 (-7.80%)
4096	0.937 (+0.82%)	0.853 (+1.83%)	0.333 (-27.19%)	0.267 (-25.25%)

### IV. CONCLUSIONS

In this paper, we presented a methodology to collect primary components of current parallel codes at low core counts and infer their expected behavior when scaled to larger core counts. We evaluated the method with 3 applications from the CORAL suite and a CFD application in two different

### REFERENCES

- [1] A. Geist and R. Lucas, «Major Computer Science Challenges At Exascale,» *Int. J. of High Perform. Comput.*, vol. 23, n° 4, pp. 427-436, 2009.
- [2] M. Casas, R. M. Badia and J. Labarta, «Automatic Analysis of Speedup of MPI Applications,» de *Proc. 22nd Intl. Conf. on Supercomputing*, 2008.
- [3] S. Habib, V. Morozov, H. Finkel, A. Pope, K. Heitmann, K. Kumara, T. Peterka, J. Insley, D. Daniel, P. Fasel, N. Frontiere and Z. Lukic, «The Universe at Extreme Scale: Multi-petaflop Sky Simulation on the BG/Q,» de *Proc. Intl. Conf. on High Performance Computing, Networking,*

*Storage an Analysis.*, 2012.

- [4] Center for Exascale Simulation of Advanced Reactors. Reactors, «Proxy-Apps for Thermal Hydraulics.» [En línea]. Available: [https://cesar.mcs.anl.gov/content/software/thermal\\_hydraulics](https://cesar.mcs.anl.gov/content/software/thermal_hydraulics). [Last access: May 2014].
- [5] V. E. Henson and U. M. Yang, «BoomerAMG: A parallel algebraic multigrid solver and pre-conditioner,» *Appl. Numer. Math.*, vol. 41, n° 1, pp. 155-177, 2002.
- [6] N. Gourdain, L. Gicquel, M. Montagnac, O. Vermorel, M. Gazaix, G. Stafflebach, M. Garcia, J. Boussuge and T. Poinso, «High performance parallel computing of flows in complex geometries: I. Methods,» *Comp. Sci. Disc.*, vol. 2, n° 1, p. 015003, 2009.
- [7] C. Rosas, J. Giménez and J. Labarta, «Scalability prediction from fundamental performance factors,» *Supercomputing Frontiers and Innovations. An Intl. J.*, vol. 1, n° 2, pp. 4-19, 2014.

# Measurements or computer simulations: to whom do we turn for atmospheric composition estimates?

Enza Di Tomaso<sup>1</sup>, Nick Schutgens<sup>2</sup>, Oriol Jorba<sup>1</sup>

<sup>1</sup>Earth Sciences Department, Barcelona Supercomputing Center, Spain

<sup>2</sup>Atmospheric, Oceanic and Planetary Physics, University of Oxford, United Kingdom  
*enza.ditomaso@bsc.es*

**Abstract-** *Computer simulations are able to provide a description of the atmosphere at unprecedented fine spatial and temporal scales, thanks to the most recent advances made in model developments and to the increasingly available computational resources. However, due to our limited understanding of physical processes in the atmosphere, their accuracy is not at the level of some atmospheric measurements, which, on the other hand, might suffer of a lower temporal resolution and sparser coverage and, by their nature, lack predictive skills. Given that limitations and uncertainties exist both in models and measurements, whom do we trust for the correct estimation of atmospheric variables? The Bayesian paradigm provides a probabilistic approach for optimally combining the two pieces of information (model simulations and observations) weighted by their respective uncertainties. We report here on the developments in this field at the Earth Sciences Department of the Barcelona Supercomputing Center (BSC) for retrievals of aerosol optical depth from two satellite products (NRL MODIS and MODIS Deep Blue) and model simulations related to mineral dust. We use a data assimilation technique particularly suited to high-performance computing applications, and show that it improves significantly the characterisation of the atmospheric dust load.*

## I. INTRODUCTION

Computer simulations of the atmosphere aim at predicting the state of atmospheric variables by numerically integrating equations of atmospheric motion and of physical processes starting from given initial conditions. Accurate initial conditions are particularly important in numerical weather prediction due to the chaotic nature of the problem. The skill of weather prediction has been enhanced substantially in the last decades thanks to the usage of more exact initial conditions (Rabier 2005). This has been obtained through advances in the observing system, in particular of satellite-based observations, and developments of data assimilation techniques which estimate optimal initial conditions, namely analyses, from meteorological observations and a prior first-guess, or background. A short-term forecast valid at the observation time is generally used as background information. Data assimilation has been key also in the monitoring of current and past states of the atmosphere. More recently also the atmospheric composition community has turned to data assimilation for a better characterisation and prediction of atmospheric constituents that are present in the atmosphere in reduced concentrations compared to the main atmospheric components, such as liquid and solid particles (also known as aerosols) other than cloud and precipitation particles and trace gases, as ozone and greenhouse gases for example (Bocquet 2014). Despite their relatively low global concentrations, both aerosols and trace gases play an important role in the Earth energy balance and have a relevant impact on climate, on the ecosystem and on health. Though their dynamic is mainly

driven by forcings such as emissions, recent studies showed that the usage of observations through data assimilation has improved significantly the accuracy of short-term forecast and the global monitoring of both aerosols and trace gases (Benedetti et al. 2009, Elbern and Schmidt 2001). The Earth Sciences Department of the BSC (ES-BSC) is implementing a computationally-efficient gas-aerosol module able to predict atmospheric composition at different spatial and temporal scales within a state-of-art meteorological model NMMB (Janjic and Gall, 2012). The new modelling system is known as the NMMB/BSC Chemical Transport Model (NMMB/BSC-CTM, Pérez et al. 2011, Jorba et al. 2012, Spada et al. 2013). We report here on the enhancement of this chemical transport model with the assimilation of aerosol satellite products following what is the state-of-art in the field, with a focus on mineral dust only, a prominent type of aerosol.

## II. METHODOLOGY

We have coupled the in-house developed chemical transport model NMMB/BSC-CTM with an ensemble-based data assimilation scheme called Local Ensemble Transform Kalman Filter (LETKF, Hunt et al. 2007, Miyoshi et al. 2007, Schutgens et al. 2010). The LETKF scheme requires an ensemble of model runs whose spread around the mean represents the model background uncertainty. As uncertainty in the emission term is particularly high for mineral dust (Huneeus et al. 2011), we here choose to create an ensemble according to perturbations in the model's dust emission. We have perturbed two emission parameters: the calibration factor for the vertical flux of dust and the threshold on the friction velocity used for the production of the horizontal flux of dust. The LETKF is a technique particularly suited to high-performance computing applications: the execution of an ensemble of independent model integrations is embarrassingly parallel, while the analysis calculations are explicitly parallelised in the code (the analysis can be performed locally in the model grid). Data assimilation experiments have been run using a set of 12 model runs on a global domain and at the resolution of 2.8 by 2 degree and 40 vertical layers. The assimilated observations are Aerosol Optical Depth (AOD) retrievals from the U.S. Naval Research Laboratory (NRL) MODIS Level 3 product (Zhang et al. 2006, Hyer et al. 2010, Shi et al. 2011) and the MODIS Deep Blue Level 3 product (Hsu et al. 2013, Sayer et al. 2013). Observations have been satellite products and undergone quality control during the

assimilation process according to their departures from the model background.

## V. RESULTS

As a sanity check we have assessed the agreement between the analysis and the assimilated observations after a spin-up period of one month. Figure 1 shows for May 2007 the scatter plots of the counts of assimilated NRL MODIS observations versus model first-guess AOD values (with correlation coefficient 0.67) and versus analysis AOD values (with correlation coefficient 0.81). A smaller scatter and a higher correlation coefficient for the analysis indicate a good performance of our data assimilation system.

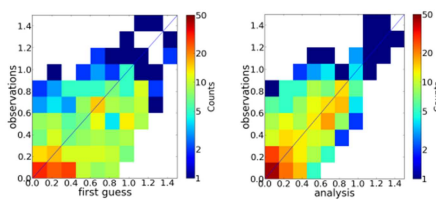


Fig. 1. Scatter plot of the counts of assimilated observations and first guess (left plot) and analysis (right plot) for May 2007 for the data assimilation experiment assimilating quality controlled NRL MODIS values.

Here we visually compare dust analyses with independent ground-station (Figure 2) and satellite (Figure 3 and 4) observations. Figure 2 shows the time series of first-guess AOD values (in blue) and analysis AOD values (in red) when NRL and Deep Blue MODIS observations are assimilated (circles), and of independent ground-station measurements (red triangles) from the AERONET network (Holben et al. 1998), for a station affected by short-range dust transport (Dakar, Senegal). The analysis is in general in better agreement than the first-guess with observations that have not been assimilated and that are quite accurate.

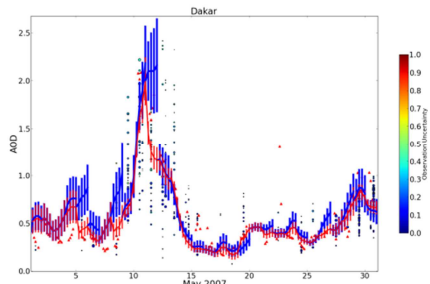


Fig. 2. Time series of AOD values for May 2007 in Dakar for the mean ensemble first-guess (blue) and for the mean ensemble analysis (red) with the assimilation of NRL MODIS and MODIS Deep Blue AOD and for the AERONET independent product (red triangles). The error bars represent the ensemble spread.

Figure 3 shows for May (top) and June (bottom) 2007 mean model background at 00UTC in a free run experiment without data assimilation (left), mean analysis with the assimilation of

NRL MODIS observations (centre) and with the assimilation of NRL MODIS and MODIS Deep Blue observations (right). The experiments with the assimilation of satellite observations show a better representation of dust concentration, more in agreement with independent satellite dust retrievals from the CALIPSO sensor (Figure 4) than the experiment without data assimilation. The assimilation of MODIS Deep Blue product in particular has a relevant impact near dust source regions.

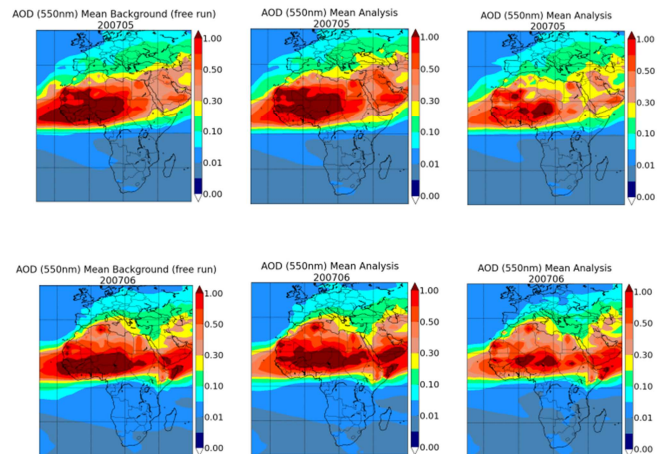


Fig. 3. Mean AOD values for the month of May (top) and June (bottom) 2007 for the free run experiment with no data assimilation (left), for the data assimilation experiment assimilating NRL MODIS AOD (centre) and NRL MODIS plus Deep Blue MODIS AOD (right).

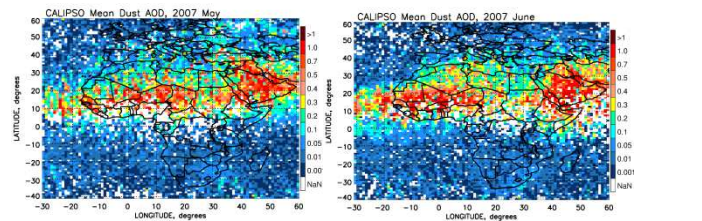


Fig. 4. Mean dust AOD values for the month of May (left) and June (right) 2007 for independent satellite dust retrievals from CALIPSO sensor.

## VI. CONCLUSIONS

Before this work, the NMMB/BSC-CTM did not have an aerosol data assimilation capability and dust, for example, was produced uniquely from model estimated surface emission fluxes. As emissions are recognized as a major factor limiting the accuracy of dust modelling, remote sensing observations from satellites have been used to improve the description of the atmospheric dust load. An ensemble-based technique (namely the Local Ensemble Transform Kalman Filter - LETKF) has been utilised to statistically combine model background and satellite retrievals. Data assimilation experiments have been evaluated against independent observations from ground stations and satellite sensors, and showed that aerosol optical depth products retrieved from satellite measurements can help to better characterise atmospheric dust. We plan to implement this data assimilation capability in the ES-BSC dust and aerosol forecast system.

## ACKNOWLEDGMENT

This work was funded by the SEV-2011-00067 grant of the Severo Ochoa Program and the CGL-2013-46736-R grant awarded by the Spanish Government. The authors thank E. Marinou and V. Amiridis for their CALIPSO retrievals, all the Principal Investigators and their staff for maintaining the AERONET sites, NRL-University of North Dakota, the MODIS and OMI mission scientists and associated NASA personnel for the production of the AOD, AAI and AE data used in this investigation. All the simulations have been run on the MareNostrum supercomputer hosted by BSC.

# HPC and edge elements for geophysical electromagnetic problems: an overview

Octavio Castillo, Josep de la Puente, Vladimir Puzyrev, José María Cela  
CASE Department, Barcelona Supercomputing Center  
octavio.castillo@bsc.es

**Abstract-** In Finite Element Methods for solving electromagnetic problems, the use of Nédélec Elements has become very popular. In fact, Nédélec Elements are often said to be a cure to many difficulties that are encountered, particularly eliminating spurious solutions, and are claimed to yield accurate results. In this paper, we present our first steps in developing a Nédélec Elements code for simulation of geophysical electromagnetic problems and first ideas on how implement the key issues of Edge Elements in an efficient way on HPC platforms.

## I. INTRODUCTION

Electromagnetic Methods (EM) are an established tool in geophysics, finding application in many areas such as hydrocarbon and mineral exploration, reservoir monitoring, CO2 storage characterization, geothermal reservoir imaging and many others. The marine controlled-source electromagnetic method (CSEM) has become an important technique for reducing ambiguities in data interpretation in the offshore environment and a commonplace in the industry. In the traditional configuration, the sub-seafloor structure is explored by emitting low-frequency signals from a high-powered electric dipole source towed close to the seafloor. By studying the received signal, the subsurface structures could be detected at scales of a few tens of meters to depths of several kilometers.

On the other hand, in the Finite Element Method for solving electromagnetic field problems, the use of Edge-based elements (Nédélec elements) has become very popular. In fact, Nédélec elements are often said to be a cure to many difficulties that are encountered, particularly eliminating spurious solutions, and are claimed to yield accurate results. In Edge elements, the simulated quantity is interpolated by vector shape functions that provide several characteristics important for the representation of the electric field.

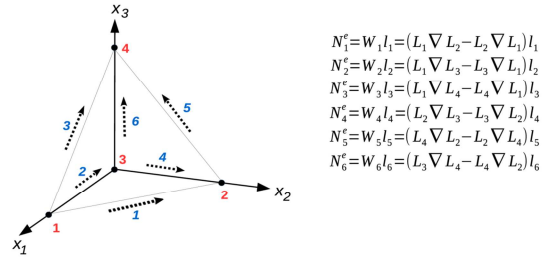
Based on previous ideas and considering the societal value of exploration geophysics, since this process is essential to among others, we present the key issues of edge element to simulate geophysical electromagnetic problems, namely, we present our first steps in developing a Nédélec Elements code for simulation of geophysical electromagnetic problems and first ideas on how implement the key issues of Edge Elements in an efficient way. Strong emphasis is placed on three aspects not easily found in the literature: vector shape functions, orientation of geometric entities and the technique to avoid the source singularities in CSEM simulations. Finally, we make some concluding remarks and show the performance of the code in terms of efficiency and accuracy.

## II. NÉDÉLEC ELEMENTS FORMULATION

The Nédélec Finite Element Method uses vector basis functions defined on the edges of the corresponding elements. Similar to the conventional node-based finite element method, the modeling domain can be discretized using rectangular, tetrahedron, hexahedron or other complex elements [1].

Fig. 1 (left) is an illustration of the tetrahedron grid that we used with node (red numbers) and edge indexing (blue numbers). Following the work of [1], the tangential component of the electric field is assigned to the center of each edge. The edge basis functions defined by [1] are divergence free but not curl free. The vector basis functions are also continuous at the element boundaries. As a result, the divergence free condition of the electric field in the source free region and the continuity conditions are imposed directly in the Nédélec Finite Element formulation. Thus, the element vector basis functions for the field associated with the edges are given by Fig. 1 (right) where  $L_i$  are defined in [1] and  $l_i$  represents the length of edge  $i$ .

Fig. 1. Nédélec element description with nodes and edges (left). Vector basis functions of linear edge elements right).



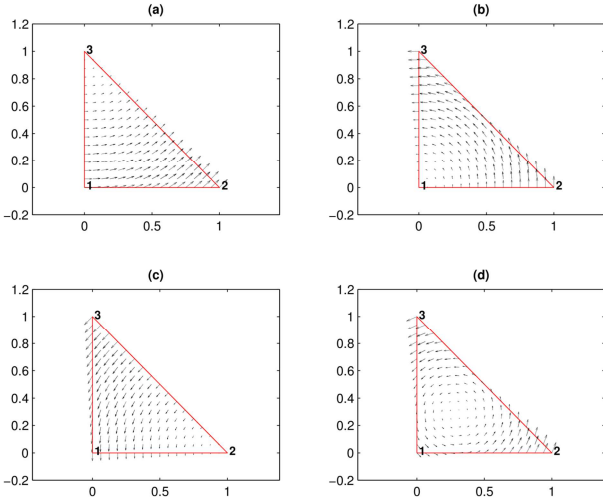
In order to guarantee global continuity of tangential components, special care has to be taken with regard to the edge direction. In short, the edge direction strategy works as follows. A global index is assigned to each node of the mesh. If an edge connects two nodes  $n_i$  and  $n_j$ , we define the direction of the edge as going from node  $n_i$  to node  $n_j$  if  $i < j$ . This gives a unique orientation of each edge. The same philosophy is used locally to determine the directions of the local edges on each element.



### III. SUMMARY OF RESULTS

The following images are the result of our first steps in developing a Nédélec Elements code for simulation of geophysical electromagnetic problems.

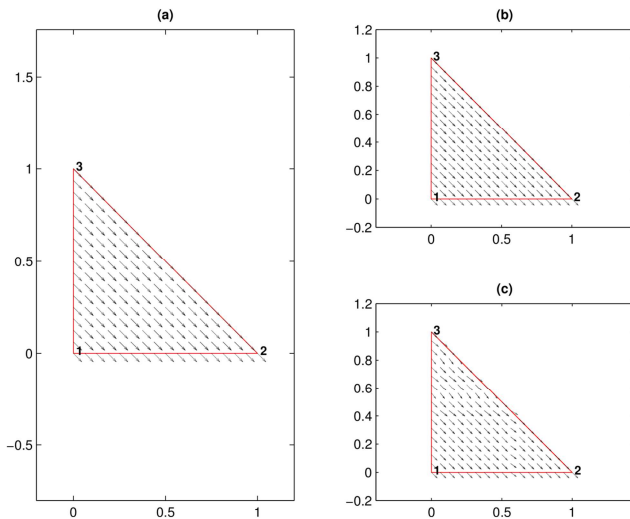
Fig. 2 Vectorial basis functions. Subplot (a) describes the shape function along edge 1 – 2, Subplot (b) describes the shape function along edge 2 – 3, Subplot (c) describes the shape function along edge 3 – 1 and subplot (c)



In geophysical simulations with CSEM, the primary field is calculated analytically for a background layered-earth model. The secondary field is discretized using edge-elements. The primary field computation stage is a classic example of embarrassingly parallel workload because no exists dependency or communication between those parallel tasks.

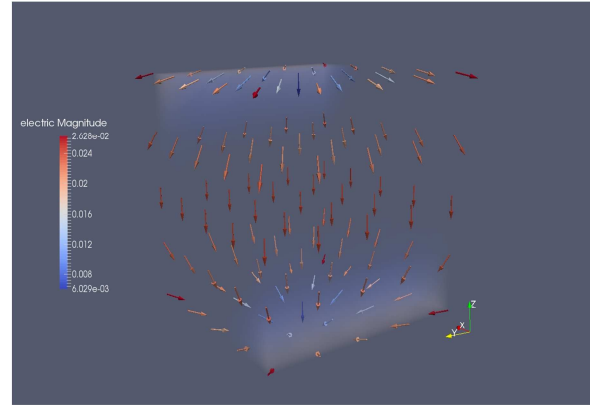
On the other hand, numerical modeling of geophysical electromagnetic problems impose a significant

Fig. 3 Ability of vectorial basis functions to interpolate a given field. (a) constant field to test, (b) field within the element, (c) error of approximation (maximum error=0.5551 - 0.3311 e-15)



computational burden. In solving large-scale models, single processor-based methods are limited and often incapable of managing the required large memory and computational time requirements. Therefore, we are exploring parallel environments such as MPI and OpenMP in order to improve the performance of these simulations.

Fig. 4 Primary field of a z-directed dipole in a homogeneous domain with 384 elements.



# T	# E	Assembly time	Solver time	h	$L^2$	$O_L$
128	208	7.66e-03	6.76e-03	3.3e-01	3.969e-01	---
512	800	6.75e-03	7.72e-03	2.0e-01	9.831e-02	2.016
2,048	3136	2.22e-02	4.47e-02	1.1e-01	2.430e-02	2.017
8,192	12,416	1.82e-01	2.08e-01	5.8e-02	6.001e-03	1.999
32,768	49,408	4.27e-01	6.68e-01	3.0e-02	1.501e-03	2.004
131,072	197,120	1.51e00	2.27e00	1.5e-02	3.741e-04	2.002
524,288	787,456	5.83e00	6.71e00	7.8e-03	9.340e-05	2.002
2,097,152	3,147,776	21.32e00	24.7e00	3.9e-03	2.333e-05	2.001

Table 1: Summary of results for the 2D case. Number of elements (T), number of edges (E), assembly time (seconds), solver time (seconds), mesh spacing (h),  $L^2$  error and convergence order.

### ACKNOWLEDGMENT

Authors gratefully acknowledge the support from the Mexican National Council for Science and Technology (CONACYT). This research was made in the facilities of the CASE Department-Barcelona Supercomputing Center.

### REFERENCES

- [1] Hongzhu, C. (2014). 3D controlled-source electromagnetic modeling in anisotropic medium using edge-based finite element method. *Computer & Geosciences* 73, 164-176.
- [2] Jin, J. (2002). *The finite element method in electromagnetics*. John Wiley & Sons.
- [3] Puzyrev, V., Koldan, J., de la Puente, J., Houzeaux, G., Vázquez, M., & Cela, J. M. (2013). A parallel finite-element method for three-dimensional controlled-source electromagnetic forward modelling. *Geophysical Journal International*, ggt027.
- [4] Weiss, C. J., & Newman, G. A. (2002). Electromagnetic induction in a fully 3-D anisotropic earth. *Geophysics*, 67(4), 1104-1114.

# From Imaging to Simulation: A framework applied to simulate the blood flow in the carotids

R. Arís<sup>1</sup>, V. Mani<sup>2</sup>, H. Dyvorne<sup>2</sup>, M. Jowkar<sup>1</sup>, A. Santiago<sup>1</sup>, G. Houzeaux<sup>1</sup>, Z. Fayad<sup>2</sup>, M. Vázquez<sup>1</sup>

<sup>1</sup>Barcelona Supercomputing Center (BSC), Spain

<sup>2</sup>Translational and Molecular Imaging Institute (TMII), The Mount Sinai Hospital, USA

ruth.aris@bsc.es

**Abstract-***In this work we present a methodology to extract information from medical imaging and use it for hemodynamical simulation in arteries. Based on in-vivo magnetic resonance images (MRI), the velocity of the blood flow has been measured at different positions and times. Also, the anatomy of the vessel has been converted into a volume mesh suitable for numerical modeling. This data has been used to solve computationally the dynamics of the fluid inside the artery in healthy and pathologic cases. As an application, we have developed a computational model within the carotids. The next step in the pipeline will be to extend the simulation to fluid-structure interaction (FSI) to find the parameters in an atherosclerotic plaque that could lead to rupture.*

## INTRODUCTION

The complex structure of an atherosclerotic plaque affects the dynamics of the fluid inside the vessel. Usually, it is composed of different materials such as a lipid pool, calcification or hemorrhage. In the last decades, magnetic resonance imaging (MRI) has emerged as an imaging modality that enables accurate quantification of the main components of the plaque. Size, shape and plaque constituents can be identified noninvasively [1, 2, 3]. Other MRI techniques like phase-contrast have also been developed to acquire multidirectional blood velocity data in-vivo. 2D and 3D phase-contrast MRI has been used for the characterization of normal and pathological velocity profiles in those areas [4].

To study the dynamics of the flow in the cardiovascular system, the simulation tool has been proved to be a powerful complement. It has helped to investigate the problem through simple and complex geometry models, most of them based on in-vivo measurements [5, 6]. Particularly, image-based computational modeling can provide valuable information of the hemodynamical conditions and can assist the prediction of possible outcomes.

The present work aims to bring imaging and simulation closer in a robust framework. In this work, 3D MRI data of human carotids is acquired at Mount Sinai Hospital of New York. For each case, the geometries of the lumen and the vessel wall are segmented from images and converted into a computational mesh ready to run the simulations. Velocity measurements are also acquired and included as initial conditions in the computational model. As a first approach, only fluid simulations are performed. Once the framework will be validated, it will be extended to FSI to obtain an accurate simulation of the phenomenon occurring between the wall and the blood flow, to study the effect of the atherosclerotic plaque in the carotids.

## METHODOLOGY

The problem is divided in three different parts:

**1. MRI acquisition.** Following the study protocols, patients are imaged on a Siemens MR/PET (3T) mMR. The imaging protocol is defined as follows: (i) the bilateral carotid arteries extending 3 cm below and above the carotid bifurcations are imaged using an 8-channel carotid coil. (ii) To delineate the vessel lumen, localization with gradient echo sequences, time-of-flight (TOF) images are acquired. (iii) To obtain black blood images and measure carotid stenosis, 3D SPACE with T2 weighting is acquired in free breathing conditions, untriggered with fat suppression.

As a result, 2D images in the coronal planes are obtained. Each slice has an in-plane resolution of 0.7x0.7 mm<sup>2</sup>, with a slice thickness of 0.7 mm. The flow in a vessel section is measured with a 2D, single-slice, cardiac-gated phase contrast sequence with through-plane velocity encoding (Venc=100cm/s). The acquisition plane is placed perpendicular to the carotid artery, 2 cm before the bifurcation, using a 3D time of flight acquisition to localize the vessels. The slice has the same resolution as the previous imaging protocol.

**2. Image analysis and mesh generation.** The geometry of the carotids is extracted from the MRI raw data performing a manual segmentation in ITK-SNAP [7]. The vessel wall is reconstructed computationally using the information in three different planes in space. In case of plaque, the corresponding materials are also identified. Once the segmentation is finished, the software generates the surface mesh of the lumen and the vessel wall in the STL (STereoLithography) file format. Finally, the mesh generator Iris, developed and supported at Barcelona Supercomputing Center (BSC), creates the unstructured finite element meshes for both fluid and solid geometries (Fig. 1).

**3. Simulation.** The appropriate framework for the simulation of fluid-structure interaction (FSI) in blood vessels is the arbitrary Lagrangian-Eulerian (ALE) description of continuous media. In this description, the fluid and solid domains are allowed to move to follow the distensible vessels and deforming fluid domain [8]. In this preliminary work, only

fluid simulations are presented. In this case, the fluid is simulated as laminar, viscous, incompressible, Newtonian flow in a fixed domain solving the Navier-Stokes equations and choosing suitable initial and boundary conditions [5]. The 3D computational model is performed in Alya, the multi-physics and parallel code developed at BSC [9, 10, 11, 12, 13] and solved in parallel in the Spanish MareNostrum supercomputer.



Fig. 1. Computational mesh created for a healthy case. Left: meshes of the lumen (grey) and the vessel wall (red), composed of tetrahedral elements. Right: view of the bifurcation in the carotids.

Fig. 2 shows a pathological carotid artery where different components have been identified. The result of the simulation is presented in Fig. 3 and it is compared to the healthy carotid artery of the same patient. In both cases, MRI data has been used to provide the geometry and boundary conditions (i.e. inflow, and outflow) for a computational fluid dynamics calculation.



Fig. 2. Surface mesh of a pathologic case in the carotids bifurcation. Lumen (blue) and wall (brown) are represented in colors. The atherosclerotic plaque is drawn in yellow.

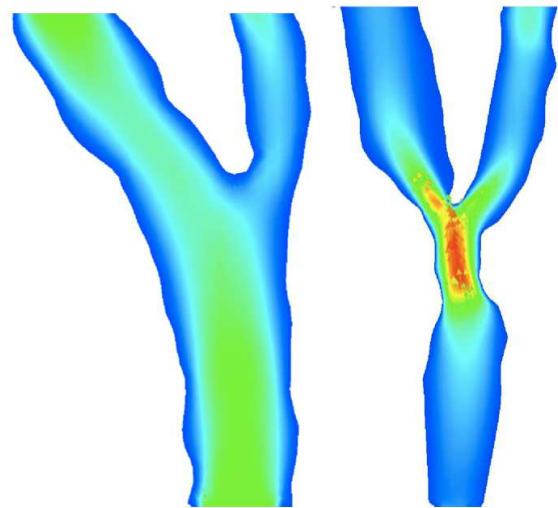


Fig. 3. The image-based computational framework is used to solve the hemodynamics of the problem. We compare the dynamics of the fluid in healthy and pathological arteries. Colors represent velocities. Right: healthy case. The velocity profile is normal, achieving a maximum of 70 cm/s at the systolic peak (PSV). Left: pathologic case. Maximum velocities are found in the bifurcation. Values are high in this case, achieving 200 cm/s at PSV.

## VII. FUTURE WORK

Nowadays, there's a lack of realistic geometries used in numerical simulation. The pipeline presented here will constitute an efficient bridge to connect image and simulation. Given the increasing resolution and accuracy of information obtained from clinical imaging, the simulation tool is the complement that can provide an efficient way of research, prior to clinical trials which are expensive and can be a risk to the patients. Simulations can be performed on a patient's data and the results can be delivered to the doctor in an ad-hoc visualization manner (virtual lab) to get a deeper insight of the problem.

To identify the wall shear stress in the atherosclerotic material, flow will be simulated on a moving mesh using an (ALE) strategy. The artery wall and the plaque components will be modeled hyperelastic, isotropic, incompressible and homogeneous. The coupled FSI model will be solved in parallel in MareNostrum.

## ACKNOWLEDGMENT

This work has been written with the support of the postdoctoral fellowship grant under the Cardio-Image Program, awarded by the Spanish Research Center CNIC.

## REFERENCES

- [1] Z. A. Fayad, J. T. Fallon, M. Shinnar, S. Wehrli, H. M. Dansky, M. Poon,

J. J. Badimon, S. A. Charlton, E. A. Fisher, J. L. Breslow, and V. Fuster, "Noninvasive In vivo high-resolution magnetic resonance imaging of atherosclerotic lesions in genetically engineered mice," *Circulation*, vol. 98, no. 15, pp. 1541–1547, 1998.

[2] C. Yuan, W. S. Kerwin, M. S. Ferguson, N. Polissar, S. Zhang, J. Cai, and T. S. Hatsukami, "Contrast-enhanced high resolution MRI for atherosclerotic carotid artery tissue characterization," *J Magn Reson Imaging*, vol. 15, no. 1, pp. 62–67, 2002.

[3] D. Tang, C. Yang, J. Zheng, P. K. Woodard, G. A. Sicard, J. E. Saffitz, and C. Yuan, "3D MRI-based multicomponent FSI models for atherosclerotic plaques," *Ann Biomed Eng*, vol. 32, no. 7, pp. 947–960, 2004.

[4] M. D. Hope, D. D. Purcell, T. A. Hope, C. von Morze, D. B. Vigneron, M. T. Alley, and W. P. Dillon, "Complete intracranial arterial and venous blood flow evaluation with 4D flow MR imaging," *AJNR Am J Neuroradiol*, vol. 30, no. 2, pp. 362–366, 2009.

[5] C. A. Taylor, T. J. Hughes, and C. K. Zarins, "Finite element modeling of three-dimensional pulsatile flow in the abdominal aorta: relevance to atherosclerosis," *Ann Biomed Eng*, vol. 26, no. 6, pp. 975–987, 1998.

[6] C.A. Taylor and C.A. Figueroa, "Patient-specific modeling of cardiovascular mechanics," *Annu Rev Biomed Eng*, vol. 11, pp. 109–134, 2009.

[7] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[8] T. J. Hughes, W. K. Liu, and T. K. Zimmermann, "Lagrangian-eulerian finite element formulation for incompressible viscous flows," *Computer Methods in Applied Mechanics and Engineering*, vol. 29, no. 3, pp. 329 – 349, 1981.

[9] G. Houzeaux, M. Vázquez, R. Aubry, and J. Cela, "A massively parallel fractional step solver for incompressible flows," *Journal of Computational Physics*, vol. 228, 2009.

[10] G. Houzeaux, M. Vázquez, X. Sáez, and J. Cela, "Hybrid mpi-openmp performance in massively parallel computational fluid dynamics," *Parallel Computational Fluid Dynamics*, vol. 74, 2011.

[11] M. Vázquez, G. Houzeaux, S. Koric, A. Artigues, J. Aguado-Sierra, R. Arís, D. Mira, H. Calmet, H. O. F. Cucchietti, A. Taha, and J. Cela, "Alya: Towards exascale for engineering simulation codes," in *SIAM Journal on Scientific Computing*. Also as preprint in arXiv:1404.4881, 2014.

[12] M. Vázquez, R. Arís, J. Aguado-Sierra, G. Houzeaux, A. Santiago, M. López, P. Córdoba, M. Rivero, and J. Cajas, "Alya red ccm: Hpc-based cardiac computational modeling," *Selected Topics of Computational and Experimental Fluid Mechanics*, 2015.

[13] J. R. Cebal, M. Vázquez, D. M. Sforza, G. Houzeaux, S. Tateshima, E. Scrivano, C. Bleise, P. Lylyk, and C. M. Putman, "Analysis of hemodynamics and wall mechanics at sites of cerebral aneurysm rupture," *J Neurointerv Surg*, 2014.

# Optimizing a coarse-grained model for the recognition of protein-protein binding

Agustí Emperador, Modesto Orozco  
Institute for Research in Biomedicine  
agusti.emperador@irbbarcelona.org

**Abstract-** We are optimizing a force-field to be used with our coarse-grained protein model for the recognition of protein-protein binding. We have found that, apart from ranking correctly the ligand-receptor conformations generated in a protein-protein docking algorithm, our model is able to distinguish binding (experimental structure) from nonbinding (false positive) conformations for many complexes. This suggests us that the model could have a good performance in complete cross-docking, a method aimed to recognize the possible binding between any two proteins that are unknown to interact.

## VIII. INTRODUCTION

We have applied our coarse-grained protein model [1] to the simulation of protein complexes. The protein model uses a force field parametrized to reproduce correct association and dissociation rates of intrinsically disordered proteins [1]. Our aim in this work is to test the stability of experimental protein complexes during molecular dynamics simulations with our protein coarse-grained model and force-field, and to test if our force field is able to identify false positives. Stability of experimental complex structures together with dissociation of false positives after simulation would suggest that our model might be able to recognize which proteins associate with each other and which is their binding conformation [2].

## IX. METHODS

We have used the Discrete Molecular Dynamics (DMD) algorithm [3] for the simulation of the proteins complexes. In DMD the particles interact through stepwise potentials, therefore move with constant velocity until they reach the pairwise distance corresponding to a potential energy step, when the collision happens. At this point a transfer of momentum between the two particles occurs, changing the velocity of them. The velocity of the particles is updated by imposing conservation of linear momentum and energy at the collision, instead of integrating the equations of motion as made in standard molecular dynamics (MD). The simplicity of the calculations made in DMD makes it much faster than MD, with a negligible loss of accuracy respect to it. This together with the low number of particles in the simulation due to the use of implicit solvent and coarse-graining of the system make the simulations orders of magnitude faster than conventional explicit solvent MD simulations.

We have used the coarse-grained model of Marrink et al. [4] for the sidechains plus an atomistic representation of the amide atoms N,H,C,O, in order to be able to include explicitly

hydrogen bonds in the model. The number of beads of each sidechain depends on the aminoacid, mapping generally four heavy atoms in each bead.

We have reparametrized our force field [1] in order to improve the predictive performance in protein-protein complexes. Following the same idea as used in the optimization of scoring functions for protein-protein docking [5], we have modified the nonbonded part of the force-field changing the weight of the following terms: i) Van der Waals; ii) electrostatic; iii) implicit solvation

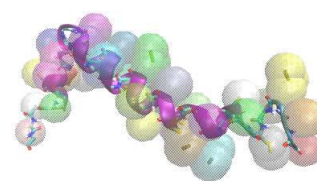


Fig. 1 Coarse-grained protein structure in our model. An atomistic representation is used for the amide atoms of the backbone, and a coarse-grained representation for the side chain beads (transparent spheres with a different color for each residue)

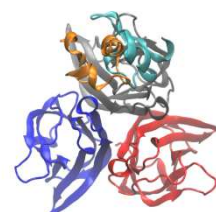


Fig.2 Movement of the ligand along the DMD simulation with our force field for the experimental structure of the complex of PDB code 1XU1. The ligand in the experimental structure is in gold color, and after the simulation in cyan color.

## X. RESULTS

We have made simulations on all the experimental complex structures of Weng's protein docking benchmark 4.0 [6]. In figure 2 is shown the movement of the ligand of an experimental complex structure after a simulation of 1 ns. The simulation of the best scored false positive for this complex generated by a standard docking algorithm [5] produces a fast dissociation of the ligand and the receptor.

In a previous work [7] we simulated this benchmark using an atomistic representation of the proteins, finding that all the

docking conformations kept stable during the simulation,[7] because we were using a standard molecular dynamics with implicit solvent force field not parametrized to give the correct association/dissociation rates in systems of interacting proteins.

Here we found that the majority of the experimental structures keep stable during the simulation, as shown in the figure 3. We found that only 20% of the complexes of the benchmark showed dissociation.

A. Emperador et al., "Efficient relaxation of protein-protein interfaces by discrete molecular dynamics simulations", *J. Chem. Theory and Comput.* **9** 1222-1229 (2012)

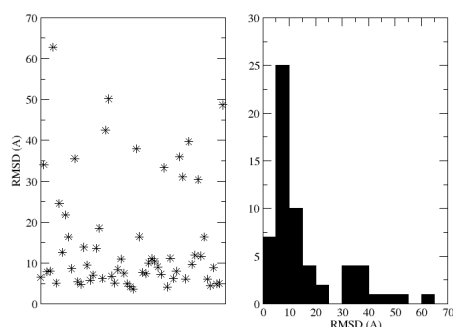


Fig. 3 RMSD respect to the experimental structure after the DMD simulation with our force-field for the complexes of the protein benchmark.

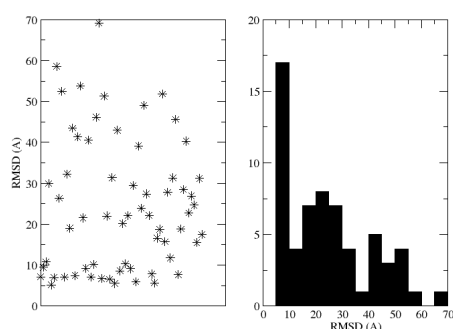


Fig. 4 RMSD respect to the initial structure of the best scored false positive generated by a docking algorithm for the complexes of the protein benchmark.

When we made the simulations of the best scored false positives, we found (figure 4) that in 60% of the complexes ligand and receptor dissociated. This suggests us that our approach is able to recognize stable receptor-ligand configurations and discard false positives for the majority of the complexes in the benchmark, opening the perspective to obtain a higher performance after further optimization of the force field.

#### REFERENCES

- [1] A. Emperador et al., "A coarse-grained model for the study of protein structure, dynamics and interactions", unpublished.
- [2] S. Sacquin-Mora et al., "Identification of protein interaction partners and protein-protein interaction sites", *J. Mol. Biol.* **382**, 1276-1289 (2008)
- [3] A. Emperador et al., "Protein feasibility from discrete molecular dynamics simulations using quasi-physical potentials", *Proteins* **78**, 83-94 (2010)
- [4] L. Monticelli et al., "The MARTINI coarse-grained force field: extension to proteins", *J. Chem. Theory and Comput.* **4**, 819-934 (2008)
- [5] J. Fernandez-Recio et al., *Protein Science* **11** 280-291 (2002)
- [6] H. Hwang, *Proteins* **78** 3111-3114 (2010)

# Study of Preconditioners based on Markov Chain Monte Carlo Methods

Oscar A. Esquivel-Flores  
CONACYT-BSC  
oscar.esquivel@bsc.es

***Abstract***-Nowadays, analysis and design of novel scalable methods and algorithms for fundamental linear algebra problems such as solving Systems of Linear Algebraic Equations with focus on large scale systems is a subject of study. This research focuses on the study of novel mathematical methods and scalable algorithms for computationally intensive problems such as Monte Carlo and Hybrid Methods and Algorithms.

## I. INTRODUCTION

Solving systems of linear algebraic equations (SLAE) or inverting a real matrix are well-known problems. Iterative or direct methods to solve these systems are costly approach. One option of reducing the effort of solving these systems is to apply preconditioners before using an iterative method. Standard deterministic preconditioners computed by the optimized parallel variant- Modified SParse Approximate Inverse Preconditioner (MSPAI) have been changed by a Monte Carlo preconditioner that relies on the use of Markov Chain Monte Carlo (MCMC) methods [1]. The study of this methods and their parallel implementation is subject of current extreme scale computing research.

Preconditioning refers to transform a complex problem into another whose solution can be tractable; preconditioner is the operator that is responsible for such transformation. A way to increase the robustness and the computational efficiency of iterative methods is based on preconditioning.

Several physical problems imply to solve SLAE as result of discretization of partial differential equations. Iterative solvers are often the method of choice due to their predictability and reliability when considering accuracy and speed. They are, however, prohibitive for large-scale problems as they can be very time consuming to compute. These methods are dependent on the size of the matrix and so the computational effort grows with the problem size [2]. On the other hand, Monte Carlo (MC) methods performing random sampling of a certain variable whose mathematical expectation is the desired solution, for some problems an estimate is sufficient or even favorable, due to the accuracy of the underlying data. MC methods can quickly yield a rough estimate of the solution.

Iterative Methods are dependent on the size of the matrix and so the computational effort grows with the problem size. The complexity of these methods is  $O(kn^2)$  for dense matrices in the iterative case and  $O(n^3)$  for direct methods with dense matrices while solving SLAE if common elimination are employed [3] in contrast MC methods for matrix inversion (MI) only require  $O(NL)$  steps to find a single element or a row of the inverse matrix. Here  $N$  is the number of Markov chains and  $L$  is an estimate of the chain length in the stochastic process. These computations are independent of the matrix

size  $n$  and also inherently parallel. Note that in order to find the inverse matrix or the full solution vector in the serial case,  $O(nNL)$  steps are required.

## II. RELATED WORK

Research efforts in the past have been directed towards optimizing the approach of sparse approximate inverse preconditioners (SPAI) [4]. In the past there have been differing approaches and advances towards a parallelisation of the SPAI preconditioner. The method that is used to compute the preconditioner provides the opportunity to be implemented in a parallel fashion. In recent years the class of Frobenius norm minimizations that has been used in the original SPAI implementation [5] was modified and is provided in a parallel SPAI software package. One implementation of it, by the original authors of SPAI, is the Modified SParse Approximate Inverse (MSPAI) [6].

The proposed Monte Carlo algorithm has been developed and enhanced upon in the last decades, and several key advances in serial and parallel Monte Carlo methods for solving such problems have been made. There is an increased research interest in parallel Monte Carlo methods for Linear Algebra in the past year [7], [8].

## III. MONTE CARLO APPROACH

Monte Carlo methods are probabilistic methods, which use random numbers to either simulate a stochastic behaviour or to estimate the solution of a problem. They are good candidates for parallelisation because of the fact that many independent samples are used to estimate the solution. These samples can be calculated in parallel, thereby speeding up the solution finding process. We design and develop parallel Monte Carlo methods with the following main generic properties:

- efficient distribution of the compute data
- minimum communication during the computation
- increased precision achieved by adding extra refinement computations

Consideration of all these properties naturally leads to scalable algorithms. The Procedure to get Monte Carlo algorithm has been presented in [9] and allows extending Monte Carlo algorithm for processing diagonally dominant matrices.

## IV. MONTE CARLO APPROACH

We compared matrices from different sets that have been obtained from two collections - The Matrix Market and The University of Florida Sparse Matrix Collection. These

matrices are used as inputs to both the MSPAI and our Monte Carlo based application to compute preconditioners. The results from those calculations are two intermediate matrices MSPAI and MCSPAI, one for each type of preconditioner. In the last step these preconditioners are used as an input to the BiCGSTAB (biconjugate gradient stabilized method) solver that is provided by the SPAI application. Numerical experiments have been executed on the MareNostrum supercomputer, located at the Barcelona Supercomputing Center (BSC). It currently consists of 3056 compute nodes that are each equipped with 2 Intel Xeon 8-core processors, 64GB RAM and are connected via an InfiniBand FDR-10 communication network. The experiments have been run multiple times to account for possible external influences on the results. The computation times for both the preconditioner calculated by MSPAI, as well as our Monte Carlo based result, have been noted. While conducting the experiments, we configured the parameters for probable errors in both programs to produce preconditioners with similar properties and therefore producing residuals within similar ranges when used as preconditioners for BiCGSTAB. A random starting pattern has been chosen in MSPAI for best analogy to the stochastic nature of the Monte Carlo approach. A basic experiment was carried out on various classes of matrices from the matrix market, for example, on real non-symmetric matrix as **bcstm13** (fig. 1)

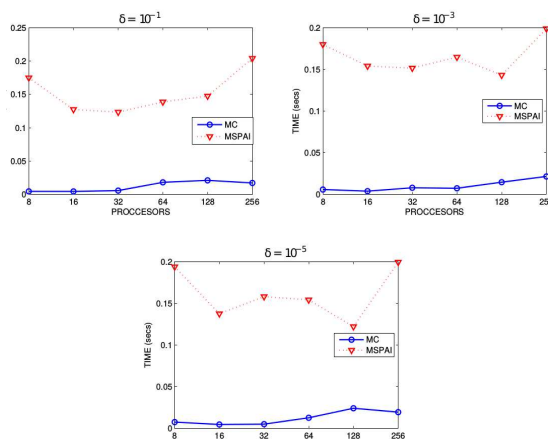


Fig. 1. Matrix bcstm13. Run times for MC preconditioner and MSPAI.

The number of compute cores used to calculate the preconditioners has been selected to match the problem size and provide meaningful comparisons between the two differing approaches. For the small test set experiments have been run on 6 to 30 cores of the computer system. The larger examples were calculated starting from 32 or 256 cores and scaling up to 2048 cores to investigate the differing scaling behavior of the deterministic method and our stochastic algorithm.

## V. SOME CONCLUSIONS AND FUTURE WORK

A Monte Carlo based preconditioner for general matrices has been proposed as an alternative to the MSPAI algorithm and its applicability demonstrated. It has been shown that the stochastic Monte Carlo approach is able to produce

preconditioners of comparable quality to the deterministic approach used in MSPAI. The proposed approach enabled us to generate preconditioners efficiently (faster) and with good scaling properties outperforming the deterministic approach, especially for larger problem sizes.

Monte Carlo Method approach is a promising tool to solve the steady state heat transfer equation  $-k\Delta T = 0$  in order to obtain effective homogenized heat conductivity coefficient for the porous Siliconized Silicon Carbide material, Margenov [10] propose to use MPI parallelized preconditioned conjugate gradient method with a preconditioner.

## ACKNOWLEDGMENT

The author would like to thank CONACYT-México for supporting a postdoctoral position in BSC.

## REFERENCES

- J. Strassburg and V. Alexandrov, "Enhancing Monte Carlo Preconditioning Methods for Matrix Computations", *Procedia Computer Science*, Vol. 29, pages 1580-1589, 2014.
- G.H. Golub, C. Loan, "Matrix computations", *Johns Hopkins University Press*, ISBN 9780801854149, 1996 .
- M. Ferronato, E.Chow, K.Phoon, "Preconditioning Techniques for Sparse Linear Systems", *Journal of Applied Mathematics*, Hindawi Publishing Corporation, Vol. 2012, article-id:518165, 2012.
- G. Allon, M. Benzi, and L. Giraud, "Sparse approximate inverse preconditioning for dense linear systems arising in computational electromagnetics", *Numerical Algorithms*, vol. 16, num. 1, pages 115, 1997.
- M. Benzi, C. Meyer, and M. Tuma, "A Sparse Approximate Inversepreconditioner for the Conjugate Gradient Method", *SIAM Journal on Scientific Computing*, Vol.5, pages 11351149, 1996.
- T. Huckle, A. Kallischko, A. Roy, M. Sedlacek, and T. Weinzierl, "Anefficient parallel implementation of the MSPAI preconditioner, Parallel Computing", *Parallel Matrix Algorithms and Applications*, vol.36, num. 56, pages 273284, 2010.
- V. Alexandrov, E. Atanassov, I. Dimov, S. Branford, A. Thandavan and C. Weihrauch, "Parallel Hybrid Monte Carlo Algorithms for Matrix Computations", In V. Sunderam, G. Albada, P. Sloot, and J. Dongarra, editors, *Computational Science ICCS 2005*, vol. 3516 of Lecture Notes in Computer Science, pages 752759. Springer Berlin / Heidelberg, 2005.
- V. Alexandrov and A. Karaivanova, "Parallel Monte Carlo algorithms for sparse SLAE using MPI", in *Recent Advances in Parallel Virtual Machine and Message Passing Interface*, pages 283290, Springer, 1999.
- S. Branford, "Hybrid Monte Carlo Methods for Linear Algebra Problems", PhD thesis, *School of Systems Engineering*, The University of Reading, April 2009.
- S. Margenov, "Robust multilevel methods for strongly heterogeneous problems and supercomputing applications in porous media", plenary talk, *5th International Conference on Porous Media*, Prague, Czech Republic, May 21-24, 2013.



# Supercomputing simulations for Beyond Standard Model theories at the TeV scale: non-standard gravity and electroweak interactions

Rafael L. Delgado

Dept. Física Teórica I, Universidad Complutense de Madrid, Madrid (Spain)

*rdelgado@ucm.es*

At the TeV scale, two new pieces of information come from the LHC. The first one, a new bosonic resonance [1-2] at about 125 GeV which, until now, fits the properties of the Standard Model (SM) Higgs boson. The second one [3], a mass gap for the presence of new physics until 600-700 GeV (and even higher for new vector resonances).

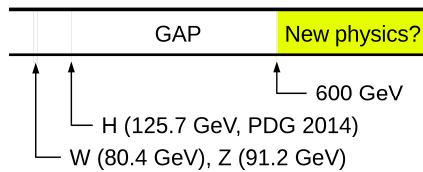


Figure 1: Electroweak-symmetry breaking sector of the SM after the LHC run I.

So, this is the image which provides us the Standard Model from some decades ago: three fundamental interactions (electromagnetic, weak and strong) mediated by photons, gauge bosons  $W^\pm$  and  $Z$ , and gluons. Photons and gluons are massless. Gauge bosons mediating the weak interaction are massive. However, these masses, and actually all masses of elementary particles (i.e., electrons, quarks,...), are in conflict with the symmetries of Quantum Field theories formalisms.

The simplest solution is the so-called Higgs mechanism, the SM Higgs boson being a relic of such a mechanism. So, if the Higgs-like boson were actually this SM Higgs, would our understanding of the micro-world complete? Well, of course there is another interaction we have not yet talked about: gravity. Actually, there is no *widely accepted and fully coherent* description of gravity in terms of a Quantum Field Theory.

Thus, we have two “fundamental theories”: General Relativity, which is a classical (i.e., *non-quantum*) theory; and SM (Quantum Chromo Dynamics-Electroweak Theory), which describes the remaining known interactions, the SM Higgs field mediating a fifth interaction which gives mass to all the elementary particles. Several difficulties arise when dealing with Quantum Gravity theories. The first one is the extremely high energy scale where Quantum Gravity effects are expected to dominate. According to a naïve approach, such scale would be the Planck one ( $1.22 \times 10^{19}$  GeV). Compare with the  $1.4 \times 10^4$  GeV that LHC will reach when running at full energy. So, is this all? The SM until the Planck scale? Actually, there are two effects which fail to be described by this approach: Dark Matter and Dark Energy.

In this work, we follow two research lines. The first one [4] is a Kaluza-Klein model (i.e., extra-dimensions). The world would be five-dimensional, the SM particles being confined fields into three spatial dimensional manifolds or branes. Gravitational interaction would have access to the total space. This model leads to a fundamental scale of gravitation  $M_D$  which can be much lower [5-6] than the Planck one  $M_P$ . Depending on the flexibility of the brane, two different regimes at *low energy* arise.

If the brane is rigid, the first excitation at low energy corresponds to gravitons. The fact that gravitons propagate through the *extra-dimensions*, produces a tower of KK excitations which roughly corresponds to the eigen-energies of a quantum oscillator in these dimensions. This can be approximated by the coupling with massive gravitons.

On the contrary, if the brane is flexible, the first effect at low energy will be the coupling with *branons*, which are the particle associate with the fluctuations of the brane.

The second research line considered [7-9] is a strongly interacting Higgs sector. The SM lore would have the property of being a weakly interacting theory. In particular, the presence of the SM Higgs on the theory would guarantee that the cross section of vector boson scattering, even computed at first order in perturbation theory, does not grow breaking unitarity (i.e., until predicting a scattering probability higher than 1).

What if we change slightly the SM? This “fine-tuning” cancellation would no longer occur, so this first order computation would break at the TeV scale. Is this incorrect? Not exactly: it only means that the perturbation theory is not usable. This is a similar case to the low energy limit of QCD, which has been a big deal for decades. Several techniques, like unitarization procedures, lattice QCD... have been used to overcome this failure of the perturbative analysis. So, we use the unitarization methods, tested for the low energy limit of QCD, to unitarize an effective Lagrangian which leads to a strongly interacting EWSBS.

In both cases, to have an estimation of the signals that the LHC may find, using Monte Carlo (MC) simulation is mandatory. The input of the MC methods are the scattering amplitudes computed with the theoretical models. And the output, a set of MC events whose statistics (i.e., counting of particles emitted within certain parameter space, a particular measure,...) should be (approximately) described by the same statistical distribution than the actual data of the LHC.

Several programs, like Pythia [10], MadGraph [11], Delphes, GEANT, ROOT,... have become a standard, as a consequence of the big effort due to the LHC experiment. However, it is also usual that, when trying to set a “non-standard” model, some of these programs fail to give an implementation in an easy way. With the only exception of “fashionable” models like the Minimal Super Symmetry Model (MSSM).

When running, these programs can be easily parallelized if different seeds are used for the pseudo-random number generator of each process. No communication between processes is necessary, until summing up all the statistics. Normally, the output of a program is the input of the next one. Two approaches are possible: either computing the statistics “on the fly”, or storing all the events. The last possibility, although is the most flexible one, requires also a huge amount of hard disk.

The goal of this presentation is showing the results obtained by the usage of the BSC facilities. For the KK models, parameter bounds for both the branon and KK-graviton regimes have been found, by comparing a MC simulation with ATLAS data [12]. New bounds over the brane tension parameter  $f$  (defined on [4]) have been published by CMS Collaboration [13].

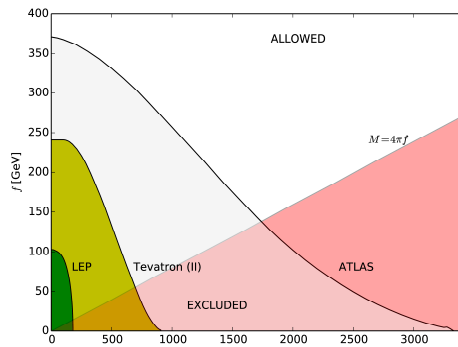


Figure 2: Computed exclusion region (according to [4]) for the value of  $f$  parameter of the branon (gray area) versus the limits of the second run of Tevatron (dark yellow area  $\sqrt{s} = 1.96$  TeV and  $\int L dt = 200$  pb $^{-1}$ ) and LEP (green area,  $\sqrt{s} = 189 - 209$  GeV). See ref. [4] for more details.

For the strongly interacting EWSBS, a graph showing how small are the contribution of longitudinal W modes to the WW scattering (as measured by the LHC) has been computed. If the SM were true, scattering of longitudinal modes of W and Z's would be negligible. However, if the EWSBS were strongly interacting, the cross section of  $W_L W_L$  scattering would saturate unitarity and, therefore, this signal would be enhanced. An interesting continuation of this work would be simulating events with our unitarized matrix elements, although some difficulties arise due to the necessity of modifying the MC programs.

#### ACKNOWLEDGMENT

The author thankfully acknowledge the computer resources, technical expertise, and assistance provided by the Barcelona Supercomputing Centre—Centro Nacional de

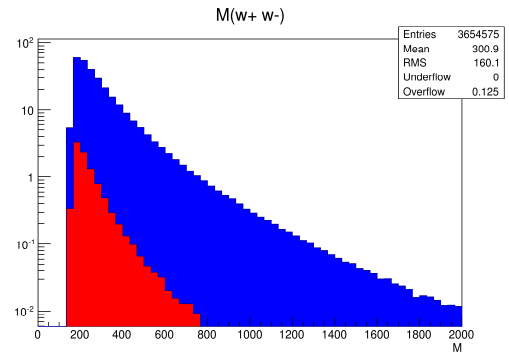


Figure 3: Production of  $W^+W^-$  (blue) vs.  $W^+_L W^-_L$  (red). X-axis in GeV. Y-axis in events/33.3 GeV. MadGraph5\_aMC@NLO used.

Supercomputación—and the Tirant supercomputer support staff at Valencia. We are also pleased to thank Professors Antonio L. Maroto, Torbjörn Sjöstrand, J.R. Peláez for their

useful help and comments. RDL thanks the HEP group at the University of Southampton (NEXT institute) for its hospitality. This work has been supported by MICINN (Spain) Project Nos. FIS 2008-01323, FIS2011-23000, and FPA2011-27853-01 and Consolider-Ingenio MULTIDARK CSD2009-00064, and by the Spanish grant UCM:910309 R. L. D. is a fellow of FPI, Ref. BES-2012-056054.

#### REFERENCES

- [1] G. Aad et al. [ATLAS Collaboration], Phys. Lett. B 716, 1 (2012)
- [2] S. Chatrchyan et al. [CMS Collaboration], Phys. Lett. B 716, 30 (2012)
- [3] S. Chatrchyan et al. [CMS Collaboration], Phys. Lett. B 704, 123 (2011); G. Aad et al. [ATLAS Collaboration], Phys. Lett. B 712, 22 (2012); G. Aad et al. [ATLAS Collaboration], Phys. Lett. B 722, 305 (2013).
- [4] J. A. R. Cembranos, R. L. Delgado and A. Dobado, “*Brane-Worlds at the LHC: Branons and KK-gravitons*”, 2013, Phys. Rev. D, 88, 075021
- [5] V.A. Rubakov and M.E. Shaposhnikov, Phys. Lett. 125 B, 136 (1983)
- [6] N. Arkani-Hamed, S. Dimopoulos, and G.R. Dvali, Phys. Lett. B 429, 263 (1998); Phys. Rev. D 59, 086004 (1999); I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, and G.R. Dvali, Phys. Lett. B 436, 257 (1998)
- [7] Rafael L. Delgado, Antonio Dobado, Felipe J. Llanes-Estrada, “*Unitarity, analyticity, dispersion relations and resonances in strongly interacting  $W_L W_L$ ,  $Z_L Z_L$  and  $hh$  scattering*”, arXiv:1502.04841 [hep-ph]
- [8] Rafael L. Delgado, Antonio Dobado, Felipe J. Llanes-Estrada, “*One-loop  $W_L W_L$  and  $Z_L Z_L$  scattering from the electroweak Chiral Lagrangian with a light Higgs-like scalar*”, JHEP 1402 (2014) 121
- [9] Rafael L. Delgado, Antonio Dobado, Felipe J. Llanes-Estrada, “*Light ‘Higgs’, yet strong interactions*”, J.Phys. G 41 (2014) 025002
- [10] T. Sjostrand, S. Mrenna, and P.Z. Skands, J. High Energy Phys. 05 (2006) 026; Comput. Phys. Commun. 178, 852 (2008).
- [11] J. Alwall et al, JHEP 1407 (2014) 079
- [12] G. Aad and T. Abajyan (ATLAS Collaboration), Phys. Rev. Lett. 110, 011802 (2013)
- [13] V. Khachatryan et al. [CMS Collaboration], CMS-EXO-12-047, CERN-PH-EP-2014-253, arXiv:1410.8812 [hep-ex]

# Discovering most significant news using Network Science approach

Ilya Blokh<sup>1</sup> and Vassil Alexandrov<sup>2</sup>

<sup>1</sup>Perm State University, Russia,

<sup>2</sup>ICREA-Barcelona Supercomputing Centre, Spain.

ilyablokh@gmail.com, vassil.alexandrov@bsc.es

## Abstract

The role of Internet mass media increased greatly in the recent years. Internet media attracts more and more audience coverage and thereby its role in forming of public opinion strengthens. Public organizations, political parties, companies as well as public figures, politicians and entrepreneurs are interested in any mentioning of them or of some specific facts concerning them in media including Internet media. They often trace a quantity of web-sites citing them, and different sources where some significant news could appear. Significant news finding could be also useful for common Internet uses since the amount of daily media publications is large and some news which are significant for someone could be missed. Thus significant (in some context) news identification is essential and actual problem.

Approaches for analyzing various types of news are widely different. From traditional search engines based on indexing to semantic methods based on idea of Semantic Web, keywords and key phrases processing, NLP methods. Ontology based approaches have a remarkable prevalence too. In addition, some methods based on Data Mining technologies and machine learning used to resolve a problem of news search, interpretation and analysis. This paper is devoted to the use of Network Science approach for news analysis.

The origin of news provider could be some Internet mass-media (such as cnn.com, bbc.com and others) as well as a page in a social network or blog. In our research we decided to take Twitter as a typical representative of the social networking phenomenon as it is one of most subscribed social networks. Messages spread in Twitter after their appearance and a quantity of references to the message (tweet) appears. One of the spreading instruments of this social network is retweet mechanism. On the first stage of our research we focus on this mechanism and base our approach on small entities of Twitter - on tweets. We use network science approach to analyze messages and find significant news. Tweets act as nodes of network. Each tweet could be retweeted many times by some external twitter users. We define each retweet being one outgoing tie from a node. Thus for each tweet feed we have some degree distribution. Next we examine this distribution for scale-free property, define the significance and the

threshold for it. Then we reveal a set of significant news from a given sample.

We interpret the news significance as a qualitative property of some piece of news and in the same way try to define some qualitative threshold to measure it. Significance analysis could help to characterize how different events impact on society or on separate social groups, to make research about events interconnection and to analyse news information spreading features. We define a quantitative metric of news significance as its retweet count for some time period. In the subsequent research we expect that this metrics will become more complex by aggregating information not only about retweet count but also include the information about number of user replies on the given tweet, number of users who put the tweet in favorites and some additional data.

We crawled data from Twitter accounts of the ones of the popular news agencies represented in social network: CNN, BBC, NYTimes, Mashable and TechCrunch. Collected data covers period from July 2014 up to January 2015 and includes information about 16500 news retweets. We looked at each agency news feed as at network and using Network Science apparatus we built and plotted degree distributions for each agency. It was determined in other research works that the whole Twitter network has scale-free property. Our goal was to specify degree distribution of sub-networks based on news messages retweets and check the presence of scale-free property. So we revealed that composed networks are scale-free and then supposed that the news is significant thus meeting the condition that the corresponding node in the network is a hub. It is crude initial measure and will be refined further taking into account a variety of additional factors. We discovered a number of significant news from initial data and presented them in the paper.

Afterwards we determine the future research scope. First we plan to make the significance property more complex and take into attention more factors and increase the number of data sources and the volume of processed data. Then we will describe the optimization approach for significant news finding and analyzing. In general we project to extend the

problem of significant news discovering and interpretation into the problem of world informational war analysis. Here we define informational war as a demonstration of multidirected regular actions forwarded into influencing at social opinion in some area and consisted in mass media publication's policy, in Internet media particularly.

#### REFERENCES

- [1] Quint Randle: A historical overview of the effects of new mass media: Introductions in magazine publishing during the twentieth century. Volume 6, Number 9 - 3 September 2001.
- [2] H. Kwak, C. Lee, H. Park, S. Moon. What is Twitter, a social network or a news media?, Proceedings of the 19th international conference on World wide web, 591-600, 2010.
- [3] Albert-László Barabási (2013). Network Science.
- [4] Gourab Ghoshal, Albert-László Barabási (2011). Ranking stability and super-stable nodes in complex networks.
- [5] Dashun Wang, Zhen Wen, Hanghang Tong, Ching-Yung Lin, Chaoming Song, Albert-László Barabási (2011)/ Information Spreading in Context.
- [6] G. Ghoshal, L. Chi, A.-L. Barabási. (2013) Uncovering the role of elementary processes in network evolution. Scientific Reports 3, pp.1-8.

# Linompss – A Linear Algebra Library on OMPSS

Sicong Zhuang  
Computer Science Department - BSC  
sicong.zhuang@bsc.es

**Abstract-** Exascale performance will require massive parallelism and asynchronous execution (DARPA, DOE, EESI2). The former pertains to the design choice to increase hardware performance through growing core counts. The latter ensures that the associated software scales well. As a result, traditional, bulk-synchronous parallel programming models like OpenMP or MPI will likely fall out of grace in favor of more amenable variants. At the time of writing, the general consensus inclines towards task-based programming models that support dynamic scheduling and use graph-based models to record task dependencies.

In this context, we present LINOMPSSs, a parallel linear algebra library built on top of OMPSSs. Currently, we offer most of the BLAS-3 functionality and the three main factorizations (QR, LU, Cholesky). There is limited support for sparse matrices. For systems of equations, LINOMPSSs disposes of implementations of both direct methods as well as iterative solvers.

At the lowest level, LINOMPSSs can be linked with most BLAS or LAPACK libraries (MKL, ESSL, Netlib, ATLAS...). These functions are used, in turn, to implement the OMPSSs tasks for the parallel implementations of the BLAS, or the Blocked BLAS. The Blocked BLAS are used to write blocked versions of the factorizations in OMPSSs, which can then be combined to write solvers or full-fledged numerical simulations with finite element methods, multigrid methods, etc. These applications benefit from the asynchronous parallelism developed by OMPSSs at run time. Moreover, the layered design allows for the concurrent and out-of-order execution of tasks that stem from previously distant regions of computation.

We not only consider LINOMPSSs as an environment for the development of linear algebra applications and benchmarks for the Computer Science Department. In our own department for example, we use LINOMPSSs to experiment with mixed-precision and incomplete/inaccurate computations. With the expertise of the CASE and Earth Science Department at BSC, we envision LINOMPSSs as a tool for cross-disciplinary collaboration and the development of industrial-strength engineering applications.

## A. INTRODUCTION

LINOMPSSs (LINEar algebra on OMPSSs) is a numerical linear algebra library developed in the Computer Science department of BSC. It strives to provide high performance and scalability on heterogeneous multi-core architectures, as well as to offer a clear interface as a math kernel for building large scientific/engineering applications.

## B. GENERAL STRUCTURE

At its core LINOMPSSs relies on an optimized implementation of the BLAS and LAPACK (ATLAS, Intel MKL, etc.). Each of the functions in those libraries has a corresponding wrapper in LINOMPSSs which is compiled with OMPSSs task support. A parallelized version of the BLAS functions (GEMM(), SYRK(), TRSM() etc.) is then realized by blocking the input matrices and passing the blocks to the aforementioned OMPSSs tasks. The OMPSSs runtime handles

the task dependencies and dynamically schedules the tasks in an asynchronous fashion.

Along the same lines, we have implemented several LAPACK functions (Cholesky factorization, LU factorization) and solvers (Jacobi, Gauss-Seidel, Conjugate Gradient).

## C. FUNCTIONALITY

The very basic level exposed to the user is a set of wrappers of the BLAS and LAPACK functions compiled with OMPSSs tasks as a dynamic library. This set of wrappers also serves as the building block for the other LINOMPSSs functions.

Going one level above are three blocked parallel BLAS level 3 functions (GEMM(), SYRK(), TRSM()). From this level on we also build a standalone program for benchmarking along with the dynamic libraries.

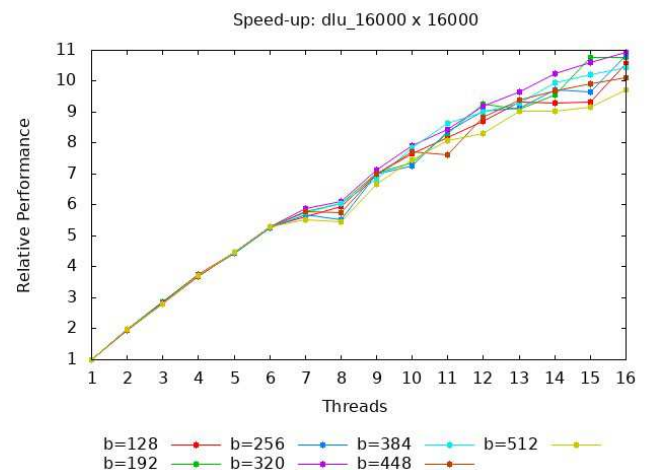
With these blocked functions at hand plus a few wrappers we are able to build two blocked parallel matrix factorization functions LU and Cholesky.

Subsequently, two blocked parallel linear equation solvers (one for general matrix and the other for a positive-definite matrix respectively) are built on top of that. Using the LU/Cholesky factorization to decompose the input matrix and applying forward-, backward-substitution to solve the system.

The current version of Linompss also ships with some iterative solvers (Conjugate Gradient, Iterative Refinement, Jacobi and Gauss-Seidel).

## D. PERFORMANCE

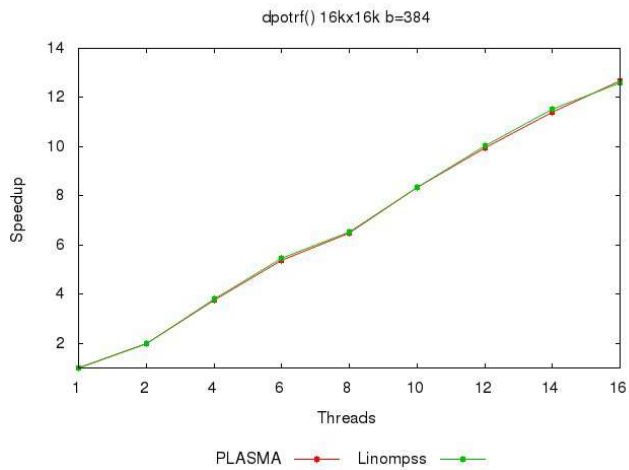
Experimenting on the standalone benchmark programs from Linompss we were able to extract some speedup plots (on MareNostrum III – BSC).



## E. COMPARISON AND FUTURE WORK

We are aware of another state-of-the-art numerical linear algebra library PLASMA that takes a rather similar approach (blocked/tiled algorithm and a dynamic scheduler).

We thus drew a comparison on two functions `dpotrf()` from both libraries with the same set of parameters. As is shown in the figure below both libraries demonstrate a fairly good scalability trend.



As of now Linompss is under active development. We are planning to further extend its supported functions (QR factorization, various sparse functions to name a few) and to exploit the OMPSS programming model to a wider extent (priority, support for heterogeneous devices etc.).

# Criticality-Aware Dynamic Task Scheduling for Heterogeneous Systems

Kallia Chronaki<sup>\*</sup>, Alejandro Rico<sup>\*</sup>, Rosa M. Badia<sup>\*†</sup> and Eduard Ayguadé<sup>\*‡</sup>

<sup>\*</sup>Barcelona Supercomputing Center, Barcelona, Spain

<sup>†</sup>Research Institute (IIIA) - Spanish National Research Council (CSIC), Barcelona, Spain

<sup>‡</sup>Universitat Politècnica de Catalunya - BarcelonaTech, Barcelona, Spain

*kallia.chronaki@bsc.es, alejandro.rico@bsc.es, rosa.m.badia@bsc.es,*

*eduard.ayguade@bsc.es*

## INTRODUCTION

In the search of performance and energy efficiency, heterogeneous multi-core architectures are an appealing option for next-generation high-performance computing.

These architectures combine different types of processing cores designed at different performance and power optimization points, thus exposing a performance-power trade-off.

Current and future parallel programming models need to be portable and efficient when moving to such systems. Load balancing and scheduling are two of the main challenges in utilizing such heterogeneous platforms. The use of task-based programming models with dynamic scheduling is a way to tackle these challenges. Some of these programming models allow the specification of inter-task dependencies that enable automatic scheduling and synchronization by the runtime system.

OmpSs is a powerful task-based programming model with dependency tracking and dynamic scheduling. In this talk we will describe the OmpSs approach on scheduling dependent tasks onto the asymmetric cores of a heterogeneous system. The proposed dynamic scheduling policy uses information discoverable at runtime and reduces total execution time. It first prioritizes the newly-created tasks at runtime according to the shape of the task dependency graph; it then detects the longest path of the dynamic task dependency graph, and finally it assigns critical tasks to fast cores and non-critical tasks to slow cores.

Previous works on scheduling for heterogeneous systems by using task priorities require the prior knowledge of various parameters of the workload; for example the task execution time, which cannot be discoverable without profiling or the task dependency graph of the workload.

The experimental evaluation proves that our implementation speeds up the execution of four scientific kernels on the ARM big.LITTLE heterogeneous chip. Our proposal outperforms a dynamic implementation of the state-of-the-art Heterogeneous Earliest Finish Time scheduler by up to 1.15x, and the default breadth-first OmpSs scheduler by up to 1.3x in the 8-core heterogeneous

ARM big.LITTLE and up to 2.7x in a simulated 128-core chip.

Previous criticality-aware schedulers for heterogeneous systems are static and based on the knowledge of profiling information. Our proposal performs dynamic scheduling using information discoverable at runtime, is implementable and works without the need of an oracle or profiling.

In our evaluation using four dependency-intensive applications, our proposal outperforms a dynamic implementation of Heterogeneous Earliest Finish Time by up to 1.15x, and the default breadth-first OmpSs scheduler by up to 1.3x in a real 8-core heterogeneous platform and up to 2.7x in a simulated 128-core chip.

## RELATED WORK

Several previous works propose scheduling heuristics that focus on the critical path in a task dependency graph (TDG) to reduce total execution time [1, 2, 3, 4]. To identify the tasks in the critical path, most of these works use the concept of *upward rank* and *downward rank*. The upward rank of a task is the maximum sum of computation and communication cost of the tasks in the dependency chains from that task to an exit node in the graph. The downward rank of a task is the maximum sum of computation and communication cost of the tasks in the dependency chain from an entry node in the graph up to that task. Each task has an upward rank and downward rank for each processor type in the heterogeneous system, as the computation and communication costs differ across processor types.

The Heterogeneous Earliest Finish Time (HEFT) algorithm [4] maintains a list of tasks sorted in decreasing order of their upward rank. At each schedule step, HEFT assigns the task with the highest upward rank to the processor that finishes the execution of the task at the earliest possible time. Another work is the Longest Dynamic Critical Path (LDPC) algorithm [1]. LDPC also statically schedules first the task with the highest upward rank on every schedule step. The difference between LDPC and HEFT is that LDPC updates the computation and communication costs on multiple processors

of the scheduled task by the computation and communication cost in the processor to which it was assigned.

The Critical-Path-on-a-Processor (CPOP) algorithm [4] also maintains a list of tasks sorted in decreasing order as in HEFT, but in this case it is ordered according to the addition of their *upward rank* and *downward rank*. The tasks with the highest *upward rank* + *downward rank* belong to the critical path. On each step, these tasks are statically assigned to the processor that minimizes the critical-path execution time.

The main weaknesses of these works are that (a) they assume prior knowledge of the computation and communication costs of each individual task on each processor type, (b) they operate statically on the whole dependency graph, so they do not apply to dynamically scheduled applications in which only a partial representation of the dependency graph is available at a given point in time, and (c) most of them use randomly-generated synthetic dependency graphs that are not necessarily representative of the dependencies in real workloads.

#### REFERENCES

- [1] M. Daoud and N. Kharm. Efficient Compile-Time Task Scheduling for Heterogeneous Distributed Computing Systems. In Parallel and Distributed Systems, 2006. ICPADS 2006. 12th International Conference on, volume 1, pages 9, 2006.
- [2] M. Hakem and F. Butelle. Dynamic Critical Path Scheduling Parallel Programs onto Multiprocessors. In In the Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium, IPDPS'05, pages 203b-203b, April 2005.
- [3] C.-H. Liu, C.-F. Li, K.-C. Lai, and C.-C. Wu. A dynamic Critical Path Duplication Task Scheduling Algorithm for Distributed Heterogeneous Computing Systems. In Parallel and Distributed Systems, 2006. ICPADS 2006. 12<sup>th</sup> International Conference on, volume 1, pages 8, 2006.
- [4] H. Topcuoglu, S. Hariri, and M.-Y. Wu. Performance-Effective and Low-Complexity Task Scheduling for Heterogeneous Computing. IEEE Transactions on Parallel and Distributed Systems, 13(3):260-274, Mar 2002.



# High Performance Computing Based Simulation for Healthcare Decision Support

Zhengchun Liu, Emilio Luque  
Computer Architecture & Operating Systems  
University Autònoma of Barcelona (UAB), Spain  
E-mail: lzhengchun@caos.uab.es, emilio.luque@uab.es

**Abstract** - Due to the complexity and crucial role of an Emergency Department (ED) in the healthcare system, the ability to more accurately represent, simulate and predict performance of ED will be invaluable for decision makers to solve management problems. One way to realize this requirement is by modeling and simulation. The objective of this research is to grasp the non-linear association between macro-level features and micro-level behavior with the goal of better understanding the bottleneck of ED performance and provide ability to quantify such performance on defined condition. Agent-based modeling approach was used to model the healthcare staff, patient and physical resources in ED. Instead of describe all the potential causes of this complex issue. Rather, in this thesis, a layer-based application framework will be presented to discover knowledge of a complex system through simulating micro-level behaviors of its components to facilitate a systematic understanding of the aggregate behavior.

## I. INTRODUCTION

An emergency department (ED), also known as accident & emergency (A&E), emergency room (ER), or casualty department, is a medical treatment facility specializing in acute care of patients who present without prior appointment, either by their own means or by ambulance, some of which may be life-threatening and must be treated quickly. Such that, ED must operate to provide 24/7 year-round service. Since emergency department is the main entrance to a healthcare system that faces uncertainty every day. The efficiency and quality of service in ED have big influence on the whole healthcare system. Making decisions under multi-constraint to meet uncertainty is a big challenge for ED managers. Simulation methods have long been used to model elements of healthcare systems with a view to analyzing new system designs, retrofitting to existing systems and proposing changes to operating rules. The modeling and simulation provide a quantitative way to analyze the behavior and predict the performance of an ED. There have been fruitful efforts in developing simulation models for solving healthcare management problems. The emergency department is a typical complex system. To perform intensive study, a realistic computational model is compulsory. An approach to modeling this kind of system is by using agent-based modeling and simulation, which is a kind of bottom-up modeling approach to investigate macro-level behavior from macro-level interactions among its components.

The final object of this study is to grasp the association between individual interaction and aggregate behavior with the goal of better understand a complex system. It includes two main parts, a pure agent based model for full emergency department and an application framework to discover knowledge from micro-level interaction data generated by the

agent based model. Due to the complexity of both the real system and model, high performance computing techniques are used to deal with the big number of simulation scenarios and the massive data analysis (see publication [3]).

## II. RESULTS SO FAR

### A. A Generalized Agent-Based Model to Simulate ED

As a typical complex system, the functionality of an emergency department is reflected by the interaction of its components. Therefore, the most important part of the ED model is the behavior model of these components (i.e. patients, sanitary staff and test technicians). Based on this, we created a generalized agent-based model of the emergency departments. It was designed based on the survey of different EDs and with the participation of sanitary staff in ED. More specifically, the following sanitary staff was considered: admission staff for registration service, triage nurses for classifying patients according to their body condition, doctors, nurses and auxiliaries for helping patients move around the ED for tests. All the sanitary staff is modeled as junior or senior according to their expertise. As for the environment model, there are two areas work independently for high acuity patients (area A) with some careboxes (a room with bed and essential equipment) and low acuity patients (area B) with some chairs. Doctors and nurses are specified for different zones but the two zones share all the test services.

The proposed generalized model is a parametric model (same model, different parameter value configuration), the behavior of the agents as well as state transferring model is configurable. Therefore, it is not dedicated with one specific ED, which can be used to simulate different EDs through calibration process. The details about the generalized agent-based model can be found in our publication [2].

### B. The Simulator for a Real ED

The flexibility and adaptability features of this generic model provide a platform for emergency department simulation to accommodate different scenarios without significant modification of the underlying model. It enables the simulation researchers to focus their effort on the understanding of ED behavior rather than developing a theoretical model each time. Based on the generalized model and real data provided by Hospital of Sabadell, we implemented the agent-based model by using Netlogo simulation environment and validated for simulating a real ED in Hospital of Sabadell. Several case studies have been carried out for proving the potential uses of the simulator. For example, to meet the increasing patient arrival overcrowding

problem, a quantitative analysis of the influence of ambulance response time (for departure) over the ED behavior. The features of this validated model as well as case studies can be found in our publication [1].

### C. Application Framework for Knowledge Discovery

We proposed a framework to discover knowledge through simulating individual behavior of system components. The agent-based model was considered as the core as well as data source of the knowledge discovery system. The behavior simulation model can generate interaction information under various configuration scenarios. Analyzing this interaction enables knowledge discovery towards better understanding the complex systemic behavior thoroughly. This makes it possible to explore association between micro-level behaviors of individuals and macro-level patterns that emerge from their interactions, thus enabling users to better understand system's behavior under various conditions without any influence on the real system. Additionally, a layer-based architecture was used to achieve flexibility and metadata monitoring configurability. The application framework to Discover knowledge from micro-level behavior is an in-depth the root causes way to fully insight into the complex system, it not only able to do prediction, but also provide knowledge on how predictions are made. Several demonstrated case studies were provided to show the potential use of the presented framework as well as how small changes in procedure yield important changes in flow.

### III. FUTURE WORK

ED is the main entrance to the healthcare service; some problems of the healthcare service system are caused by the performance of ED. However, the ED is not independent, all the departments of healthcare system influence each another. Thus, our future work also includes creating the simulator of other healthcare departments, for example the hospital wards to close the simulation loop of the whole healthcare service system.

In addition, in order to assess the impact of the parameters and decision rules within the model, a sensitivity analysis was performed to determine how model is sensitive to changes in the parameters value. Sensitivity analyses are necessary to explore the behavior of complex system models, because the structural complexity of the modeled process and the model is coupled with a high degree of uncertainty in estimating the values of many of the input parameters. Sensitivity analysis for agent-based models provides understanding of the influence of the different input parameters and their variations on the model outcomes. Therefore, global sensitivity analysis of the emergency department will be a scope of my future work.

Furthermore, look deep into the system in one side is for providing information, another challenge for supporting critical decision is to find a balance or tradeoff between different parameter configurations, for instance service for patients and efficiency for providers. Accordingly, optimization based framework for managing complex

processes in the healthcare domain is another scope of future work.

### PUBLICATIONS

[1] **Zhengchun Liu**, Eduardo Cabrera, Manel Taboada, Francisco Epelde, Dolores Rexachs and Emilio Luque. Quantitative Evaluation of Decision Effects in the Management of Emergency Department Problems. International Conference on Computational Science, Reykjavík, Iceland. June, 2015. (*Accepted*)

[2] **Zhengchun Liu**, Eduardo Cabrera, Dolores Rexachs and Emilio Luque. A Generalized Agent-Based Model to Simulate Emergency Departments. The Sixth International Conference on Advances in System Simulation, Nice, France. October, 2014. (*Published*)

[3] **Zhengchun Liu**, Eduardo Cabrera, Dolores Rexachs and Emilio Luque. Study of Emergency Department by Using High Performance Computing. XXV Jornadas de Paralelismo Valladolid, Spain. Septiembre 2014. Jornadas sarteco 2014 (*Published*)

[4] **Zhengchun Liu**, Emilio Luque, Dolores Rexachs, Francisco Epelde, Eduardo Cabrera, Manel Taboada González. A Simulator of Emergency Departments for Decision Support and QoS Improving. V CONFERENCE: R+D+I Research and Development in ICT and Health. Girona, Spain, June 2014. (*Poster*)

### ACKNOWLEDGMENT

This research has been supported by the MINECO (MICINN) Spain under contract TIN2011- 24384 and has been partially supported by China Scholarship Council (CSC) under reference number: 201306290023.

# Hardware Scheduling Algorithms for Asymmetric Single-ISA CMPs

Nikola Markovic, Daniel Nemirovsky, Osman Unsal, Mateo Valero, Adrian Cristal  
Barcelona Supercomputing Center  
Universitat Politècnica de Catalunya  
{nikola.markovic, daniel.nemorivsky, osman.uunsal, mateo.valero, adrian.cristal}@bsc.es

## Abstract

As thread level parallelism in applications has continued to expand, so has research in chip multi-core processors. Since more and more applications become multi-threaded we expect to find a growing number of threads executing on a machine. Consequently, the operating system will require increasingly larger amounts of CPU time to schedule these threads efficiently. Instead of perpetuating the trend of performing more complex thread scheduling in the operating system, we propose a two lightweight hardware thread scheduling mechanisms. First is a Hardware Round-Robin Scheduling (HRRS) policy which is influenced by Fairness Scheduling techniques thereby reducing thread serialization and improving parallel thread performance. Second is a Thread Lock Section-aware Scheduling (TLSS) policy which extends HRRS policy. TLSS policy is influenced by the Fairness-aware Scheduling and bottleneck identification techniques. It complements the HRRS scheduler by identifying multi-threaded application bottlenecks such as thread synchronization sections. We show that HRRS outperforms Fairness scheduler by 17 percent while TLSS outperforms HRRS by 11 percent on an ACMP consisted of one large (out-of-order) core and three small (in-order) cores.

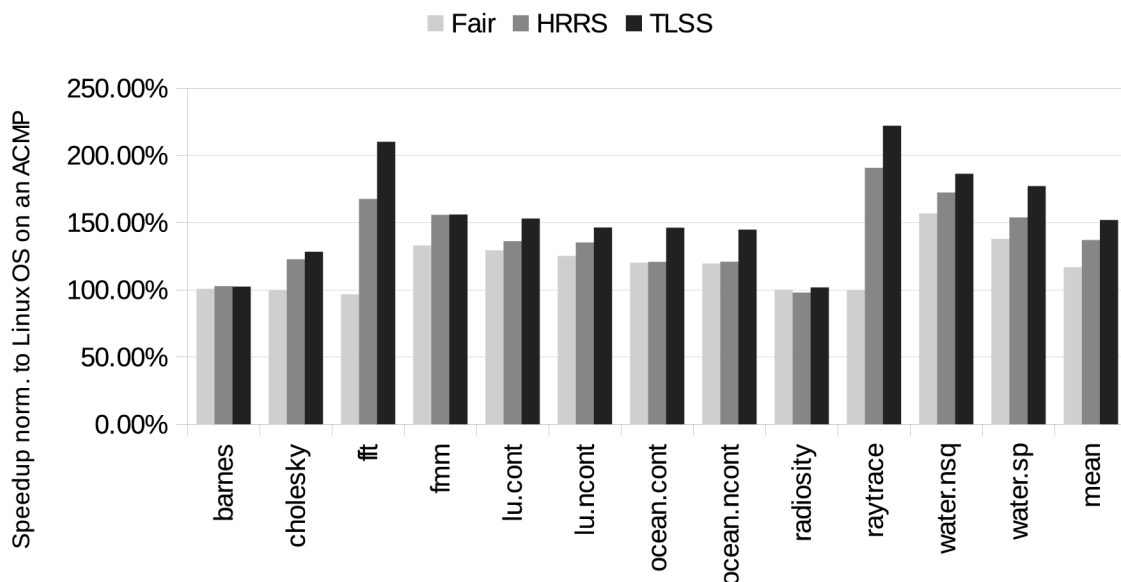


Fig1. Speedup comparison of the TLSS, the HRRS and the Fairness scheduler normalized to Linux OS scheduler for the SPLASH-2 benchmark suite running on four cores (1 OoO+3 InO).

# Single amino acid mutation controls hole transfer dynamics in DNA-methyltransferase HhaI complexes

Marina Corbella,<sup>a</sup> Alexander A. Voityuk<sup>bc</sup> and Carles Curutchet<sup>a</sup>

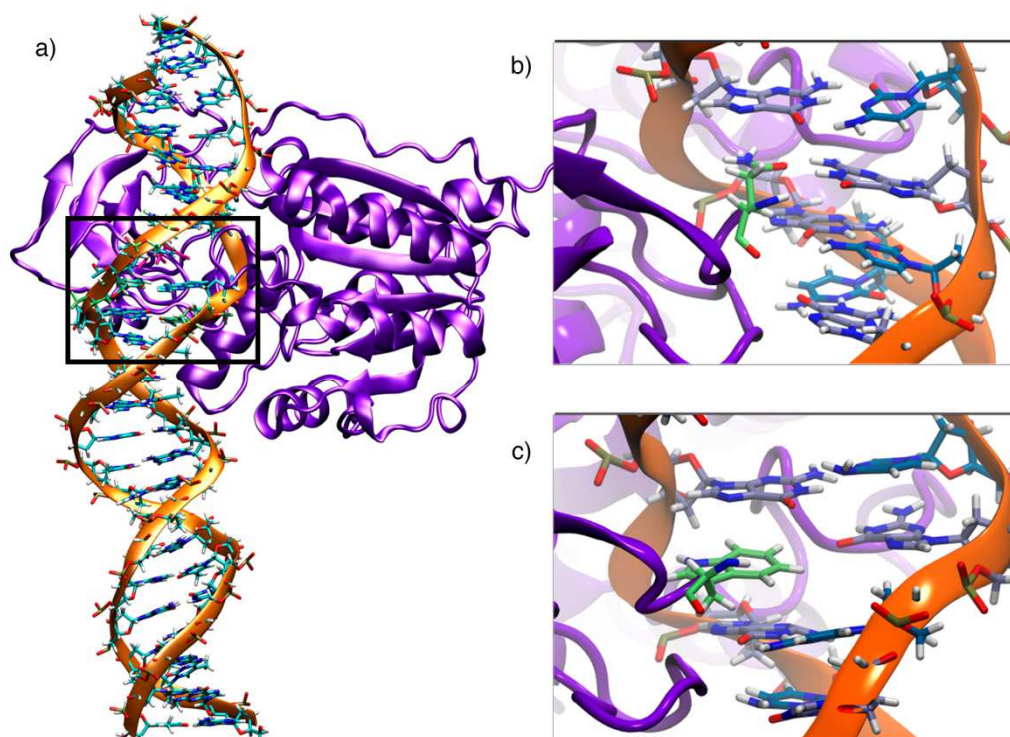
<sup>a</sup>Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain

<sup>b</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

<sup>c</sup>Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Campus Montilivi, 17071 Girona, Spain

*marinacorbella@ub.edu*

**Abstract-** Different mutagenic effects are generated by DNA oxidation that implies the formation of radical cation states (so-called holes) on purine nucleobases in the  $\pi$  stack. The interaction of DNA with proteins may protect DNA from the oxidative damage owing to hole transfer (HT) from the stack to aromatic amino acid residues. The HT dynamics in such systems are still poorly understood. Here, we report a computational study of HT in DNA complexes with methyltransferase HhaI and its mutant Q237W, which were experimentally investigated in the Barton group. We employ a combined approach based on molecular dynamics simulations and quantum mechanical calculations to estimate the rate for all individual steps involved in the HT pathways; finally the overall HT kinetics are explored using the Monte-Carlo method. Our results indicate that the HT characteristics are strongly affected by structural deformations of DNA upon its binding to the protein. In the wild-type enzyme complex, a Gln residue inserted in the DNA  $\pi$ -stack is shown to destabilize the radical cation states of neighboring guanines and thereby inhibits the long-range HT in line with experimental findings. In contrast, the HT is estimated to be quite fast in a complex of the Q237W mutant where Trp237 stabilizes hole states on the adjacent G bases and enhances the electronic coupling of these sites. An alternative HT pathway that implies the formation of a Trp<sup>+</sup> radical is predicted to be less efficient. Our study provides a consistent molecular picture on how long-range HT in DNA-protein complexes is controlled by amino acids closely interacting with the  $\pi$  stack.



# Macroscopic structures generated by microorganisms swimming in a fluid

Francisco Alarcón and Ignacio Pagonabarraga

Departament de Física Fonamental, Universitat de Barcelona, C. Martí Franqués 1, 08028-Barcelona, Spain  
falarcon@ffn.ub.es, ipagonabarraga@ub.edu

**Abstract**—It is known that active particles induce emerging patterns as a result of their dynamic interactions, giving rise to amazing collective motions, such as swarming or clustering. Here we present a systematic numerical study of self-propelling particles; our main goal is to characterize the collective behavior of suspensions of active particles as a result of the competition among their propulsion activity and the intensity of an attractive pair potential. Active particles are modeled using the squirmer model. Due to its hydrodynamic nature, we are able to classify the squirmer swimmer activity in terms of the stress it generates (referred to as pullers or pushers). We show that these active stresses play a central role in the emergence of collective motion. We have found that hydrodynamics drive the coherent swimming between swimmers while the swimmer direct interactions, modeled by a Lennard-Jones potential, contributes to the swimmers' cohesion. This competition gives rise to two different regimes where giant density fluctuations (GDF) emerge. These two regimes are differentiated by the suspension alignment; one regime has GDF in aligned suspensions whereas the other regime has GDF of suspensions with an isotropic orientated state. All the simulated squirmer suspensions shown in this study were characterized by a thorough analysis of global properties of the squirmer suspensions as well as a complementary cluster analysis.

## I. INTRODUCTION

Active matter refers generically to systems composed of self-driven units, active particles, each capable of converting stored or ambient free energy into systematic movement [1]. Examples of active systems are found at all length scales and could be classified in living and nonliving systems such as microorganisms, tissues and organisms, animal groups, self-propelled colloids and artificial nanoswimmers. Specifically, at the micro and nano scale we find an enormous range of interesting systems both biological and artificial; e.g. spermatozoa that fuse with the ovum during fertilization, the bacteria that inhabit our guts, the protozoa in our ponds, the algae in the ocean; these are but a few examples of a wide biological spectrum [2]. In the artificial world we have self-healing colloidal crystals and membranes as well as self-assembled microswimmers and robots [3]. Experiments in this field are now developing at a very rapid pace [4] and new theoretical ideas are needed to bring unity to the field and identify “universal” behavior in these internally driven systems.

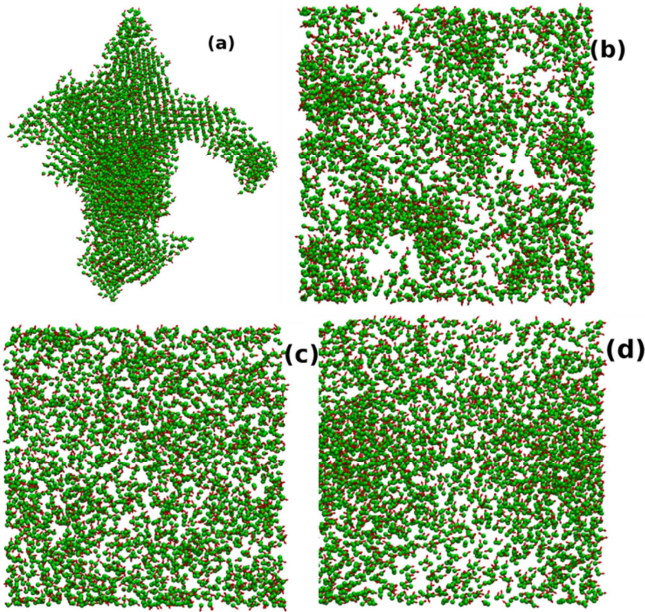
One important feature of active matter is that their elements can develop emergent, coordinated behavior; collective motion constitutes one of the most common and spectacular example [5]. Collective motion is ubiquitous and at every scale, from herds of large mammals to amoeba and bacteria colonies, down to the cooperative behavior of molecular motors in the cell. The behavior of large fish schools and the dance of starling flocks at dusk are among the most

spectacular examples. From a physical perspective collective motion emerges from a spontaneous symmetry breaking that allows for long-range orientational order [6]. The different mechanisms responsible for such symmetry breaking are still not completely understood. We have performed a systematic numerical study of interactive micro-swimmer suspensions building on the squirmer model, introduced by Lighthill [7]. Since the squirmer identifies systematically the hydrodynamic origin of self-propulsion and stress generation it provides a natural scheme to scrutinize the impact that the different features associated to self-propulsion in a liquid medium have in the collective dynamics of squirmer suspensions. In this abstract we describe the simulation scheme and how squirmers are modeled, then some of the main results are discussed and finally we conclude emphasizing the main implications of the results obtained.

## II. SIMULATION SCHEME

The squirmer model was proposed to model describe microorganisms that swim due to the synchronized motion of the cilia through metachronal waves. The minimal squirmer can be characterized only by two parameters  $B_1$  and  $B_2$ .  $B_1$  is related to the polarity associated with squirmer propulsion and  $B_2$  quantifies the active stresses that the metachronal waves induce in the fluid. To study the influence of activity we have taken the active stress relative to the polarity, quantified by  $\beta = B_2/B_1$ . In this way, swimmers can be classified as pullers if  $B_2 > 0$  or as pushers if  $B_2 < 0$ , which is related to how squirmer induces stress in the fluid. This classification applies generically to active microorganisms. In fact, many bacteria behave as pushers while certain motile algae such as *Chlamydomonas* behave as pullers [8]. Squirmers provide a useful model that identifies the basic hydrodynamic stresses associated to active swimming. In this respect, a previous study has shown the impact that active stresses have in the collective behavior of semi-dilute squirmer suspensions [9]. This study has shown that pushers decorrelate while pullers develop large aggregates for a finite range of positive  $\beta$ , when self-propulsion and stress generation competes with each other.

Fig. 1. Snapshots of simulations with  $\beta = 0.6$  at steady state of the simulations for several values of interaction strengths. The green spheres correspond to the squirmers positions while red arrow is the squirmer fixed orientation vector. In (a),  $\xi = 0.6$ , the attractive potential dominates and particles crystallize without any significant alignment. In (b),  $\xi = 1.8$ , a coexistence phase between small solid clusters, isotropically aligned, and individually swimming particles coexist. In (c),  $\xi = 4.2$ , the transition between isotropic and aligned squirmer structures appear. In (d),  $\xi = \infty$ , we recover the pure hydrodynamic swarming with a clear non-zero polar order.



We carry out this numerical study using the Lattice Boltzmann method (LBM), a well-known kinetic model that captures the hydrodynamic interactions among suspended particles [10]. LBM has a high degree of parallelization, hence it is ideal for parallel machines (computational clusters) [11]. We simulate our squirmers as spherical solid particles and modify locally the boundary condition at their surfaces to impose the active propulsion and stress.

We have carried out a thorough numerical study of semi-dilute squirmer suspensions where additionally to the active stress that characterize them, we introduce an isotropic Lennard-Jones (LJ) potential between the particles. We can then analyze the competition between the intrinsic propulsion and active stresses and the cohesive tendency of the LJ interactions to form dense packings. This competition will be quantified by the interaction strength parameter  $\zeta = F_d/F_{LJ}$ , where  $F_d = 6\pi\eta R_p v_s$ , is the drag force generated by one active particle of radius  $R_p$  that move with velocity  $v_s = 2/3 B_l$  in a fluid with viscosity  $\eta$  and  $F_{LJ} = F_{LJ}(r=r_0) = 2.4\epsilon_{LJ}/\sigma_{LJ}$  is the LJ-force at the minimum of the interaction ( $r_0$ ). Thus,  $\zeta$  is the ratio between squirmer propulsion force and the characteristic LJ force intensity. We can see in Fig. 1 some snapshots for a few representative scenarios: In (a)  $\zeta$  is very low and the cohesion is so strong that squirmers build up a big crystallized cluster. As the attraction decreases its strength, a coexisting liquid- solid phase like in (b) is allowed, a liquid phase like (c) develops in (c), and eventually active coherent swarming ( $\beta = 0.6$ ) are observed for  $\zeta = \infty$ , in (d).

### III. RESULTS

The collective properties of squirmer suspensions can be rationalized in terms of two order parameters. The suspension alignment can be quantified in terms of the global polar order, and the growth of squirmer density fluctuations with system size can be quantified in terms of its exponent. A scaling exponent closer to 1, corresponding to giant density fluctuations (GDF), correlates with swarm formation [12]

while systems in thermal equilibrium have an exponent of  $1/2$ . We find that the alignment depends completely on the squirmer hydrodynamic signature, and observe non-zero polar order for pullers with  $\beta < 3$ . The attractive potential plays a key role in the polar order only when its strength is comparable to squirmer activity. For larger values of  $\zeta$ , GDF correlate with clusters that display global orientational ordering. However, for smaller values of  $\zeta$ , around 1.8 we have found another GDF area related to a competition between the crystal formed when  $\zeta = 0.6$  and the isotropic fluid observed at  $\zeta > 1.8$ .

### IV. CONCLUSIONS AND REMARKS

We have analyzed the emerging behavior of squirmer suspensions. We cover all the cases at this density regime, and are able to know when collective motion emerges depending on the interaction and the hydrodynamic squirmer signature. It is important to remark that we find special density fluctuations at  $\zeta = 1.8$  where polar order is not present, which means that the origin of this GDF are totally different from the GDF found it for high values of  $\zeta$ . GDF can be related to a coexistence phase between solid and liquid, since we have a solid in  $\zeta = 0.6$  and a liquid in  $\zeta = 3.0$ . Although LBM has high degree of parallelization and since it is solving the velocity of the solvent explicitly, it requires of a lot of processors to run all the simulations shown above. We have found that the macroscopic polar order parameter of active suspensions decays inversely with the simulation box and reach a plateau at large enough simulation boxes. Hence, we have shown and quantified the intrinsic polar order in squirmer suspensions without any significant finite size effects.

### ACKNOWLEDGMENT

This work was financially supported by CONACYT-CECTI (México) through Grants No. 214151, MINECO through FIS2011- 22603, DURSI Project 2009SGR-634, and Generalitat de Catalunya under program Icrea Academia. Projects. This work was possible thank to the access to MareNostrum Supercomputer at Barcelona Supercomputing Center (BSC) and also through the Partnership for Advanced Computing in Europe (PRACE).

### REFERENCES

- [1] F. Schweitzer, *Brownian Agents and Active Particles*, Springer, Berlin, (2003).
- [2] E. Lauga and T.R. Powers, *Rep. Prog. Phys.* 72, 096601 (2009).
- [3] I.S. Aranson, *C.R. Physique* 14, 518 (2013).
- [4] A. Rabani, G. Ariel, A. Be'er, *PLoS ONE* 8, 12 (2013).
- [5] T. Vicsek and A. Zafeiris, *Phys. Rep.* 517, 71 (2012).
- [6] H. Chate et. al, *Eur. Phys. J. B* 64, 451 (2008).
- [7] M. J. Lighthill, *Communications on Pure and Applied Mathematics*, 5, 109 (1952).
- [8] I. Pagonabarraga, I. Llopis, *Soft Matter* 9, 7174 (2013).
- [9] F. Alarcón and I. Pagonabarraga, *Journal of Molecular Liquids* 185, 56 (2013).
- [10] S. Succi, *The Lattice Boltzmann equation for Fluid dynamics and Beyond*, Oxford University Press, (2001).
- [11] K. Stratford and I. Pagonabarraga, *Comput. Math. Appl.* 55, 1585 (2008).
- [12] S. Ramaswamy et. al, *EPL (Europhysics Letters)*, 69 196 (2003).

# Inter-residue interactions in membrane proteins

Eduardo Mayol<sup>1\*</sup>, Mireia Olivella<sup>2</sup>, Arnau Cordomí<sup>1</sup> and Leonardo Pardo<sup>1</sup>  
<sup>1</sup>Laboratori de Medicina Computacional. Unitat de Bioestadística, Facultat de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain.  
<sup>2</sup>Departament de Biologia de Sistemes, Universitat de Vic, Catalonia, Spain.  
\*Correspondence to: [luiseduardo.mayol@uab.cat](mailto:luiseduardo.mayol@uab.cat)

**Abstract-** Knowing the precise 3D-structure of a protein is crucial to understand its functional mechanism at the molecular level and to develop new pharmacological agents to targeting it. Nowadays only a few hundred integral membrane protein structures have been solved at high resolution due to the associated technical difficulties. In the present study we aim to characterize the main interactions in alpha and beta membrane proteins that are responsible of the maintenance of the overall structure. With this purpose, two non-redundant databases of alpha and beta transmembrane segments were constructed and analysed. The interactions that stabilize the structure of alpha and beta membrane proteins were quantified. The results reveal important differences in inter-residues interactions between alpha and beta membrane proteins. This novel structural information may be useful in predicting 3D models of proteins lacking structural information or in refining initial models of alpha and beta membrane proteins.

## I. INTRODUCTION

Membrane proteins (MPs) are located in the cell membrane, mediating the interaction between the cell and its surrounding. They include receptors, ion channels, transporters, and enzymes and are involved in multiple cellular processes. Membrane proteins constitute 20%–30% of human genes <sup>[1]</sup> and represent the targets of over half of known drugs targets <sup>[2-5]</sup>. With the explosive growing of sequence information that results of massive parallel sequence technology, the gap between sequences and protein 3D structures is still widening. Since experimental structure determination of MPs is such a major endeavor, computational approaches that predict 3D structures of membrane proteins are a valuable tool to complement existing experimental data. The fact that hydrophilic environment conditions the structure and features of membrane proteins, implies that water-exposed regions of membrane proteins will differ from the membrane-embedded ones. Thus, to study specific structural features of the transmembrane (TM) proteins, it is necessary to distinguish the membrane embedded from the water exposed regions. Several algorithms have been developed to identify the transmembrane spanning regions of membrane proteins <sup>[6-7]</sup>. Inter-residue interactions have been one of the main focuses to understand the mechanisms of protein folding and stability. Consequently, many methods have been described to explore the amino acid content of a protein and their inter-residue interactions for a variety of goals <sup>[8-14]</sup>. These studies mainly rely on globular proteins, as these proteins are overrepresented in the Protein Data Bank <sup>[15]</sup>. The aim of this study is to characterize the inter-residue interactions in membrane proteins that are responsible for maintaining the overall fold. The difference in the nature of interactions for alpha and beta membrane proteins provide clues to the different characteristics of these two types of membrane proteins and its role in stabilization of structures. The quantification of these

interactions can be a valuable tool to refine initial molecular models.

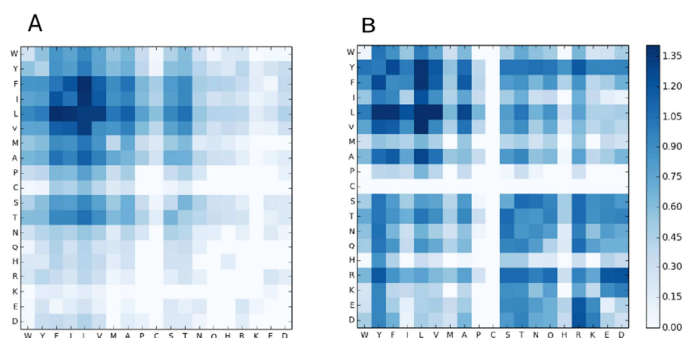
## II. RESULTS AND DISCUSSION

### A. Residue interactions in alpha and beta membrane proteins

The interactions involving hydrophobic amino acids represent the majority of the residue-residue contacts in alpha membrane proteins: L-F, I-F, L-V, L-L, followed by F-I, I-V, F-V, A-L, T-L and F-A (Figure 1-A). Interactions involving polar (S and T) and hydrophobic amino acids also show a high frequency. The frequency of contacts between polar and/or charged amino acids is very low. These results confirm that hydrophobic interactions are the most important contacts that stabilize alpha helical membrane proteins.

In beta membrane proteins, the interactions are spread among a wider type of residue interactions (Figure 1-B). The most prevalent interaction is L-L, followed by L-V, F-L and Y-L and then L-A, L-I, and F-Y. Contacts between R and D, E and Y are also found with a high frequency. Polar amino acids interact with hydrophobic ones, like in alpha membrane proteins, but they also interact with charged and polar amino acids with a higher frequency. Contacts between charged residues are also frequently present. On the other hand, R and Y do not show marked preferences for specific amino acids.

Figure 1. Inter-residue interactions matrix. A) alpha membrane proteins and B) beta membrane proteins



## III. CONCLUSIONS

- Alpha membrane proteins concentrate residue contacts in one type of interactions, while in beta membrane protein, substantial inter-residue contacts are distributed in a wider type of interactions.

- Hydrophobic interactions appear to be the most prevalent interaction in alpha and beta membrane proteins, although some differences in the prevalence of these interactions are observed. Polar or charged inter-residues interactions have an important role in beta membrane proteins.
- The analysis of inter-residue contacts can be a valuable tool for the prediction of structural models for transmembrane alpha and beta proteins.

## REFERENCES

- [1] Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M., & Berglund, L. (2010). Prediction of the human membrane proteome. *Proteomics*, 10(6), 1141-1149.
- [2] Arinaminpathy, Yalini, Ekta Khurana, Donald M. Engelman, and Mark B. Gerstein. "Computational Analysis of Membrane Proteins: The Largest Class of Drug Targets." *Drug Discovery Today* 14, no. 23 (2009): 1130-1135.
- [3] Hopkins, Andrew L. and Colin R. Groom. "The Druggable Genome." *Nature Reviews Drug Discovery* 1, no. 9 (2002): 727-730.
- [4] Overington, John P., Bissan Al-Lazikani, and Andrew L. Hopkins. "How Many Drug Targets are there?" *Nature Reviews Drug Discovery* 5, no. 12 (2006): 993-996.
- [5] Yıldırım, M. A., Goh, K. I., Cusick, M. E., Barabási, A. L., & Vidal, M. (2007). Drug—target network. *Nature biotechnology*, 25(10), 1119-1126.
- [6] Tusnády, Gábor E., Zsuzsanna Dosztányi, and István Simon. "TMDET: Web Server for Detecting Transmembrane Regions of Proteins by using their 3D Coordinates." *Bioinformatics* 21, no. 7 (2005): 1276-1277.
- [7] Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: Orientations of Proteins in Membranes database. *Bioinformatics* 22, 623-625.
- [8] Bahar, I., Kaplan, M., & Jernigan, R. L. (1997). Short-range conformational energies, secondary structure propensities, and recognition of correct sequence-structure matches. *Proteins Structure Function and Genetics*, 29(3), 292-308.
- [9] Faure, G., Bornot, A., & de Brevern, A. G. (2008). Protein contacts, inter-residue interactions and side-chain modelling. *Biochimie*, 90(4), 626-639.
- [10] Gromiha, M. M., & Selvaraj, S. (2004). Inter-residue interactions in protein folding and stability. *Progress in biophysics and molecular biology*, 86(2), 235-277.
- [11] Miyazawa, S., & Jernigan, R. L. (1996). Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *Journal of molecular biology*, 256(3), 623-644.
- [12] Punta, M., & Rost, B. (2005). PROFcon: novel prediction of long-range contacts. *Bioinformatics*, 21(13), 2960-2968.
- [13] Selvaraj, S., & Gromiha, M. M. (1998). Importance of long-range interactions in ( $\alpha/\beta$ ) 8 barrel fold. *Journal of protein chemistry*, 17(7), 691-697.
- [14] Meysman, P., Zhou, C., Cule, B., Goethals, B., & Laukens, K. (2015). Mining the entire Protein DataBank for frequent spatially cohesive amino acid patterns. *BioData mining*, 8(1), 4.
- [15] Berman, Helen M., et al. "The protein data bank." *Nucleic acids research* 28.1 (2000): 235-242.



# A novel approach to reconstruct the plinian and co-ignimbrite phases of large eruptions - Campanian Ignimbrite

Alejandro Marti (\*)(1), Arnau Folch (1), Antonio Costa (2), and Samantha Engwell (3)

(1) CASE Department, Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain

(2) Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Italy

(3) Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Italy

(\*).Alejandro.Marti@bsc.es

**Abstract-** Reconstructing the volume and tephra dispersal from volcanic super-eruptions is necessary to assess the widespread impact of these massive events on climate, ecosystems and humans. Recent studies have demonstrated that volcanic ash transport and dispersion models are unrivaled in accurately constraining the volume of material ejected and provide further insight about the eruption dynamics during these gigantic events. However, the conventional simplified characterization of caldera-forming super-eruptions as a single-phase event can lead to inaccurate estimations of the eruption dynamics and its impacts. Here, we apply a novel computational inversion method to reconstruct, for the first time, the two phases of the largest eruption of the last 200 ky in Europe, the Campanian Ignimbrite (CI) super-eruption. Additionally, we discuss the eruption's contribution to the Middle to Upper Palaeolithic transition by evaluating its environmental and climate implications.

## I. INTRODUCTION

Volcanic super-eruptions, typically associated with caldera-forming events, are often multiphase. The typical scenario begins with a Plinian column (Fig. 1a) which destabilizes to produce collapsing fountains that shed pyroclastic flows spreading laterally along the ground, eventually leading to secondary, co-ignimbritic plumes (Fig. 1b). Source conditions for co-ignimbrite plumes vary considerably from those of Plinian plumes, with much larger source radii and lower initial velocities. In eruptions where both phases have occurred, tephra deposits are typically bimodal. Such deposits are commonly separated into constituent phases based on their grain size characteristics, with the coarse mode attributed to the Plinian phase, and the fine mode attributed to the co-ignimbrite phase<sup>2</sup>. In order to correctly evaluate the magnitude of each eruptive phase, it is critical to constrain their eruption dynamics and quantify their Eruption Source Parameters (ESPs) independently.

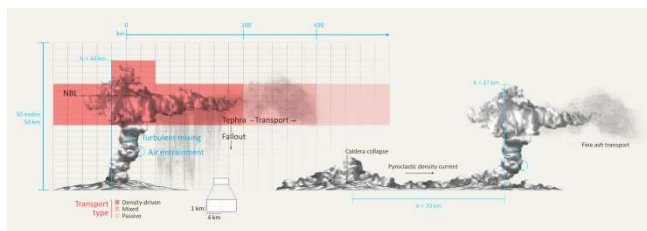


Figure 1. . Illustration of a typical multi-phase super-eruption with an initial (a) Sustained Plinian phase; followed by a (b) Caldera-collapse and large pyroclastic density currents eventually leading to secondary, co-ignimbritic plumes offset from the vent. Nodes with grey outline represent the model domain used for the CI model simulations. Our simulations used a grid spacing 0.2 degrees latitude and longitude (~4km), and 1 km cell height. Red nodes illustrate the effect of the density-current and the corresponding transport regimes in the umbrella cloud based on the best-fit model parameters (blue text) associated to the CI event.

Here, we apply a novel computational approach to reconstruct, for the first time, the two phases of the largest eruption of the last 200 ky in Europe, the Campanian Ignimbrite (CI) super-eruption<sup>3</sup>. It has been debated if the CI eruption, dated approximately 39.000 calendar years ago, was related to the disappearance of the remaining Neanderthals in Europe and, its interference with the territorial expansion of anatomically modern humans in the more southern parts of western Eurasia<sup>4</sup>. The event coincided with the onset of an extremely cold climatic phase (Heinrich Event 4) and the attendant episode of the Freno-Scandinavian ice cap and peripheral tundra on land. To conclude this work, we evaluate the environmental and climate-forcing implications associated to the eruption to provide new insights of the eruption's contribution to the Middle to Upper Palaeolithic transition.

The aims of this conference paper are:

- To quantify the volume and tephra dispersal across Europe and Mediterranean area during each phase of the event.
- To describe the density-current effects in the umbrella cloud
- To discuss the eruption's contribution to the Middle to Upper Palaeolithic transition
- To provide an outreach interactive website for more information on this work ([http://www.bsc.es/viz/campanian\\_ignimbrite](http://www.bsc.es/viz/campanian_ignimbrite))

## II. METHODOLOGY

To reconstruct the tephra dispersal and the ESPs of the CI eruption, we used the FALL3D tephra dispersal model<sup>5</sup> in conjunction with a downhill simplex inversion method (DSM) to infer values of ESPs such as erupted mass, mass flow rate, plume height and total grain-size-distribution (TGSD). We compare the tephra dispersal and volume results from reconstructing the CI eruption as a two-phase event (Method 1), with those from the classical single-phase approach<sup>6</sup> (Method 2). Optimal inversion values for the ESPs were obtained by best-fitting two independent datasets containing tephra deposits separated into constituent phases. To account for the gravitational cloud spreading associated to the CI eruption, we couple FALL3D with a model that accounts for the density-driven transport in the umbrella cloud<sup>7</sup>.

## III. RESULTS

### A. Method 1: two-phase approach

Results for the best-fit model parameters for Method 1 suggest that the eruption began with a short, high-intensity ultra-Plinian explosive phase that yielded a column height of

44 km and a mass eruption rate (MER) of  $3.75 \times 10^9$  kg/s. The Plinian phase lasted for 4 h, depositing a total of  $54 \text{ km}^3$  ( $\sim 22 \text{ km}^3$  DRE) of fallout material. After that time, the column would have collapsed to form radially spreading pyroclastic density currents, thereby superseding the Plinian column phase and leading to a co-ignimbrite phase. The co-ignimbrite column would have reached 37 km in height, fed by an average MER of  $2.25 \times 10^9$  kg/s over 19 h, depositing  $\sim 154 \text{ km}^3$  ( $\sim 62 \text{ km}^3$  DRE) of fallout material. Assuming that  $\sim 35$ – $40\%$  of the erupted material was elutriated from PDC<sup>8</sup> the total MER would have reached up to  $\sim 5$ – $6 \times 10^9$  kg/s.

### B. Method 2: single-phase approach

Best-fit simulation results from Method 2 suggest a column height of 38 km, a mean MER of  $2.55 \times 10^9$  kg/s and duration of 23 hours. The amount of material deposited as tephra fallout totaled  $\sim 211 \text{ km}^3$  ( $\sim 84 \text{ km}^3$  DRE).

### C. Method 1 vs. Method 2

Method 1 is more accurate than the classical single-phase approach used in Method 2 (correlation coefficients 0.96 vs. 0.74) and in previous studies<sup>6</sup>. Both methods confirm the dominant role of the co-ignimbrite phase in the total bulk volume of the eruption. The amount of material deposited as fallout amongst experiments differed in less than 1.5%, a very consistent result. Method 1 simulated the dispersal of tephra (0.5 cm thick or larger) to be 15% smaller than Method 2, and 20% smaller than previous studies<sup>6</sup>.

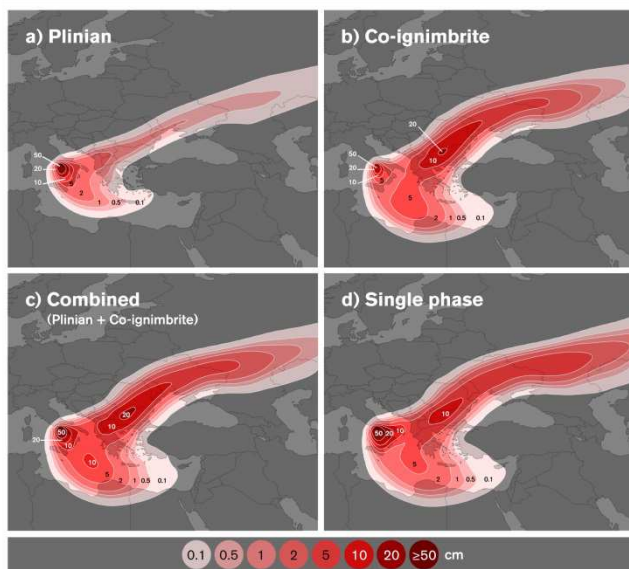


Figure 2. Model isopach maps showing the best-fit FALL3D tephra dispersal deposit (cm) for each phase of the CI super-eruption. Tephra dispersal is shown for (a) the 44km high ultra-plinian plume accumulating a total of  $54 \text{ km}^3$  of fallout material; (b) the 38 km co-ignimbrite plume accounting for 74% of the total fallout material; (c) the combined Plinian and co-ignimbrite phases accumulating a total of  $208 \text{ km}^3$  of fallout material over  $\sim 3$  million  $\text{km}^2$ ; (d) the classical single-phase approach for an eruption duration of 23h.

### D. Density-current effects

We find the density-driven transport to be dominant for the first hour with an effective spreading radial velocity of  $\sim 130$

m/s and producing an umbrella cloud radius of  $\sim 100$  km. Model results show the effect of the density-driven transport to be significant in proximal areas, where tephra deposition is 1.5–2 times higher NE from the source and up to 50% lower in the east Mediterranean region due to the wind direction. These values are lower than expected for an eruption of such a magnitude. This is due to the strong stratospheric wind velocity of  $\sim 90$  m/s found at the transport height above the vent.

## IV. CONCLUSIONS

We bring to light valuable new results and methods in reconstructing the 39 ky CI super-eruption, accounting for both Plinian and co-ignimbrite phases. Simulation results show tephra fallout predominantly originated from the co-ignimbrite clouds. Density-current processes would have been significant in proximal areas. However, tephra transport would have been primarily dominated by wind advection, most likely as a result of intense stratospheric winds corresponding to best-fit synoptic conditions. Ecosystem recovery would have required tens to thousands of years depending on the thickness, nature and distance from source of the tephra deposits.

We suggest that the eruption would have cause a demographic crash among modern human populations boosted by the impact of the synchronous Heinrich Event 4 and the attendant episode of glacial advance on land. Such crash would have favoured the persistence of the Neanderthals in the Iberian Peninsula and not their disappearance as previously thought.

## ACKNOWLEDGMENT

A.M. thanks the CASE visualization team at the BSC for their contribution to this work.

## REFERENCES

- Sparks, R. S. J., Wilson, L. & Hulme, G. Theoretical modeling of the generation, movement, and emplacement of pyroclastic flows by column collapse. *J. Geophys. Res. Solid Earth* 83, 1727–1739 (1978).
- Engwell, S. L., Sparks, R. S. J. & Carey, S. Physical characteristics of tephra layers in the deep sea realm: the Campanian Ignimbrite eruption. *Geol. Soc. London, Spec. Publ.* 398, 47–64 (2014).
- Fedele, F., Giaccio, B., Isala, R. & Orsi, G. The Campanian ignimbrite eruption, Heinrich event 4, and Palaeolithic change in Europe: a high-resolution investigation. In *Volcanism and the Earth's Atmosphere* 139, 301–325 (American Geophysical Union, 2003).
- Zilhão, J. Neandertals and moderns mixed, and it matters. *Evol. Anthropol.* 15, 183–195 (2006).
- Costa, A., Macedonio, G. & Folch, A. A three-dimensional Eulerian model for transport and deposition of volcanic ashes. *Earth Planet. Sci. Lett.* 241, 634–647 (2006).
- Costa, A. et al. Quantifying volcanic ash dispersal and impact of the Campanian Ignimbrite super-eruption. *Geophys. Res. Lett.* 39, L10310 (2012).
- Costa, A., Folch, A. & Macedonio, G. Density-driven transport in the umbrella region of volcanic clouds: Implications for tephra dispersion models. *Geophys. Res. Lett.* 40, 4823–4827 (2013).
- Sparks, R. S. J. & Walker, G. P. L. The significance of vitric-enriched air-fall ashes associated with crystal-enriched ignimbrites. *J. Volcanol. Geotherm. Res.* 2, 329–341 (1977).

# Plasma Physics Code Contribution to the Mont-Blanc Project

Xavier Sáez

Barcelona Supercomputing Center (BSC-CNS), Spain  
*xavier.saez@bsc.es*

Alejandro Soba

Centro de Simulación Computacional para Aplicaciones Tecnológicas (CSC-CONICET), Argentina  
*soba@cnea.gov.ar*

Mervi Mantsinen

Barcelona Supercomputing Center (BSC-CNS), Spain  
Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain  
*mervi.mantsinen@bsc.es*

**Abstract**-This work develops strategies for adapting a particle-in-cell code to heterogeneous computer architectures and, in particular, to an ARM-based prototype of the Mont-Blanc project using OmpSs programming model and the OpenMP and OpenCL languages.

- **pull**: the particle properties are interpolated to neighboring points in the computational mesh.
- **solve**: the moment equations are solved on the mesh.
- **push**: the momentum of each particle is calculated by interpolation on the mesh. The particle properties are updated.

## I. INTRODUCTION

The scientific grand challenges, such as fusion reactors, have been the driving force for the evolution of High-Performance Computing (HPC).

During the last two decades, the supercomputers have grown rapidly in performance to provide scientists the required computing power, at the cost of a similar growth in power consumption.

However, nowadays the computer's performance is limited by the power consumption and power density, so new developed platforms to construct a future sustainable exaflop supercomputer will have to be based on the power efficiency. The Mont-Blanc project appeared with the aim to design computer architectures capable of delivering an exascale performance using 15 to 30 time less energy than present architectures [1].

Particle-in-cell (PIC) is one of the most used methods in plasma physics simulations [2]. The quality of results achieved by this method relies on tracking a very large number of particles. Therefore, PIC codes (such as EUTERPE) are good candidates to be adapted to new HPC platforms since they require intensive computation.

## II. PARTICLE-IN-CELL CODE

### A. Particle-in-cell Methods

PIC methods are used to model physical systems whose behavior varies over different ranges of spatial scales. The individual particles are tracked in a continuous phase space, whereas densities and the current are computed concurrently on stationary mesh points.

A PIC algorithm can be summarized in three steps (Fig.1) repeated at each time step [3]:

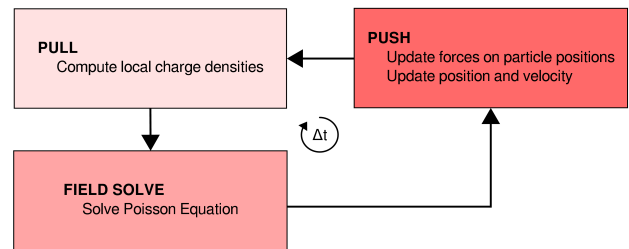


Fig. 1. The steps of a PIC algorithm.

### B. EUTERPE Code

EUTERPE is a gyrokinetic PIC code for global linear and non-linear simulations of fusion plasma instabilities in three-dimensional geometries, in particular in tokamaks and stellarators [4, 5].

It has been written to target traditional HPC clusters using MPI and a domain cloning technique to increase the number of processors without boosting the interprocessor communications to prohibitive levels [6].

## III. COMPUTER PROTOTYPE

Mont-Blanc is a European exascale computing approach to develop a full energy-efficient HPC prototype. The project is coordinated by Barcelona Supercomputing Center (BSC) since October 2011.

The aim is reducing energy consumption using low-power commercially available processors that were designed for mobile and embedded systems. In this way, we exploit their cheapness due to the large volume of these platforms and their high accessibility in the commodity market.

The prototype used in this work is based on a system-on-chip (SoC) Samsung Exynos 5 which contains an ARM Cortex-A15 dual core and an ARM Mali T604 GPU [7].

This embedded SoC with integrated GPU accelerator is the first that could be used for HPC, since it supports 64-bit floating point arithmetic and provides support for parallel programming languages (such as OpenCL 1.1).

#### IV. ACHIEVEMENTS

We have developed three versions of the EUTERPE code for this prototype:

##### A. Hybrid version

EUTERPE was only parallelized at task level using MPI. For that reason, we developed a hybrid version of the code introducing OpenMP to take advantage of all the levels of parallelism that a multi-core architecture can offer [8].

##### B. OpenCL + Hybrid version

In order to obtain the best possible results on the prototype, all the available resources were used by the application: the dual core CPU (using previous hybrid version) and the GPU (using a new OpenCL version).

We only wrote the most time-consuming routines to OpenCL with the aim to minimize the necessary changes. In the *push* part, one work-item was assigned to each particle. The *pull* part was more challenging to implement, since different particles could contribute to the charge density on the same mesh point. To avoid these memory conflicts, a mesh copy was created per work-group, so the lock contention was far minor.

##### C. OmpSs version

When we port a code to a new platform not only is important the possibility to reach the maximum performance of the platform, but also the ease of programming it. The main drawback of OpenCL is its low programmability because of it is a low-level programming language.

To address this shortcoming, we tested OmpSs, a task-based programming model developed by Barcelona Supercomputing Center (BSC) [9]. It provides an abstraction to the user thereby reducing programmer effort and unifies the SMP, heterogeneous and cluster programming in one model.

Although OmpSs version is a bit slower than the hybrid version, it is a simpler version and its productivity has improved considerably (see Table I).

#### V. CONCLUSIONS

This work confirmed that is possible to port this kind of plasma physics codes to an ARM-based platform and we can say that OmpSs simplifies the porting of codes to this new platform.

#### REFERENCES

- [1] Mont-Blanc project: European approach towards energy efficient high performance. <http://www.montblanc-project.eu>
- [2] C.K. Birdsall and A.B. Langdon. *Plasma Physics via Computer Simulation*. Institution of Physics, Bristol, 1991.
- [3] E. Akarsu et al. "Particle-in-cell Simulation Codes in High Performance Fortran", *Proceedings of the 1996 ACM/IEEE Conf. on Supercomputing (IEEE Computer Society)*, 38, 1996.
- [4] V. Kornilov, R. Kleiber, R. Hatzky, et al., Gyrokinetic global three-dimensional simulations of linear ion-temperature-gradient modes in Wendelstein 7-X", *Physics of Plasmas*, 11, 2004.
- [5] E. Sánchez, R. Kleiber, R. Hatzky, et al., Linear and non-linear simulations using the EUTERPE gyrokinetic code, *IEEE Transaction on Plasma Science* 38, 9, 2010.
- [6] R. Hatzky, "Domain Cloning for a Particle-in-Cell (PIC) Code on a Cluster of Symmetric-Multiprocessor (SMP) Computers". *Parallel Computing*, 32, 2006.
- [7] ARM: The architecture for the Digital World. <http://www.arm.com>
- [8] X. Sáez, A. Soba, E. Sánchez et al. "Particle-In-Cell algorithms for Plasma simulations on heterogeneous architectures", *Proc. of the 19th Euromicro Conf. on Parallel, Distributed and Network-based Processing (IEEE)*, 385-389, 2011.
- [9] A. Duran, E. Ayguadé, R.M. Badia, J. Labarta et al. "OmpSs: A proposal for programming heterogeneous multi-core architectures". *Parallel Processing Letters*, 21, 2, 173-193, 2011.

TABLE I  
COMPARISON BETWEEN THE DIFFERENT VERSIONS

Routine	Performance - Best time (s)		Programmability / Productivity	
	Hybrid version + OpenCL	OmpSs version	OpenCL API calls	OmpSs directives
Push	6.02	6.92	161	12
Pull	10.31	10.83	167	18

# Folding: reporting instantaneous performance metrics and source-code references

Harald Servat, Jesús Labarta  
BSC – Computer Sciences Department  
UPC-DAC – Computer Architecture Department

**Abstract-** Despite supercomputers deliver huge computational power, applications only reach a fraction of it. There are several factors limiting the application performance, and one of the most important is the single processor efficiency because it ultimately dictates the overall achieved performance. We present the folding mechanism, a process that combines measurements captured through minimal instrumentation and coarse-grain sampling ensuring low time dilation (less than 5%). The mechanism reports instantaneous performance and source-code references for optimized binaries accurately by taking advantage of the repetitiveness of many applications, especially in HPC. The mechanism enables the exploration of the application performance and guides the analyst to source-code modifications.

## I. MOTIVATION

Nowadays, supercomputers deliver an enormous amount of computational power; however, it is well-known that applications only reach a fraction of it. There are several factors that limit the achieved application performance. One of the most important factors is the single processor performance (*i.e.* how fast a processor executes a work unit) because it ultimately dictates the overall achieved performance. Performance analysis tools are pieces of software that help locating performance inefficiencies and identifying their nature within applications to ultimately improve the application performance. These tools rely on two collection techniques to invoke their performance monitors: instrumentation and sampling. Instrumentation refers to the ability to inject performance monitors into concrete application locations whereas sampling deals with invoking the installed monitors periodically. Each technique has its advantages. The measurements obtained through instrumentation are directly associated with the application structure while sampling allows a simple way to determine the volume of the measurements captured. In any case, the granularity of the measurements provides valuable insight cannot be easily determined *a priori*. Should analysts study the performance of an application for the first time, they may consider using a performance tool and instrument every routine or use high-frequency sampling rates in order to provide the most detailed results. More often than not, these approaches lead to large overheads that impact on the application performance and thus alter the measurements gathered and, therefore, mislead the analyst.

## II. FOLDING

The folding mechanism overcomes the overhead by taking advantage of the repetitiveness of many applications. This mechanism smartly combines instrumented and inexpensive coarse-grain sampled information that dilates the application

runtime less than 5% on optimized binaries. The results of the mechanisms include rich reports that show the instantaneous performance evolution and source-code progression within instrumented regions of code accurately. To this end, the captured metrics should contain information regarding performance metrics associated to the processor such as number of instructions executed, number of L1 cache misses, and stalled cycles, as well as, call-stack information to correlate with the application structure.

The folding projects the collected samples into a synthetic instance preserving their time since the start of their respective instance; thus, a sample fired at time  $T_s$  within an instance that starts at  $T_i$  gets mapped into the representative region at time  $\delta_s$ , where  $\delta_s = T_s - T_i$ . Figure 2 provides a visual description of the process in order to help understand it. The top of the Figure depicts a time-line of an application with a repetitive region (e.g. a routine, a loop body, ...) that has been executed three times during the whole execution whereas the bottom part schematizes how the folding works. The analyst instruments the application to determine when each routine invocation begins (shown as  $I_x$ , where  $x=\{1,2,3\}$ ) and ends (shown as  $I'_x$ , where  $x=\{1,2,3\}$ ). The analyst also enables a periodic sampling mechanism that freely runs no matter the application activity (in the example, providing measurements every seven units of time). As a consequence of using the folding mechanism, the synthetic instance contains more samples than any other instance, and therefore, it is capable of depicting the progression within the instance more accurately.

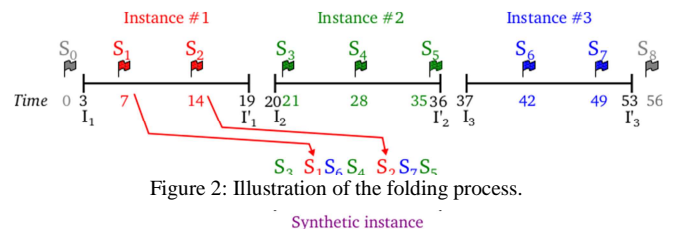


Figure 2: Illustration of the folding process.

### A. Detailed performance counter evolution

The folding depicts instantaneous progression of the performance counters using a fitting mechanism on the performance metrics associated to the folded samples. We have evaluated two fitting algorithms used in different areas to report continuous metrics: a Gaussian interpolation process named Kriging [1], and piece-wise linear regressions. No matter the approach taken, the folding results include the

progression of the performance counter rates. From all these counter rates, the MIPS rate indicates whether a region of code runs at good pace but it does not unveil the reasons for performance bottlenecks, if any. Our work has addressed this topic by taking advantage of performance analytical models based on performance counters and applying them to the folding results. For instance, **Figure 3** shows the progression of the MIPS rate (in black, on the right Y-axis) and other performance counter ratios per instruction (in other colors, on the left Y-axis) in the most time-consuming region of CGPOP [2]. The plot exposes two phases within this region. The first phase runs approx. at 500 MIPS and seems to be limited by the amount of cache misses per instruction (approx. 6%). The second phase starts running at 2000 MIPS after decreasing the cache miss ratio from the first phase. The subsequent decrease of L3 cache misses turns into an increase to the instruction rate up to 3500 MIPS.

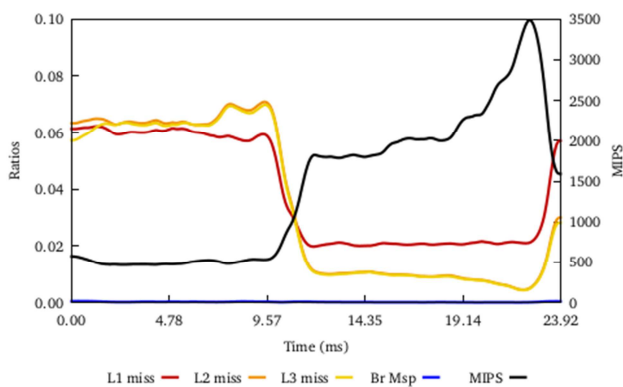


Figure 3: Folding results for the most time-consuming region in the CGPOP application.

### B. Detailed correlation with source-code

The correlation of performance inefficiencies and their associated source code is a cornerstone to understand why the efficiency of an application falls behind the computer’s peak performance and to enable optimizations on the application code, ultimately. We have explored the opportunities of the folding mechanism when treating source code references captured in the samples to allow the analyst easily understand the application’s bottlenecks. Despite source code references do not benefit from fitting models as the performance counters, our work has also explored two approaches to establish an approximate correlation between performance and source code. One of these approaches takes benefit of the phases delimited by the piece-wise linear regressions when fitting the performance counters, and associates source code references to each phase. The other approach has been inspired by MultiSequence Alignment techniques [3,4] and finds similarities in consecutive sampled call-stacks to unveil the active routine.

Figure 4 exemplifies the results in which the performance results are collocated with their source-code when exploring one of the most time-consuming regions in Arts\_CF [5]. This region exposes two phases in terms of the instruction rate and each of these phases is correlated with a particular region of code (shown in a different background color). It is worth to notice that the first phase (shown in red), reveals that the routine `scalar_weno5_coeff` (line 667) runs approx. at 2600 MIPS, which is far below the processor’s peak performance. Despite we did not have access to the application source-code, we asked the developer for the surrounding lines and after exploring them we were able to suggest optimizations that improved the performance of the region by 17.4% by applying well-known optimization techniques to tiny regions of code.

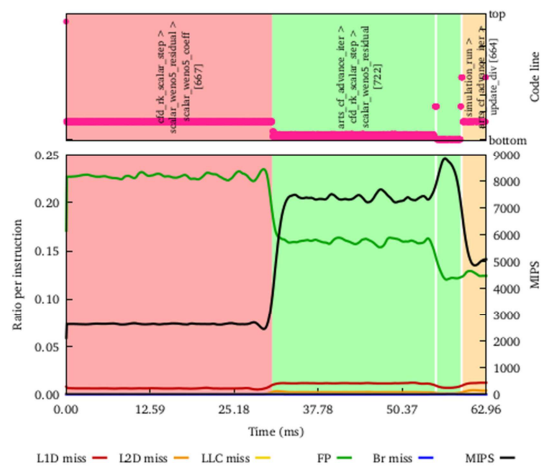


Figure 4: Results of the Arts\_CF application showing the source-code collocated with the performance results.

### CONCLUSIONS

We have presented a mechanism that reports the instantaneous progression of the performance metrics and source code references accurately for already optimized binaries. The mechanism is capable of reporting the performance progression of very small delimited regions of code by using coarse-grain sampling, ensuring low intrusion during the application execution. Despite not every application may benefit from manually changing the source-code, the mechanism has proven valuable to guide the analyst to small code changes based on well-known optimization techniques to further increase the performance.

### REFERENCES

- [1] F. Trochu. “A contouring program based on dual kriging implementation.” *Engineering with computers* 9.4 (1,993), pp. 160-177.
- [2] Andrew Stone, John Dennis and Michelle Mills-Strout. “The CGPOP miniapp, v1.0” Tech report CS-11-103- Colorado State University 2,011.
- [3] Cedric Notredame, Desmond G. Higgins and Jaap Hering. “T-Coffee: a novel method for fast and accurate Multiple-Sequence Alignment” *Journal of molecular biology* 302.1 (2,000) pp 205-217.
- [4] Andreas Doring, David Weese, Tobias Rausch and Knut Reiner. “SeqAn an efficient, generic C++ library for sequence analysis” *BMC bioinformatics* 9.1 (2,009) pp 11.

[5] Olivier Desjardins, Guillaume Blanquart, Guillaume Balarac, and Heinz Pitsch. "High order conservative finite difference scheme for variable density low Mach number turbulent flows" *Journal of Computational Physics* 227.15 (2,008), pp-7125-7159.

#### LIST OF PUBLICATIONS

Harald Servat, Germán Llort, Judit Giménez, Jesús Labarta "Detailed Performance analysis using Coarse Grain Sampling" In proceeding of the Workshop on Productivity and Performance (PROPER), in conjunction with Euro-PAR 2,009: 185-198.

Harald Servat, Germán Llort, Judit Giménez, Kevin A. Huck, Jesús Labarta "Unveiling Internal Evolution of Parallel Application Computation Phases" In proceedings of International Conference on Parallel Processing (ICPP) 2,011: 155-164.

Harald Servat, Germán Llort, Judit Giménez, Kevin A. Huck, Jesús Labarta. "Folding: Detailed Analysis with Coarse Sampling" In proceedings of the Parallel Tools Workshop 2,011: 105-118.

Harald Servat, Germán Llort, Judit Giménez, Kevin A. Huck, Jesús Labarta. "Framework for a productive performance optimization" In *Parallel Computing Journal*, 39(8) 2,013: 336-353.

Harald Servat, Germán Llort, Juan González, Judit Giménez, Jesús Labarta. "Identifying code phases using piece-wise linear regressions" In proceedings of the 28<sup>th</sup> IEEE International Parallel & Distributed Processing Symposium (IPDPS) 2,014: 941-951.

Harald Servat, Germán Llort, Juan González, Judit Giménez, Jesús Labarta: "Bio-inspired call-stack reconstruction for performance analysis" Technical report UPC-DAC-RR-2014-20.

# Probabilistically Time-Analyzable Complex Processors in Hard Real-Time systems

Mladen Slijepcevic<sup>1,2</sup>, Jaume Abella<sup>2</sup>, Francisco J. Cazorla<sup>2,3</sup>

<sup>1</sup>Universitat Politècnica de Catalunya (UPC),

<sup>2</sup>Barcelona Supercomputing Center (BSC-CNS),

<sup>3</sup>Spanish National Research Council (IIIA-CSIC)

*mladen.slijepcevic@bsc.es*

**Abstract**-Critical Real-Time Embedded Systems (CRTES) feature performance-demanding functionality. High-performance hardware and complex software can provide such functionality, but the use of aggressive technology challenges time-predictability. Our work focuses on the investigation and development of (1) hardware mechanisms to control inter-task interferences in shared time-randomized caches and (2) manycore network-on-chip designs meeting the requirements of Probabilistic Timing Analysis (PTA).

## I. INTRODUCTION

Industry developing CRTES, such as Aerospace, Space, Automotive, and Railways, face relentless demands for increased processor performance to support advanced new functionalities. Multiple indicators suggest that these demands will continue to grow across almost all sectors of the CRTES industry. Multicores are well accepted as one of the main design paradigms to increase performance. However, current generation CRTES, based on relatively simple single-core processors, are already extremely difficult to analyse in terms of their temporal behaviour. The advent of multicore and manycore platforms exacerbates this problem, rendering traditional temporal analysis techniques unable to scale and ineffectual, with potentially dire consequences for the quality and reliability of future products. In this context multicores for CRTES must balance the achievement of trustworthy and low Worst-Case Execution Time (WCET) estimates, high performance and low design complexity while meeting the needs of mixed-criticality workloads.

## II. BACKGROUND

### A. PTA

Timing analysis techniques for time-deterministic hardware [1] deliver a single WCET estimate. However, the pessimism of the WCET estimate grows if not enough information about hardware internal behaviour is available or hardware is complex and not amenable to WCET analysis.

Conversely, PTA [2,3,4,5] provides a distribution of WCET estimates so that the particular value at a given exceedance probability - a so called Probabilistic WCET (pWCET) estimate - can be theoretically exceeded with a probability upper-bounded by the exceedance threshold chosen, which can be arbitrarily low (e.g.,  $10^{-12}$  per hour), thus largely below the probability of hardware failures. In this work we focus on the Measurement Based version of PTA (MBPTA) as it is closer to industrial practice.

MBPTA uses Extreme Value Theory (EVT)[6] on the execution time measurements obtained at analysis time. EVT

is a well-known statistical method to approximate the tail of distributions, and so to derive the pWCET distribution. Thus, the execution time value at the desired exceedance threshold can be used as the pWCET estimate for the program under analysis. Figure 1 shows a hypothetical result of applying EVT to a collection of 1,000 observed execution times.

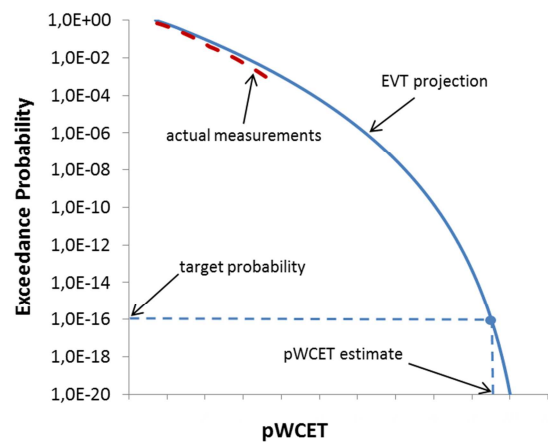


Fig. 1. Example of the 1-CDF and tail projection.

## III. PROBABILISTICALLY UPPER-BOUNDING INTER-TASK INTERFERENCE FEATURES

Shared caches in multicores challenge WCET estimation due to inter-task interferences. Hardware and software cache partitioning [7,8,9,10] address this issue although they complicate data sharing among tasks and the Operating System task scheduling and migration. We propose a technique [11] that overcomes the limitations of cache partitioning by enabling the estimation of trustworthy and tight WCET estimates for systems equipped with fully-shared (non-partitioned) time-randomised last level caches (LLCs). The main principle behind our proposal is that, while in a time-deterministic LLC interferences depend on when (time) and where (the particular cache in which) misses occur, a time-randomised LLC removes any dependence on the particular addresses accessed and its assigned cache set. This makes that the LLC interferences that a task suffers only depend on how often (frequency) its co-runner tasks miss in cache - thus evicting data - and not the particular address generating the miss. Based on this analysis we propose a simple hardware mechanism that limits the miss frequency of tasks in each core at analysis and deployment time in a manner that probabilistic upper-bounds can be obtained for the effect in the LLC of one



task on the other co-running tasks. Our approach removes cache partitioning constraints while making WCET estimates tighter.

### III. NoC

Among manycore shared resources the network-on-chip (NoC) has prominent impact on programs' execution time and WCET estimates, as it connects cores to memory and/or shared cache levels. Among existing NoC designs, only buses have been proven MBPTA-compliant [12] for different arbitration policies. However, bus scalability is limited since bus latency increases rapidly with the number of cores. We propose a new tree-based NoC design that is compatible with MBPTA requirements and that delivers scalability towards medium/large core counts.

### IV. CONCLUSIONS

Guaranteed performance needs of CRTES require using high-performance hardware designs but those jeopardise time predictability needed by conventional timing analyses. MBPTA has emerged recently as a powerful method to derive WCET estimates for critical tasks in safety-related systems on the top of complex hardware. We present new techniques to obtain time-composable WCET estimates on the top of shared non-partitioned LLCs, thus removing partitioning constraints, and MBPTA-compliant tree NoCs that outperform buses in multicores with 8/16 cores.

### ACKNOWLEDGMENT

The research leading to these results received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the PROXIMA Project ([www.proxima-project.eu](http://www.proxima-project.eu)) grant agreement no. 611085. This work has also been partially supported by the Spanish Ministry of Science and Innovation under grant TIN2012-34557 and the HiPEAC Network of Excellence. Mladen Slijepcevic is funded by the Obra Social Fundación la Caixa under grant Doctorado "la Caixa"—Severo Ochoa. Jaume Abella is partially supported by the Ministry of Economy and Competitiveness under Ramon y Cajal postdoctoral fellowship number RYC-2013-14717.

### REFERENCES

- [1] Wilhelm R. et al. The worst-case execution-time problem overview of methods and survey of tools. *ACM Transactions on Embedded Computing Systems*, 7:1-53, May 2008.
- [2] F.J. Cazorla et al. PROARTIS: Probabilistically analysable real-time systems. *ACM TECS*, 2013.
- [3] L. Cucu-Grosjean et al. Measurement-based probabilistic timing analysis for multi-path programs. In *ECRTS*, 2012.
- [4] S. Edgar and A. Burns. Statistical analysis of WCET for scheduling. In *RTSS*, 2001.
- [5] J. Hansen, S. Hissam, and G. A. Moreno. Statistical-based wcet estimation and validation. In *WCET Workshop*, 2009.
- [6] S. Kotz and S. Nadarajah. *Extreme value distributions: theory and applications*. World Scientific, 2000.
- [7] H. Kim, A. Kandhalu, and R. Rajkumar. A coordinated approach for practical os-level cache management in multi-core real-time systems. In *ECRTS*, 2013.
- [8] J. Liedtke, H. Hartig, and M. Hohmuth. OS-controlled cache predictability for real-time systems. In *RTAS*, 1997.

- [9] B. Ward, J. Herman, C. Kenna, and J. Anderson. Making shared caches more predictable on multicore platforms. In *ECRTS*, 2013.
- [10] M. Paolieri et al. Hardware support for WCET analysis of hard real-time multicore systems. In *ISCA*, 2009.
- [11] M. Slijepcevic et al. Time-analysable non-partitioned shared caches for real-time multicore systems. In *DAC*, 2014.
- [12] J. Jalle et al. Bus designs for time-probabilistic multicore processors. In *DATE*, 2014.

# Improving the prefetching performance through code region profiling

Martí Torrents<sup>1</sup>, Raul Martínez<sup>1</sup>, Carlos Molina<sup>2</sup>  
<sup>1</sup>UPC-BarcelonaTech, <sup>2</sup>Universitat Rovira i Virgili  
{martit,raulm}@ac.upc.edu, carlos.molina@urv.net

**Abstract-** In this work, we propose a new technique to improve the performance of hardware data prefetching. This technique is based on detecting periods of time and regions of code where the prefetcher is not working properly, thus not providing any speedup or even producing slowdown. Once these periods of time and regions of code are detected, the prefetcher may be switched off and later on, switched on. To efficiently implement such mechanism, we identify three orthogonal issues that must be addressed: the granularity of the code region, when the prefetcher is switched on, and when the prefetcher is switched off.

## I. INTRODUCTION

An imbalance in technological improvements in recent years has led to an increasing gap between processor and memory speeds. One way to address this problem is to use latency hiding techniques such as prefetching. This mechanism tries to predict which data the processor will require in the near future and bring it to the nearest cache level before it is needed by the application. Prefetching is a key technique employed by almost all current commercial high-performance processors in at least one but typically all levels of their memory hierarchy [1].

However, the prefetcher works as a double-edged sword because, if the prefetcher does not work properly, and too many of the prefetching requests are useless<sup>1</sup>, the mechanism will not increase performance but still may contribute to increment network latency, power consumption, and cache pollution<sup>2</sup>. This is particularly critical in CMP environments, where non-accurate prefetchers may perform important slowdowns since the network becomes another key resource, [2].

As a general rule, the prefetcher produces speedup as long as it generates a certain number of useful prefetch requests<sup>3</sup> respect the total number of requests generated by the prefetcher (i.e. the accuracy<sup>4</sup> is higher than a certain threshold that depends on the application and the prefetcher). Moreover, in the scope of CMP environments, we have observed that, the implementation of the prefetcher in a distributed and shared memory system is not trivial. There are some challenges [3] that must be handled in order to not increase the cases where the prefetcher does not work accurately and produces slowdowns.

<sup>1</sup> A **useless prefetch** is a request injected in the cache by the prefetcher and evicted from it without being used by the core.

<sup>2</sup> Prefetch requests **pollute** the cache when they replace data that is still required by the program.

<sup>3</sup> A **useful prefetch** is a request injected in the cache by the prefetcher that is actually requested afterwards by a regular memory request

<sup>4</sup> The **accuracy** is measured as the number of useful prefetch divided by the total prefetch requests.

## II. PROPOSAL

Although the prefetcher may contribute to global speedup, there may be periods of time and regions of code during the execution where the prefetcher may introduce slowdown. Our proposal aims to improve the global performance of the prefetcher by switching it off when it does not work properly, and switch it on again when it would provide actual performance improvements. Therefore, our target is to reduce the congestion of the system, the pollution that the prefetcher may produce in the cache module, and to carry out a better power management.

In order to achieve these objectives, there are several issues that must be properly addressed in order to obtain the maximum benefit from this technique.

1. **Code region granularity.** There are several types of code region to be considered: single instructions, basic blocs, superblocks, inner loops, outer loops, the whole dynamic code, etc. To implement this kind of mechanism, it would be mandatory to get prefetching statistics when executing code dynamically. Techniques such as [4] may be useful.
2. **Switching off the prefetcher.** Several statistics can be analyzed in order to switch off the prefetcher. In this study, we will focus on a particular prefetching statistic as the accuracy and also on cache miss ratios.
3. **Switching on the prefetcher.** The problem of switching on again the prefetcher is that, when the prefetcher is switched off, we cannot collect prefetching statistics. Therefore, we will focus on cache miss ratios and timeout mechanisms

## ACKNOWLEDGMENT

This work has been partially supported by the Spanish Ministry of Science and Innovation (MCI) and FEDER funds of the EU under the contracts TIN2010-18368 and TIN2013-47245-C2-2-R, and the Generalitat of Catalunya under grants 2009SGR1250.

## REFERENCES

- [1] M. Torrents, et al. "Comparative study of prefetching mechanisms". Jornadas del Paralelismo, Valencia (Spain), 2012.
- [2] M. Torrents, R. Martínez, C. Molina. "Network Aware Performance Evaluation of Prefetching Techniques in CMPs". Simulation Modeling Practice and Theory (SIMPAT), 2014.
- [3] M. Torrents, et al. "Prefetching Challenges in Distributed Memories for CMPs", In Proceedings of the International Conference on Computational Science (ICCS'15), Reykjavik, (Iceland), June 2015.
- [4] R. Martinez, J. Codina, E. Gibert, P. Lopez, M. Torrents, et al. "Profiling asynchronous events resulting from the execution of software at code region granularity". U.S. Patent Application 13/993,054, 2011.

# Dynamics of double-polarity subduction: application to the Western Mediterranean

PhD candidate:

Mireia Peral Millán

Institute of Earth Science Jaume Almera (CSIC)

mperal@ictja.csic.es

Supervisors:

Dr. Sergio Zlotnik

Universitat Politècnica de Catalunya (UPC)

sergio.zlotnik@upc.edu

Prof. Dr. Manel Fernàndez Ortiga

Institute of Earth Science Jaume Almera (CSIC)

mfernandez@ictja.csic.es

**Abstract-** *The Western Mediterranean tectonic setting and history is studied by means of three-dimensional numerical models. Goals are to understand its dynamics and to test the feasibility of a double-polarity subduction process that could have a key importance in the complex setup of this area.*

*The physical models, the numerical techniques involved, and some results in two and three dimensions are presented in this work.*

## I. INTRODUCTION

Present day tectonic setting of the Western Mediterranean is highly debated. Many researchers have different and conflicting views on the current setup, and, in particular in the geological history of this region. This variety results from the extremely complex data gathered in the region (seismic, petrology, volcanism, topography, gravity and surface structure) that need to be reconciled in a single model.

Recently Vergés and Fernàndez proposed a tectonic history of the region since Cretaceous (last 85 Million years) based on the interaction of two tectonic plates with opposite polarity (opposite direction of subduction, see Figure 2). They explain the formation and evolution of the Betic-Rif orogenic system and the associated Alboran back-arc basin (Fig. 1) by a lateral change in subduction polarity of the Ligurian-Thetys oceanic domain.

Double-polarity subduction and subduction reversal are proposed in many areas, for example [2-4], and studied numerically in 2D for example [5]. Nevertheless, at the moment there is no study to understand its feasibility, dynamics and geological consequences in a realistic three-dimensional study. The third dimension is particularly important in this process, as many subduction zones are oblique and therefore lateral propagation of deformation is expected. Moreover, surface deformation, for example curvature of the trench (the line on the surface of the contact between the plates) may be modified by the interaction between two closes by subducting plates.

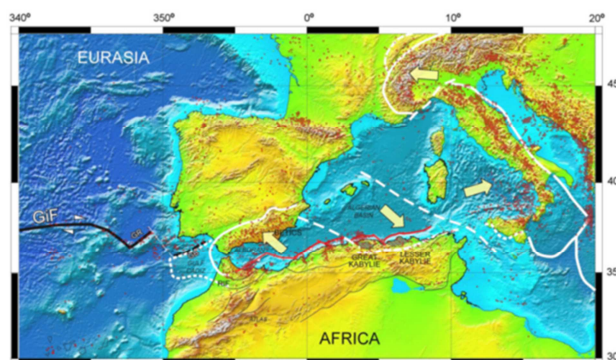


Fig. 1. Vergés and Fernàndez, 2012.

## II. OBJECTIVE

The aim of this study is to analyze the dynamic evolution of a double polarity subduction process and its consequences in order to test the physical feasibility of this interaction and provide geometries and evolutions comparable to the proposed to the Western Mediterranean. While this work is focused on numerical models, analogue modeling will be also carried out within the scope of the Thesis. An interdisciplinary group of researchers at the

Institute of Earth Science Jaume Almera (CSIC) in Barcelona is focused on the study of the Western Mediterranean; this work takes place within that framework, aiming to understand its geological history.

## III. MODEL AND METHODOLOGY

The dynamic evolution of the Earth at large temporal and spatial scales -for example subduction- is modeled as a multiphase flow problem. Its solution requires advanced numerical techniques usually consuming large amounts of computer resources. The main difficulties involved are: i) the

location of the material phases, that bring a double Lagrangian-Eulerian description of the problem, ii) a strong nonlinear character due to the viscous and viscoplastic rheologies involved and iii) the stiffness of the numerical problem produced by large contrast in viscosities that can reach 4 to 6 orders of magnitude and have sharp gradients.

In this study we are using the open source framework *Underworld* developed at Monash University [6]. *Underworld* solves a non-linear Stokes flow problem using Finite Elements combined with particle-in-cell approach, thus the discretization combines a standard Eulerian Finite Element mesh with Lagrangian particles to track the location of the phases.

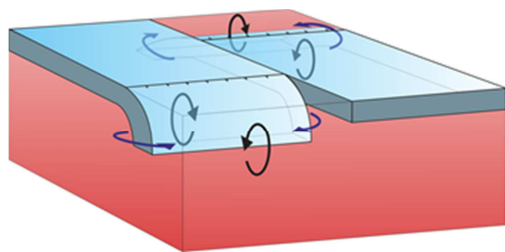


Figure 2. 3D double-polarity subduction model.

The model setup consists of two oceanic plates with viscoplastic rheology subducting into the upper mantle (Figure 2) and the problem is driven by a gravitational Rayleigh-Taylor instability. The main factors to be studied are the interaction between the two plates, the poloidal and toroidal mantle fluxes, the velocity variations of slabs, the stress distribution and the variations in the trench morphology.

With the purpose of being able to study the dynamics of the final model, several preliminary subduction models in 2D and 3D have been developed. The methodology and preliminary results will be presented in this work.

## REFERENCES

- [1] Vergés and Fernández, Tethys-Atlantic Interaction along the Iberia-Africa Plate Boundary: the Betic-Rif Orogenic System, *Tectonophysics*, Vol 579, 144-172, 2012.
- [2] Gianluca Vignaroli, Claudio Faccenna, Laurent Jolivet, Claudia Piromallo, Federico Rossetti, Subduction polarity reversal at the junction between the Western Alps and the Northern Apennines, Italy, *Tectonophysics*, Vol 450, 34-50, 2008.
- [3] Peter D. Clift, Hans Schouten and Amy E. Draut, A general model of arc-continent collision and subduction polarity reversal from Taiwan and the Irish Caledonides, *Geological Society*, London, vol 219, 81-98, 2007.
- [4] D. Brown, P.D. Ryan, J.C. Afonso, D. Boutelier, J.P. Burg, T. Byrne, A. Calvert, F. Cook, S. DeBari, J.F. Dewey, T.V. Gerya, R. Harris, R. Herrington, E. Konstantinovskaya, T. Reston, and A. Zagorevski, Arc-Continent Collision: The Making of an Orogen, Chapter 17, D. Brown and P.D. Ryan (eds.), *Arc-Continent Collision*, *Frontiers in Earth Sciences*, 477 DOI 10.1007/978-3-540-88558-0\_17, Springer-Verlag Berlin, 2011
- [5] Yury A. Mishin, Taras V. Gerya, Jean-Pierre Burg, James A.D. Connolly, Dynamics of double subduction: Numerical modeling, *Physics of the Earth and Planetary Interiors*, Vol 171, 280-295, 2008.
- [6] Moresi et al., A Lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials. *Journal of Computational Physics*, 184, 476-497, 2003.

# Characterization of coal power plants plume dynamics under typical synoptic conditions over the Iberian Peninsula

Víctor Valverde, María Teresa Pay, José María Baldasano

Earth Sciences Department, Barcelona Supercomputing Center-Centro Nacional de Supercomputación, Barcelona, Spain  
victor.valverde@bsc.es

**Abstract-** This work describes the pollution plumes of seven Spanish coal power plants under the most typical meteorological conditions that affect the Iberian Peninsula at synoptic scale. The aim is to understand how meteorology modulates the plume dynamics (length, altitude, orientation) and their contribution to  $\text{NO}_2$  and  $\text{SO}_2$  surface concentration. Using the BSC-ES operational air quality forecasting system (CALIOPE-AQFS), the behavior of the plumes is analyzed for a representative day of each of the six most common synoptic situations. The results show that the plumes from Atlantic facilities are mainly driven by synoptic conditions whereas for power plants located over the Mediterranean and on mountainous regions, mesoscale dynamics dominate. Moreover, when the injection of the pollutants is done within the planetary boundary layer there is an increase in the  $\text{NO}_2$  and  $\text{SO}_2$  surface concentrations close to (<15-20 km) the sources.

## I. INTRODUCTION

Air pollution can be defined as a situation in which substances are present in the atmosphere at concentration sufficiently high above their normal levels to produce a measurable effect on humans, ecosystems or materials [1]. The concentration of pollutants in the atmosphere depends on interlinked processes: the emission from natural and anthropogenic sources, the transport and transformation of the pollutants, and the deposition. The exceedances of air quality limit values in Spain are related to Nitrogen dioxide ( $\text{NO}_2$ ), Sulphur dioxide ( $\text{SO}_2$ ), Ozone ( $\text{O}_3$ ) and particulate matter [2]. The combustion of fossil fuels for the generation of electricity is an important contributor to  $\text{NO}_2$  and  $\text{SO}_2$  emissions. In Spain, despite the increase in the wind and solar electricity production, coal power plants still contribute with ~20% of the electricity generation [3]. In 2012, the associated annual emissions were ~14% of  $\text{NO}_2$  and ~52% of  $\text{SO}_2$  total Spanish emissions [4].

TABLE I  
CHARACTERISTICS OF THE STUDIED POWER PLANTS

Power plant / Acronym	Installed capacity (MW)	Stack height (magl)	$\text{NO}_2$ / $\text{SO}_2$ emissions (Gg.year <sup>-1</sup> )
As Pontes / ASP	1 468	356	7.46 / 4.99
Aboño / ABO	921	225	8.13 / 5.83
Compostilla / COM	1 341	290	8.38 / 3.79
Guardo-Velilla / GUA	498	176	3.07 / 0.82
Andorra / AND	1 101	343	10.00 / 11.71
Carboneras / CAR	1 159	200	9.80 / 13.99
Los Barrios / LBB	568	230	5.39 / 2.53

## III. OBJECTIVE

This work aims to characterize the dynamics of the plumes from coal power plants under the typical meteorological conditions affecting the Iberian Peninsula in order to better understand the role of these large point emission sources on air quality.

## IV. METHODOLOGY

There are currently 16 coal power plants (combustion installations with boilers > 300MWt) in Spain, seven of which have been selected in this study among those with highest electricity generation in 2012 (Table 1). The power plants have varied structural and topographical characteristics.

The present work uses a synoptic classification objectively established over the Iberian Peninsula for air quality purposes [5]. This classification identifies six typical circulation types (CTs) for the present climate (1983-2012) using a k-means clustering technique on daily sea level pressure. The six CTs consistently represent different advective patterns towards the Iberian Peninsula. The two most common CTs occur in summertime, replacing one another. NWadv (23.9% of climatic frequency) is a NW advective pattern characterized by the arrival of polar maritime air masses towards the Iberian Peninsula. IBtl (22.4%) is compatible with the development of the Iberian thermal low with net advection from North African air masses. ENEadv (21%), especially frequent in spring, is the result of a blocking anticyclone over central Europe that leads to E-NE advection towards the Iberian Peninsula. AtlHi (12%) is a winter anticyclonic situation that enables the arrival of Atlantic air masses towards the Iberian Peninsula, whereas ZonWadv (10%) is characterised by zonal Atlantic maritime advection. WNWadv (10%) is typical of transitional seasons and it presents unstable conditions over the Iberian Peninsula with W-NW advection. A representative day of each CT is objectively identified by [5] in the year 2012.

The characterization of the plume dynamics is performed on the representative day of each CT by means of the CALIOPE-AQFS (Fig. 1). Several evaluations of the modelling system give confidence to the results [6, 7]. The specific plumes are obtained following a zero-out approach. The characterization comprises a comprehensive analysis of the emissions, the synoptic and mesoscale meteorology and the surface and total column mass of  $\text{NO}_2$  and  $\text{SO}_2$ .

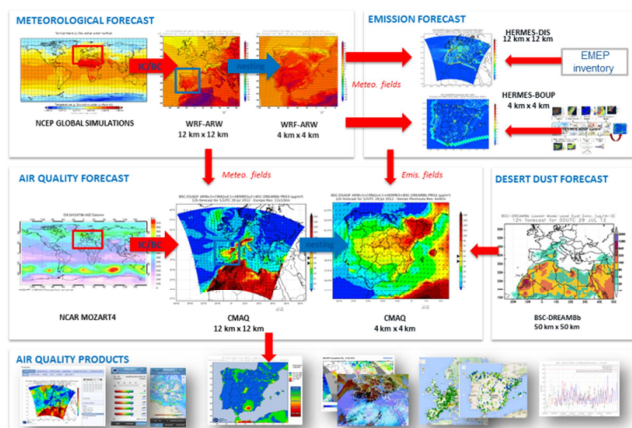


Fig. 1. Schematic view of the CALIOPE- air quality forecasting system [www.bsc.es/caliope](http://www.bsc.es/caliope). The CALIOPE couples the dedicated HERMES emission model, the WRF-ARW meteorological model, the CMAQ chemical-transport model and the BSC-DREAM8b mineral dust model [6].

## RESULTS AND DISCUSSION

The plume dynamics are comprehensively described in [7] using hourly  $\text{NO}_2$  and  $\text{SO}_2$  surface concentration maps total column mass maps (Fig. 2), vertical cross-sections, planetary boundary layer height, wind speed and direction and, the vertical vorticity at each power plant on each representative day.

The analysis of the transport dynamics of  $\text{NO}_2$  and  $\text{SO}_2$  emissions for each CT does not show a common behavior for the seven studied power plants. It is possible to define two kinds of power plants depending on their plume dynamics. On the one hand, there are facilities whose plume dynamics are mainly driven by synoptic winds (the plume follows the characteristic advective pattern of the CT). This is the case of power plants on the Atlantic region of the Iberian Peninsula ASP, ABO and COM. On the other hand, there are facilities located over the Mediterranean and in complex topographic areas (AND, CAR, LBB and GUA) whose plumes are driven by a combination of the synoptic advection and mesoscale processes (sea-land breezes and wind-channelling by river valleys).

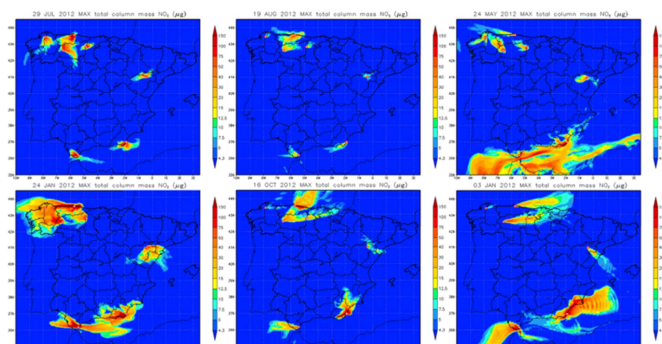


Fig. 2. Daily maximum  $\text{NO}_2$  total column mass ( $\mu\text{g}$ ) in the Iberian Peninsula associated to the emissions in the seven analyzed power plants on the representative day of each CT.

The maximum surface contribution of each power plant under the six CTs shows similar patterns for  $\text{NO}_2$  and  $\text{SO}_2$ , although  $\text{SO}_2$  contributions are generally higher. However, there are significant differences between power plants (largest contributions in CAR, ABO, and LBB) and synoptic conditions (AtlHi and NWadv present higher contributions than the other CTs).

For the synoptic-dominated power plants the plume length is larger for CTs characterised by Atlantic advection (NWadv, AtlHi, WSWadv and ZonWadv) than for CTs with non-Atlantic advection (IBtl and ENeadv). For  $\text{NO}_2$ , 76 km vs 24 km, respectively (Fig. 2). For  $\text{SO}_2$ , 81 km vs 45 km, respectively. Furthermore, the contribution to surface  $\text{NO}_2$  concentration is lower for Atlantic advection CTs than for non-Atlantic advection CTs, 11.8 and 14.8  $\mu\text{g NO}_2\cdot\text{m}^{-3}$ , respectively.

For the power plants that are along the Mediterranean (CAR, LBB) and/or in complex topographic areas (AND, GUA) the  $\text{NO}_2$  and  $\text{SO}_2$  contributions are higher under winter/autumn CTs (on average 19.2  $\mu\text{g NO}_2\cdot\text{m}^{-3}$  and 16.4  $\mu\text{g SO}_2\cdot\text{m}^{-3}$ ), during which emissions are maximal than during the summer/spring CTs (11.6  $\mu\text{g NO}_2\cdot\text{m}^{-3}$  and 5.6  $\mu\text{g SO}_2\cdot\text{m}^{-3}$ ).

Overall to all the power plants, close to the source contributions (< 15-20 km) are larger when the injection of the emissions is done within the planetary boundary layer (PBL), usually at midday under summer CTs. On autumn/winter CTs the PBL height is lower than the injection height, favouring horizontal dispersion of the plume.

## ACKNOWLEDGMENT

All simulations were performed on the MareNostrum supercomputer hosted by the Barcelona Supercomputing Center. This work is partially funded by the Iberdrola Foundation.

## REFERENCES

- [1] J.H. Seinfeld, and S.N. Pandis. *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, Inc., New York, NY, USA, 1326 pp. 1998.
- [2] European Environmental Agency. *Air quality in Europe – 2013 report*. 107 pp. 2013.
- [3] Red Eléctrica de España. *Informe del Sistema Eléctrico Español. Año 2012*. 140 pp. 2013.
- [4] Ministerio de Agricultura, Alimentación y Medio Ambiente. *Inventarios Nacionales de Emisiones a la Atmósfera 1990-2012. Volumen 2: Análisis por Actividades SNAP*. 110 pp. 2014.
- [5] V. Valverde, M.T. Pay, and J.M. Baldasano. “Circulation-type classification derived on a climatic basis to study air quality dynamics over the Iberian Peninsula,” *Int. J. Climatol*. 2014.
- [6] J.M. Baldasano, M.T. Pay, O. Jorbá, S. Gassó, and P. Jiménez-Guerrero. “An annual assessment of air quality with the CALIOPE modeling system over Spain,” *Sci. Total Environ*. 2011.
- [7] V. Valverde, M.T. Pay, and J.M. Baldasano. “A model-based analysis of  $\text{NO}_2$  and  $\text{SO}_2$  dynamics associated to coal-fired power plants under synoptic circulation types over the Iberian Peninsula”, unpublished.

# Quantum dynamics study of the hydrogen molecule confined in single-walled carbon nanotubes

M. Mondelo-Martell, F. Huarte-Larrañaga

Departament de Química Física & Institut de Química Teòrica i Computacional  
Universitat de Barcelona  
manel.mondelo@ub.edu

**Abstract**-In the present work a full-dimensional study of the dynamics of a hydrogen molecule confined in a narrow Single-walled Carbon Nanotube (SWCNT) is performed by using the Multi-configurational Time-dependent Hartree approach. New insights on the coupling between the different degrees of freedom of the molecule during the diffusion along the nanotube are found and discussed.

## I. INTRODUCTION

Storage of low-density gaseous species has become a field of intensive research in the last few years. Its interest arises with two basic objectives in mind: store large amounts of potential fuels in order to make their transport from the production centers feasible, and capture known pollutants from the atmosphere to prevent, for instance, greenhouse effect [1-5]. Concerning the first application, the paradigmatic case of hydrogen is the one in which we focus our study.

Hydrogen has a very low density that makes its storage in sufficient amounts not viable with the common storage devices like high-pressure tanks [3]. This is the reason why novel alternatives are being studied to store high amounts of hydrogen in a non-expensive, reversible way. Nanostructured materials, such as Carbon Nanotubes (CNT) or some Metal-organic Frameworks (MOFs) show the potential to be used for this purpose.

Hydrogen confined in nanostructured materials has been a matter of intensive research both theoretically and experimentally in the last few years [6-10]. The studies carried out in this kind of systems have shown not only the effectivity of some of these materials as storage devices, but also some distortions of the confined molecules at the molecular level when the cavities in which they are trapped are of the order of the nanometer. These changes of the electronic structure and the dynamics of confined species are of great importance since they not only allow a better understanding of the affinity of the confined molecules by the adsorbant, but because they have allowed the discovery of new potential applications for nanostructured materials, such as *quantum sievings*, which allow the separation of isotopomers of a given molecule, like  $H_2$  and  $D_2$ , due to the different Zero-point energy (ZPE) of molecules with different mass [11]. Also, some investigations point to the possibility of controlling chemical reactions at the molecular level using nanostructured materials [12].

There are mainly two theoretical approaches to study these systems. The first one is the study of adsorption through Molecular Dynamics simulations. This allows the study of bulk diffusion effects with hundreds of molecules in a non-expensive way, but relies completely on classical mechanics

and therefore cannot be used to study quantum phenomena, such as quantum confinement. The second approach, which is the one used in the present work, treats the problem quantum mechanically. This so-called Quantum Dynamics approach is much more expensive and has been rarely used to treat more than a few molecules, but on the other hand is the only tool which allows the study of intrinsically quantum effects.

The present work is structured in two main sections: first the computational tools, namely Multiconfigurational Time-dependent Hartree approach, as well as the model used to study our system, will be explained. Then the results of the study will be outlined and the main conclusions summarized.

## II. COMPUTATIONAL TOOLS AND MODELLING OF THE SYSTEM

### A. The Multiconfigurational Time-depend Hartree method

The MCTDH method allows an efficient propagation of multidimensional wave packets [13]. This is possible due to the use of a two-layer representation for the wave functions: it is represented in a relatively small basis set of time-dependent, low-dimensional basis functions, known as Single-Particle Functions (SPFs,  $\phi$ ), which in turn are expanded in a time-independent basis of primitive functions:

$$\Psi(Q_1, \dots, Q_p, t) = \sum_{j_1=1}^{n_1} \dots \sum_{j_p=1}^{n_p} A_{(j_1, \dots, j_p)}(t) \prod_{k=1}^p \phi_{j_k}^{(k)}(Q_k, t)$$

This *Ansatz* in combination with the Dirac-Frenkel variational principle yields a system of coupled equations of motion for the system which must be integrated to solve the dynamics of the problem. The two-layer approach allows an important decreasing of the size of the matrices to work with during the integration of the equations of motion, since the time-dependent basis set is able to adapt to the total wave function, yielding the best possible basis set at each time step.

### B. Modeling of the system:

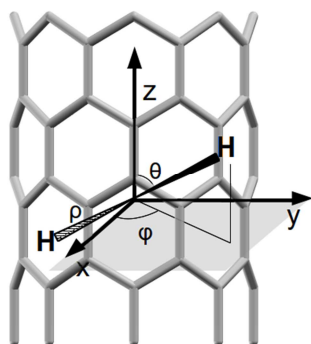
In our model, hydrogen is treated with a full-dimensional Hamiltonian. The degrees of freedom of the system are depicted in Fig. 1. This model allows seeing the effects of the coupling between the different degrees of freedom. On the other hand, the carbon nanotube is considered in the rigid cage approach, and therefore all atoms are fixed in their equilibrium geometry positions. Therefore this model will allow us to observe effects of the confinement in the structure of the

hydrogen molecule, but will not treat the energy exchange between the nanostructure and the confined molecule.

Due to the periodicity of the carbon nanotube, the computation and interpretation of the eigenstates of the hydrogen molecule in full dimension is not a trivial task. Instead, we will focus on another interesting and more straightforward property of this system such as the diffusion of the hydrogen molecule along the nanostructure. In order to study this particular feature of the system, a meaningful initial state must be propagated. In order to obtain this state, we add a harmonic trapping potential in the  $z$  coordinate. This allows us to calculate several eigenstates of the system with mainly no dependence on  $z$  and which mimic a thermal ensemble. These states can be propagated afterwards by “turning off” the trapping potential.

### III. RESULTS

The propagation of the wave packets in six dimensions for a time of 200 fs showed only a small distortion of the function in all coordinates except  $z$ . This small coupling is explained by the low corrugation of the potential along the  $z$  coordinate in the studied nanotube. Hence, the system is almost separable,



which opens a way to study the diffusion phenomena using a simplified model in the future, using the present full

Figure 5: Scheme of the degrees of freedom used to study the hydrogen molecule in the hollow cavity of a SWCNT.

dimensional study as a benchmark.

The study yielded satisfactory results for the diffusion coefficients of the hydrogen molecule at several temperatures. This property of the system was calculated following two different approaches. The first one was based on the time the wave packet needed to pass from a unit cell to the following. The second approach used the thermal flux correlation function formalism in order to obtain a reliable approximation to the rate of the process, which can be afterwards related to the diffusion coefficient.

### ACKNOWLEDGMENT

We acknowledge financial support from the Spanish Ministerio de Economía y Competitividad (Ministry of Economy and Competitiveness) (CTQ2013-41307-P) and

Generalitat de Catalunya (SGR-2009-17). M.M. further thanks a predoctoral grant from the FPU program (FPU2013/02210) of the Spanish Ministerio de Educación, Culture y Deporte (Ministry of Education, Culture and Sports, Spain).

### REFERENCES

- [1] A. Dillon, K. Jones, T. Bekkedahl, C. Kiang, D. Bethune, M. Heben, “Storage of hydrogen in single-walled carbon nanotubes,” *Nature*, vol. 386, pp 377-379, 1997.
- [2] H. Cheng, Q. Yang and C. Liu, “Hydrogen storage in carbon nanotubes,” *Carbon*, vol. 39, pp 1447-1454, 2001.
- [3] L. Schlapbach, A. Züttel, “Hydrogen-storage materials for mobile applications,” *Nature*, vol. 414, pp 353-358, 2001.
- [4] S. Dunn, “Hydrogen futures: toward a sustainable energy system,” *International Journal of Hydrogen Energy*, vol. 27, pp 235-264, 2002.
- [5] J. Atwood, L. Barbour, A. Jerga, “Storage of methane and freon by interstitial van der Waals confinement,” *Science*, vol. 296, pp 2367-2369, 2002.
- [6] S. FitzGerald, *et al.* “Quantum Dynamics of the interstitial  $H_2$  in solid  $C_{60}$ ,” *Physical Review B*, vol. 60, pp 6439-6451, 1999.
- [7] A. Harris, T. Yildirim, “Quantum dynamics of a hydrogen molecule confined in a cylindrical potential,” *Physical Review B*, vol. 67, pp 45413, 2003.
- [8] T. Lu, E. Goldfield, S. Gray, “Quantum States of hydrogen and its isotopes confined in single-walled carbon nanotubes: dependence on interaction potential and extreme two-dimensional confinement,” *Journal of Physical Chemistry B*, vol. 110, pp 1742-1751, 2006.
- [9] F. Huarte-Larrañaga, M. Albertí, “A molecular dynamics study of the distribution of molecular hydrogen physisorbed on single walled carbon nanotubes,” *Chemical Physics Letters*, vol. 445, pp 227-232, 2007.
- [10] J. Suárez, F. Huarte-Larrañaga, “Hydrogen confined in single-wall carbon nanotubes: Anisotropy effects on ro-vibrational quantum levels,” *Journal of Chemical Physics*, vol. 137, pp 064320, 2012.
- [11] Q. Wang, S. Challa, D. Scholl, J. Johnson, “Quantum sieving in Carbon Nanotubes and Zeolites,” *Physical Review Letters*, vol. 82, pp 956-959, 1999.
- [12] T. Nielsen *et al.* “A reversible nanoconfined chemical reaction,” *ACS nano*, vol 4, pp 3903-8, 2010.
- [13] H.- D. Meyer, U. Manthe, L. Cederbaum, “The multi-configurational time-dependent Hartree approach,” *Chemical Physics Letters*, vol. 165, pp 73-78, 1990.



# Sugar Conformations that Enhance Cleavage of Glycosidic Bonds in Carbohydrate-Active Enzymes

Lluís Raich<sup>1</sup> and Carme Rovira<sup>1,2</sup>

1. Departament de Química Orgànica and Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franquès 1-11, 08028 Barcelona, Spain.

2. Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08018 Barcelona, Spain.  
lraich@ub.edu

**Abstract-** Carbohydrate-active enzymes select particular conformations of one sugar of their carbohydrate substrates to enhance catalysis. The evolution of this conformation during catalysis, known as the catalytic itinerary, is of fundamental interest in glycobiology. Here we calculate the free energy landscape of  $\beta$ -xylose to predict the possible catalytic itineraries of  $\beta$ -xylan enzymes. Our results discard one of the three catalytic itineraries proposed on the basis of X-ray structures of mutant enzymes. Additional classical and quantum mechanics/molecular mechanics (QM/MM) calculations indicate that complexes obtained from modified enzymes do not necessarily represent the structures that take place in the reaction coordinate.

## I. INTRODUCTION

The covalent bond that joins a simple carbohydrate molecule (also known as sugar) to another group, which may or may not be another carbohydrate, is known in chemistry as *glycosidic bond*. These bonds are extremely stable in solution and display half-lives of several million years. However, in the presence of enzymes such as glycoside hydrolases (GHs), the half-life of glycosidic bonds can decay up to the millisecond range [1]. How do these enzymes enhance the breakage of glycosidic bonds is a topic that has been extensively studied since 1894 Emil Fisher's 'lock and key' model [2].

It is nowadays accepted that, for beta oligosaccharides (carbohydrate molecules bound by beta glycosidic bonds), one of the sugar molecules changes shape upon binding to the GH enzyme [3], facilitating catalysis by placing the leaving group of the reaction in an axial position. In particular, a key sugar molecule deforms from its ground state *chair* conformation in solution to a new conformation (typically a *boat*, *skew-boat* or *half-chair*) that is less populated in the same conditions (Fig. 1). Knowing the precise conformation of the sugar molecule in the active site of the enzyme and how this shape changes during catalysis (the catalytic itinerary) is important not only to understand how these enzymes work, but also to design new drugs for GHs involved in human health and disease [4], as well as engineering GHs for their use in food, detergent, oil, gas and biotechnological industries. In particular,  $\beta$ -xylanases are GH enzymes responsible for the hydrolysis of glycosidic bonds in  $\beta$ -xylans, a group of hemicelluloses of high biotechnological interest that are found in plant cell walls. An intriguing aspect of these enzymes is that, unlike other GHs, their  $\beta$ -xylan carbohydrate substrates are assumed to adopt more than one conformation in the enzyme active site, thus leading to several conformational itineraries during catalysis (Fig. 2 B). Based on crystallographic structures, some  $\beta$ -xylanases have even been predicted to adopt conformations that disobey previously established rules for enzymatic glycosidic bond cleavage [5].

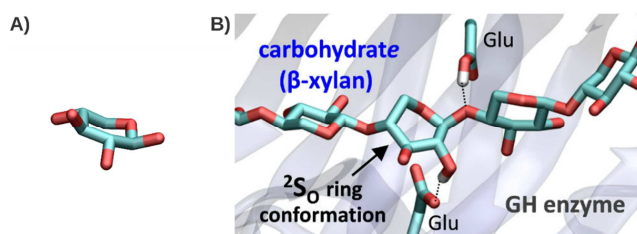


Fig. 1. A)  $\beta$ -xylose sugar in the most populated conformation in solution: *chair*  ${}^4C_1$ . B)  $\beta$ -xylose oligosaccharide bound to a family 11 GH enzyme. The key sugar, which is located between the two glutamic catalytic residues, displays a  ${}^2S_0$  skew-boat conformation. Almost all hydrogens have been omitted for the sake of clarity.

In recent years, our group has contributed to the understanding of how GHs tune carbohydrate conformations to enhance catalysis [6,8]. Specifically, we demonstrated that the computed conformational free energy landscape (FEL) of isolated monosaccharides (e.g. glucose and mannose) correlate with the observed X-ray structures of enzyme-ligand complexes [7]. In other words, one can predict the conformation and electronic properties that a given carbohydrate will exhibit in the active sites of GHs from gas phase calculations [8,9]. Here we investigate all possible conformations of a  $\beta$ -xylose sugar molecule using state-of-the-art computational techniques based on molecular dynamics (MD). In particular, we compute the conformational FEL of  $\beta$ -xylose by *ab initio* metadynamics and use it to predict the most likely catalytic itineraries for  $\beta$ -xylanases. Moreover, we employ classical and quantum mechanics/molecular mechanics (QM/MM) MD simulations to check whether common enzyme modifications used in experiments can affect the observed xylose conformation, leading to incorrect assignments of catalytic itineraries.

## II. METHODS

Metadynamics [10] is a molecular dynamics based technique that enhances phase space exploration by the addition of repulsive potentials on a set of collective variables (functions of atomic positions) that enclose the slowest modes that are relevant to the process of interest.

We used metadynamics on the phase space of  $\beta$ -xylose described by Density Functional Theory, using  $\theta$  and  $\Phi$  Cremer and Pople puckering coordinates [11] as collective variables (Fig. 2 A). All *ab initio* and hybrid QM/MM molecular dynamics simulations were carried out within the Car-Parrinello approach, as implemented in the CPMD 3.15.1

program [12]. The AMBER [13] force-field was used for the classical molecular dynamics.

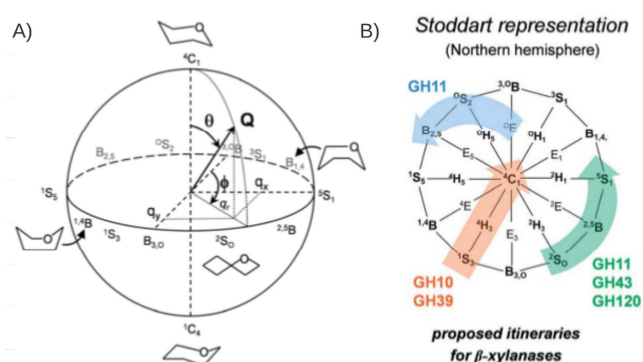


Fig. 2. A) Cremer and Pople puckering coordinates of a pyranose ring ( $Q$ ,  $\theta$  and  $\Phi$ ). This coordinates are able to unequivocally distinguish different conformations of a six-membered ring, having  $^4C_1$  chair represented at the top,  $^1C_4$  chair at the bottom, *boats* and *skew-boats* in the equator and *half-boats* and *envelopes* in between. B) Northern projection (Stoddart representation) with the proposed catalytic itineraries for different families of  $\beta$ -xylosidases (the number after GH stands for the family).

### III. RESULTS

The conformational FEL of  $\beta$ -xylose (Fig. 3) contains a deep minimum in  $^4C_1$  as well as a wide minimum between  $^1S_3$  and  $^2S_0$ . As previously found for other pyranoses, the puckering coordinates of all the key xylose sugar molecules extracted from experimental structures of  $\beta$ -xylosidase enzyme-ligand complexes are located, without exceptions, in low-energy regions of the FEL. These low-energy regions agree with two of the three conformational routes previously proposed for different  $\beta$ -xylosidase families (orange and green pathways of Fig. 2 B). The excluded one (blue pathway), which passes through a high free energy region, was further studied by classical and QM/MM MD simulations on the crystallographic complex from which was proposed [14]. The results of these calculations suggest that the  $^4C_1/{}^0E$  conformation found by X-ray experiments is an artefact due to enzyme mutation and that the most probable conformation in the wild type is  $^2S_0$ , which places the leaving group in an axial position and is energetically consistent with the computed FEL. Although not all mutant enzymes are expected to behave equally (similar calculations on the complex of a modified GH10 crystal structure [15] that agree with the FEL show both mutant and wild type with the same conformational pattern), our results indicate that complexes obtained from modified enzymes do not necessarily represent the structures that take place in the reaction coordinate.

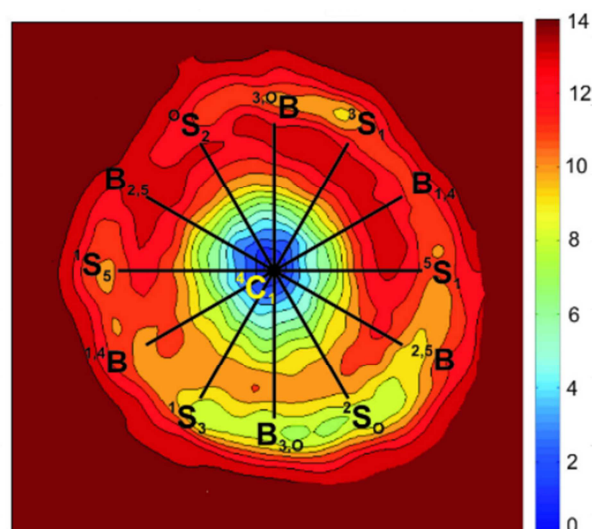


Fig. 3. Conformational free energy landscape of an isolated  $\beta$ -xylose (Stoddart representation of the puckering sphere, Northern hemisphere). Free energy is given in kcal·mol<sup>-1</sup> and each contour line of the diagram corresponds to 1 kcal·mol<sup>-1</sup>.

### ACKNOWLEDGMENT

This work was supported by the Generalitat de Catalunya (grant 2014SGR-987) and by the Ministerio de Economía y Competitividad (MINECO) (grant CTQ2011-25871). We acknowledge the computer support, technical expertise, and assistance provided by the Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS). L. R. acknowledges the University of Barcelona for a PhD studentship (APIF-2013/2014).

### REFERENCES

- [1] Wolfenden, R. and Snider, M. J.; *Acc. Chem. Res.* **2001**, 34, 938-945.
- [2] Fischer, E.; *Ber. Dtsch. Chem. Ges.* **1894**, 27, 2985-2993.
- [3] Vocadlo, D. and Davies, G. D.; *Curr. Opin. Chem. Biol.* **2008**, 12, 539-555.
- [4] Gloster and Vocadlo, D.; *Nat. Chem. Biol.* **2012**, 8, 683-694.
- [5] Iglesias-Fernández, J.; Raich, L.; Ardèvol, A. and Rovira, C.; *Chem. Sci.* **2015**, 6, 1167-1177.
- [6] Biarnés, X.; Nieto, J.; Planas, A.; Rovira, C.; *J. Biol. Chem.* **2006**, 281, (3), 1432-41.
- [7] Biarnés, X.; Ardèvol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M.; *J. Am. Chem. Soc.* **2007**, 129, (35), 10686-93.
- [8] Davies, G. J.; Planas, A.; Rovira, C.; *Acc. Chem. Res.* **2012**, 45, (2), 308-16.
- [9] Thompson, A. J.; Dabin, J.; Iglesias-Fernández, J.; Ardèvol, A.; Dinev, Z.; Williams, S. J.; Bande, O.; Siriwardena, A.; Moreland, C.; Hu, T. C.; Smith, D. K.; Gilbert, H. J.; Rovira, C.; Davies, G. J.; *Angew. Chem. Int. Ed. Engl.* **2012**, 51, (44), 10997-1001.
- [10] Bussi, G.; Laio, A. and Parrinello, M.; *Phys. Rev. Lett.* **2006**, 96, 090601.
- [11] Cremer, D. and Pople, J.A.; *J. Am. Chem. Soc.* **1975**, 97, 1354-1358.
- [12] CPMD, <http://www.cpmc.org/>, Copyright IBM Corp 1990-2008, Copyright MPI für Festkörperforschung Stuttgart 1997-2001.
- [13] Pearlman, D.A.; Case, D.A.; Caldwell, J.W.; Ross, W.S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G. and Kollman, P.; *Comput. Phys. Commun.* **1995**, 91, 1-41.
- [14] Wan, Q.; Zhang, Q.; Hamilton-Brehm, S.; Weiss, K.; Mustyakimov, M.; Coates, L.; Langan, P.; Graham, D. and Kovalevsky, A.; *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2014**, 70, 11-23.
- [15] Suzuki, R.; Fujimoto, Z.; Ito, S.; Kawahara, S.; Kaneko, S.; Taira, K.; Hasegawa, T. and Kuno, A.; *J. Biochem.* **2009**, 146, 61-70.

# pyDock performance in 5th CAPRI edition: from docking and scoring to binding affinity predictions and other challenges

Chiara Pallara, Brian Jiménez-García, Miguel Romero, and Juan Fernandez-Recio  
Joint BSC-IRB Research Programme in Computational Biology, Barcelona Supercomputing Center, Barcelona, Spain  
[chiara.pallara@bsc.es](mailto:chiara.pallara@bsc.es)

**Abstract-** *Proteins form the executive machinery underlying all the biological processes that occur within and between cells, from DNA replication to protein degradation. Although genome-scale technologies enable to clarify their large, intricate and highly dynamics networks, they fail to elucidate the detailed molecular mechanism that underlies the protein association process. Therefore, one of the most challenging objectives in biological research is to functionally characterize protein interactions by solving 3D complex structures.*

*This is, however, not a trivial task as confirmed by the large gap that exist between the number of complexes identified by large-scale proteomics efforts and those for which high-resolution 3D experimental structures are available. For these reasons, computational docking methods, aimed to predict the binding mode of two proteins starting from the coordinates of the individual subunits, are bound to become a complementary approach to solve the structural interactome.*

*Given its importance, the field of protein docking has experienced an explosion in recent years partially propelled by CAPRI (<http://www.ebi.ac.uk/msd-srv/capri/>). CAPRI (Critical Assessment of PRedicted Interaction) is a community-wide blind experiment aimed at objectively assessing the performance of computational methods for modeling protein interactions by inviting developers to test their algorithms on the same target system and quantitatively evaluating the results.*

*In order to test pyDock,<sup>1</sup> a docking scoring algorithm developed in our group, the PID (Protein Interaction and Docking) group of the BSC Life Science Department, we have participated in all the 15 targets (T46 to T58) of the 5th CAPRI edition (2010-2012). Our automated protocol confirmed to be highly successful to provide correct models in easy-to-medium difficulty protein-protein docking cases placing among the Top5 ranked groups out of more than 60 participants.*

**Key words:** Complex structure, CAPRI, protein-protein docking, pyDock, protein interactions.

## I. INTRODUCTION

One of the major challenges in structural biology is to provide structural data for all complexes formed between proteins and other macromolecules. Current structural coverage of protein-protein interactions (i.e. available experimental structures plus potential models based on homologous complex structures) is below 4% of the estimated number of possible complexes formed between human proteins.<sup>2</sup> The pace of experimental determination of complex structures is still behind the determination of individual protein structures. In addition, many of these interactions will never be determined by x-ray crystallography because of their transient nature. For these reasons, computational docking methods aim to become a complementary approach to solve the structural interactome. The field of protein docking has experienced an explosion in recent years, partially propelled by the CAPRI (<http://www.ebi.ac.uk/msd-srv/capri/>). CAPRI (Critical Assessment of PRedicted Interaction) is a community-wide blind experiment aimed at objectively

assessing the performance of computational methods for modeling protein interactions by inviting developers to test their algorithms on the same target system and quantitatively evaluating the results. This involves sampling putative association modes and modeling their atomic structure (the docking problem), and identifying those likely to be stable out of a very large pool of decoys (the scoring problem). Models submitted by participants are finally evaluated in comparison with experimental coordinates made available by their authors to the CAPRI assessors according to some criteria as described in Figure 1 of Lensink et al. *Proteins* 2007 69:704.

In order to test pyDock,<sup>1</sup> a docking scoring algorithm developed in our group, we have participated in all the 15 targets of the 5th CAPRI edition (2010-2012). In addition to the standard prediction of protein-protein targets, this edition has entered into related areas including binding affinity predictions and free energy changes upon mutation, as well as prediction of sugar binding and interface water molecules. Our overall experience has been highly rewarding and we describe here the details of our participation and the key factors of our success.

## II. MATERIALS AND METHODS

### A. Generation of rigid-body docking poses for the predictors experiment

In all targets, we used FTDock<sup>3</sup> and ZDOCK 2.1<sup>4</sup> to generate 10,000 and 2,000 rigid-body docking poses, respectively. For the final four targets of this edition (T53, T54, T57 and T58) we generated an additional pool of flexible docking poses using SwarmDock.<sup>5</sup>

### B. Scoring of rigid-body docking poses for both the predictors and the scorers experiment

We scored the docking models generated by the above described methods with our pyDock protocol, based on energy terms previously optimized for rigid-body docking. The binding energy is basically composed of ASA-based desolvation, Coulombic electrostatics and van der Waals energy (with a weighting factor of 0.1 to reduce the noise of the scoring function). Cofactors, water molecules and solvent ions were not considered for scoring.

### C. Removal of redundant docking poses

After scoring, we eliminated redundant predictions in order to increase the variability of the predictions and maximize the success chances by using a simple clustering algorithm with a distance cut-off of 4.0 Å, as previously described.<sup>6</sup>

### D. Minimization of final models

The final ten selected docking poses were minimized in order to improve the quality of the docking models and reduce the number of interatomic clashes. In the majority of the targets we used TINKER. In targets T53 and T54 we used

CHARMM<sup>7</sup> while in target T58 we used AMBER10 with AMBER parm99 force field.<sup>8</sup>

#### E. Modeling of subunits with no available structure

For several targets, the structures of the subunits were not available and needed to be modeled. We used Modeller 9v6<sup>9</sup> with default parameters based on the template/s suggested by the organizers or on other homologue proteins found by BLAST search (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). The final selected model was that with the lowest DOPE score.

### III. RESULTS AND DISCUSSION

In this CAPRI edition we submitted predictions for all the proposed targets. Our results for the standard protein-protein docking assessment are summarized in Table I and Fig. 1.

TABLE I

Target	Type <sup>a</sup>	Predictors		
		Submission rank <sup>b</sup>	Quality <sup>c</sup>	Successful Groups <sup>d</sup>
T46	HH	-	-	2 (40)
T47	HU	1	***	25 (29)
T48	UU	3	*	14 (32)
T49	UU	4	*	14 (33)
T50	UH	1	**	18 (40)
T51	DHD	-	-	3 (46)
T53	UH	3 <sup>e</sup>	**	20 (42)
T54	UH	-	-	4 (41)
T58	UU	5	**	11 (23)

Table I. Results of pyDock protocol for all protein-protein targets in predictors.

a B: bound; U: unbound; H: homology-based model.

b Rank of the best model within our submission to CAPRI.

c Quality of our best model according to CAPRI criteria: acceptable (\*), medium (\*\*), and high (\*\*\*)

d Number of successful groups for each target; in brackets, total number of participants.

e Model rank 1 had acceptable accuracy (\*).

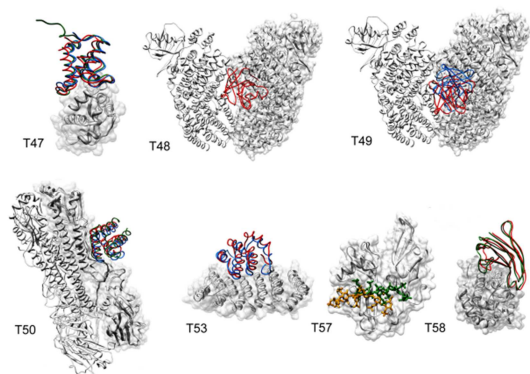


Fig. 1. Representation of our best models for targets T47, T48, T49, T50, T53, T57 and T58. For each target, receptors are superimposed and shown in white. Ligand in our best model as predictors is shown in red, and as scorers in blue. For comparison, the structure of the experimental complex (if available) is represented in green.

For the generation of docking poses, the better grid resolution used for FTDock and the use of flexible SwarmDock for the last targets were key for the success. In selected targets (T47, T48 and T58), distance restraints were used, but in most cases

this did not make a difference. In the target T58, SAXS data was used for complementary scoring with pyDockSAXS,<sup>10</sup> which slightly improved the scoring. We obtained consistently good models for all non-difficult cases, although they were far from being trivial, since their subunits were unbound or needed to be modeled based on homology templates. In all cases but one our successful models were ranked within our first five submitted solutions, being ranked 1st in several cases.

### IV. CONCLUSIONS

In this CAPRI edition we learned that our automated protocol is useful to provide correct models in easy-to-medium difficulty protein-protein docking cases, but we need further methodological development for difficult cases, especially when subunits need to be modeled based on homologues with low sequence identity. Our overall experience has been highly rewarding, pyDock docking scheme confirmed its high performance in protein complexes prediction placing among the Top5 ranked groups out of more than 60 participants (Table II).

TABLE II

Rank	Group	Summary: #Targets / *** + ** + *
1	Bonvin	9 / 1 *** + 3 ** + 5 *
2	Bates	8 / 2 ** + 6 *
3	Vakser	7 / 1 *** + 6 *
4	Vajda	6 / 2 *** + 3 ** + 1 *
<b>5</b>	<b>Fernandez-Recio</b>	<b>6 / 1 *** + 3 ** + 2 *</b>
5	Shen	6 / 1 *** + 3 ** + 2 *
7	Zou	6 / 1 *** + 2 ** + 3 *
8	Zacharias	6 / 1 *** + 5 *
9	ClusPro	6 / 4 ** + 2 *
10	Eisenstein	5 / 1 *** + 2 ** + 2 *

Table II. Overall pyDock performance among the Top10 ranked groups. Predictions are classified as acceptable (\*), medium (\*\*), and high (\*\*\*).

### REFERENCES

- Cheng, T. M.-K., Blundell, T. L., & Fernandez-Recio, J. (2007). *Proteins* 68, 503-15.
- Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., et al. (2009). *Nat Methods* 6, 83-90.
- Gabb, H. A., Jackson, R. M., & Sternberg, M. J. (1997). *J Mol Biol* 272, 106-20.
- Chen, R., Li, L., & Weng, Z. (2003). *Proteins* 52, 80-7.
- Moal, I. H. & Bates, P. A. (2010). *Int J Mol Sci* 11, 3623-48.
- Pons, C., Solernou, A., Perez-Cano, L., Grosdidier, S., & Fernandez-Recio, J. (2010). *Proteins* 78, 3182-8.
- Brooks, B. R., Brooks, 3rd, C. L., Mackerell, Jr, A. D., Nilsson, L., et al (2009). *J Comput Chem* 30, 1545-614.
- Case, D. A., Cheatham, 3rd, T. E., Darden, T., Gohlke, H., Luo, R., Merz, Jr, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). *J Comput Chem* 26, 1668-88.
- Sali, A. & Blundell, T. L. (1993). *J Mol Biol* 234, 779-815.
- Pons, C., D'Abramo, M., Svergun, D. I., Orozco, M., Bernadó, P., & Fernández-Recio, J. (2010). *J Mol Biol* 403, 217-30.

# Implementation of an Internal Coordinates Anisotropic Network Model in PELE

Victor A. Gil<sup>†\*</sup> and Victor Guallar<sup>†‡</sup>

<sup>†</sup> Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, 08034 Barcelona, Spain

<sup>‡</sup> Institutio Catalana de Recerca i Estudis Avancats (ICREA), Passeig Lluís Companys 23, E-08010 Barcelona, Spain  
\*victor.gil@bsc.es

**Abstract-** The use of computational methods to elucidate the ligand-protein binding mechanisms is of utmost importance for the pharmaceutical industry. PELE (Protein Energy Landscape Exploration) software has proved to have good predictive power. We want to further improve it by changing its conformational sampling step.

## I. INTRODUCTION

Pharmaceutical industry invested more than \$1.2 billion per each new developed drug in 2009 [1]. The accurate simulation of the interactions between drugs and their targets using computational chemistry methods would help companies to improve their productivity in drug development as well as decreasing its associated costs.

A good characterization of the conformational space of biomolecules is a key part of these methods. Unfortunately, due to the huge number of degrees of freedom studied systems usually have, this sampling cannot be exhaustive.

## II. PELE

PELE's primary application is to unveil the mechanism of protein-ligand interactions [2, 3], and it is currently being extended to handle other types of biopolymers. It implements a Monte Carlo scheme, with two main steps: perturbation and relaxation (see Fig. 1).

In the perturbation step the current protein conformation is changed to a new one by applying a linear combination of its Anisotropic Network Model (ANM) modes. After this, the ligand is rotated and translated to a new position, and its conformation is changed if needed.

During the relaxation step, the side chains with higher potential energies are changed using a rotamer library [4]. A modified Truncated Newton algorithm is then used in order to further lower the energy of the system. Finally, in the Metropolis acceptance step, the Boltzmann probability of the new state will be calculated. Depending on it, it will be accepted and used as the initial conformation of the next step, or it will be rejected and discarded.

In recent works, PELE has been successfully used to solve problems beyond protein ligand docking [5, 6] like the analysis of mutational effects on ligand delivery [7] or the description of ligand kinetics at only a percent of the computational cost that other more consolidated methods like MD have [8].

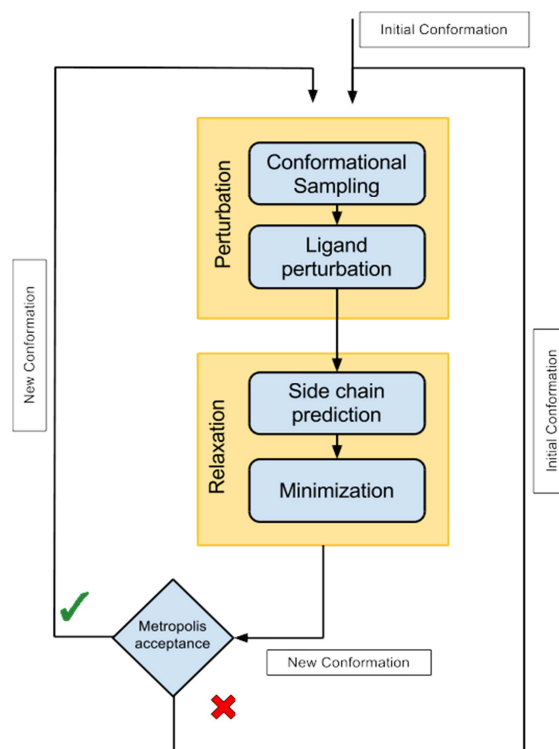


Fig. 1: PELE implements an iterative Monte Carlo scheme with two main steps: perturbation and relaxation.

## III. CONFORMATIONAL EXPLORATION IN PELE

### ANM using Cartesian coordinates

In the ANM methodology, the protein is simplified to an elastic network where each atom (or coarse grain unit) becomes a node. PELE uses an alpha-carbon based coarse grain model. Once the elastic network has been defined, the Hessian ( $H$ ) is calculated and the eigenvalues ( $V$ ) and eigenvectors ( $\Lambda$ ) resulting from solving (1) will be used to generate the new conformation. In order to do this, a subset of the eigenvectors are chosen (either randomly or user-defined) and a linear combination of them is calculated. The resultant vectors will define a set of target coordinates, the conformation proposal. The next conformation is calculated by adding harmonic constraints between the initial alpha carbon positions and the target ones and minimizing.

$$HV = V\Lambda. (1)$$

Yet being a simplification, the predicting capabilities of ANM have been demonstrated beyond doubt [9, 10]. However, there is still room for improvement when it comes to its application. For example, ANM suffers from the so-called 'tip effect', that can potentially nullify the modes in structured regions. In addition, implementations based on

linear interpolations in the mode direction can generate non-physical deformations that must be corrected by minimizing the system. Also, our use of the ANM methodology in PELE has shown to have some limitations when it comes to produce medium range deformations in structured regions or to perform folding/unfolding events.

#### Minimization bias

The extensive use of energy minimization (twice if using the current ANM scheme) entails the risk of biasing the sampling process to near minimum zones. This would trap the exploration in a way that regions of the conformational space that must be accessible can become unreachable.

#### ANM using dihedral angles

Cartesian coordinates (CC) are the usual choice to use the ANM, especially because of its simplicity. A less common option is to use the internal coordinates (IC) of the protein, more specifically the backbone dihedral angles.

The coarse grain model is composed of units delimited by the  $\phi$ ,  $\psi$  torsion angles; the elastic network is defined similarly to the CC ANM. The Hessian (H) and Kinetic (K) matrices are calculated differently [11, 12], and the modes, obtained by solving (2), will represent infinitesimal dihedral rotations equivalent to their Cartesian counterparts. New angular increments are calculated and applied using standard transformation matrices [13].

$$HV = VK\Lambda. (2)$$

Our hypothesis is that this new methodology will be able to overcome some of the problems stated above: there is no need to use a minimization to generate the next conformation, which is computationally more efficient and diminishes the minimization bias. Moreover, we expect that the use of curvilinear displacements can produce broader deformations and reproduce folding events.

This approach has already been implemented in PELE and it is currently being thoroughly tested.

#### ACKNOWLEDGMENTS

We would like to thank Dr. Christoph Grebner and Dr. Anders Hogner for their invaluable help.

#### REFERENCES

- [1] C. P. Adams and V. V. Brantner, "Spending on new drug development1," *Health Econ.*, vol. 19, no. 2, pp. 130–141, Feb. 2010.
- [2] K. W. Borrelli, A. Vitalis, R. Alcantara, and V. Guallar, "PELE: Protein energy landscape exploration. A novel Monte Carlo based technique," *J. Chem. Theory Comput.*, vol. 1, no. 6, pp. 1304–1311, 2005.
- [3] A. Madadkar-Sobhani and V. Guallar, "PELE web server: atomistic study of biomolecular systems at your fingertips," *Nucleic Acids Res.*, vol. 41, no. Web Server issue, pp. W322–W328, Jul. 2013.
- [4] M. P. Jacobson, G. A. Kaminski, R. A. Friesner, and C. S. Rapp, "Force Field Validation Using Protein Side Chain Prediction," *J. Phys. Chem. B*, vol. 106, no. 44, pp. 11673–11680, Nov. 2002.
- [5] A. Hosseini, M. Espona-Fiedler, V. Soto-Cerrato, R. Quesada, R. Pérez-Tomás, and V. Guallar, "Molecular Interactions of Prodiginines with the BH3 Domain of Anti-Apoptotic Bcl-2 Family Members," *PLoS ONE*, vol. 8, no. 2, p. e57562, Feb. 2013.
- [6] D. Linde, R. Pogni, M. Cañellas, F. Lucas, V. Guallar, M. C. Baratto, A. Sinicropi, V. Sáez-Jiménez, C. Coscolín, A. Romero, F. J. Medrano, F. J. Ruiz-Dueñas, and A. T. Martínez, "Catalytic surface radical in dye-decolorizing peroxidase: a computational, spectroscopic and site-directed mutagenesis study," *Biochem. J.*, vol. 466, no. 2, pp. 253–262, Mar. 2015.
- [7] A. Hosseini, M. Brouk, M. F. Lucas, F. Glaser, A. Fishman, and V. Guallar, "Atomic Picture of Ligand Migration in Toluene 4-Monooxygenase," *J. Phys. Chem. B*, vol. 119, no. 3, pp. 671–678, Jan. 2015.
- [8] R. Takahashi, V. A. Gil, and V. Guallar, "Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations," *J. Chem. Theory Comput.*, vol. 10, no. 1, pp. 282–288, Dec. 2013.
- [9] I. Bahar, T. R. Lezon, A. Bakan, and I. H. Shrivastava, "Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins," *Chem. Rev.*, vol. 110, no. 3, pp. 1463–1497, Mar. 2010.
- [10] I. Bahar, T. R. Lezon, L.-W. Yang, and E. Eyal, "Global dynamics of proteins: bridging between structure and function," *Annu. Rev. Biophys.*, vol. 39, pp. 23–42, 2010.
- [11] E. B. Wilson, "A Method of Obtaining the Expanded Secular Equation for the Vibration Frequencies of a Molecule," *J. Chem. Phys.*, vol. 7, pp. 1047–1052, Nov. 1939.
- [12] N. G. Tosiuyuki Noguti, "Dynamics of Native Globular Proteins in Terms of Dihedral Angles," *J. Phys. Soc. Jpn. - J PHYS SOC JPN*, vol. 52, no. 9, pp. 3283–3288, 1983.
- [13] V. Choi, "On updating torsion angles of molecular conformations," *J. Chem. Inf. Model.*, vol. 46, no. 1, pp. 438–444, Feb. 2006.

# Cation- $\pi$ -cation interactions in structural biology

Silvana Pinheiro,<sup>1</sup> Ignacio Soteras,<sup>1</sup> José Luis Gelpí,<sup>2</sup> François Dehez,<sup>3</sup> Christophe Chipot,<sup>3,4</sup> F. Javier Luque,<sup>1</sup> and Carles Curutchet<sup>1</sup>

<sup>1</sup>Departament de Físicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Spain

<sup>2</sup>Departament de Bioquímica y Biología Molecular, Faculta de Biología, Universito de Barcelona, Spain

<sup>3</sup>Laboratoire International Associé Centre National de la Recherche Scientifique et University of Illinois at Urbana –Champaign, Unité Mixte de Recherche No. 7565, Université de Lorraine, Vandoeuvre-lès-Nancy cedex, France

<sup>4</sup>Beckman Institute for Advanced Science and Engineering, University of Illinois at Urbana – Champaign, Urbana, Illinois (USA).

Corresponding author: \* Carles Curutchet; [carles.curutchet@ub.edu](mailto:carles.curutchet@ub.edu);

Phone: (+34) 934.029.00 Fax: (+34) 934.035.987

**Abstract** - Biological structures are stabilized by a variety of noncovalent interactions, such as hydrogen bonds,  $\pi$ -stacking, salt bridges or hydrophobic interactions. Besides hydrogen bonds and  $\pi$ -stacking, cation- $\pi$  interactions between aromatic rings and positively charged groups have emerged as one of the most important interactions in structural biology. Although the role and energetic characteristics of these interactions is well established, a special case involving three molecular species, termed cation- $\pi$ -cation interaction, is still poorly understood. In this contribution, we aim at advancing in the understanding of cation- $\pi$ -cation interactions and their role in the structure and stability of biosystems via two complementary approaches. The first one consists of a statistical study of the occurrence, composition and geometry of cation- $\pi$ -cation interactions identified on a non-redundant set of protein structures from the PDB (Protein Data Bank), which demonstrates that cation- $\pi$ -cation interactions are indeed common in proteins. We also analyze the degree of conservation of the interactions by inspection of similar sequences obtained through the sequence alignment tool BLAST. The second part of the study consists of an energetic analysis of the most relevant interactions at the SCS-MP2/CBS level of theory, as well as an energy decomposition analysis for representative cases performed at the SAPT2 level. Besides the well-known deficiency of standard additive force fields to describe the relevant polarization contribution to cation- $\pi$  interactions, our results indicate that non-additive three-body contributions are significant in this case, implying that cation- $\pi$ -cation interactions constitute an even more challenging case expected to be dramatically misrepresented by standard additive force fields. In vacuum, the interactions identified are strongly repulsive. Further work is being carried out in order to understand the strength of cation- $\pi$ -cation interactions in a protein environment, where the cation-cation repulsion will be strongly screened.

**Keywords:** Cation- $\pi$  interaction; interaction energy; cation- $\pi$ -cation interactions; three-body; non-polarizable FFs.

## I. INTRODUCTION

The understanding of intermolecular forces is extremely important to understand the behavior of chemical systems at the molecular level. It should be noted that such interactions and understanding gain its maximum expression in biological systems. Molecules of life (DNA, RNA, proteins, etc.) are held in their three-dimensional structures by intra and intermolecular interactions. Thus, the three-dimensional molecular structure is responsible for the specific biological activity of these molecules; note the importance of understanding of such interactions. Moreover, the correct description in the classical force field for molecular simulations used is a key to enable a realistic description of such systems.

Hydrogen bonds,  $\pi$ -stacking, salt bridges, and the hydrophobic effect all play roles in folding a protein and establishing its final structure. In addition, the cation- $\pi$  interactions is a non-covalent interaction between positively charged atoms or groups and a  $\pi$ -electron system have recognized as an important non-covalent binding interaction relevant to structural biology (1-5). The presence of a positive charge near a highly polarizable  $\pi$ -system leads to significant polarization effects, not properly described by standard non-polarizable FFs (6-10). Like this hydrogen bonding and  $\pi$ - $\pi$  stacking interactions, the cation- $\pi$  interactions have emerged as the fundamental forces that control the structure and function of macromolecules (2).

A number of studies have established a role for cation- $\pi$  interactions in biological recognition and the contribution of polarization in such interactions are very important, and may even surpass electrostatic contribution, from a monopole (cation) interacting with a quadruple (aromatic ring) (3). This is explained by the presence of a positive charge near one system as the highly polarizable  $\pi$  system. This type of interaction is common in protein structures.

Although the role and energetic characteristics of these interactions is well established, a special case involving three molecular species, termed cation- $\pi$ -cation interaction, for example, Arg-Phe-Lys interactions, is still poorly understood, are simultaneous interaction between a  $\pi$ -electron system, the aromatic sidechains of phenylalanine (Phe, F), tyrosine (Tyr, Y) or tryptophan (Trp, W) and two cations than occur between the cationic side chains of either lysine (Lys, K) or arginine (Arg, R)). In which the polarization effect are even more complex due to its nature of the three-body interactions.

In this contribution, we aim at advancing in the understanding of cation- $\pi$ -cation interactions and their role in the structure and stability of biosystems via two complementary approaches. A structural analysis, statistical study of frequency, composition and structure of cation- $\pi$ -cation interactions identified on a non-redundant set of proteins from the PDB and energetic analysis, SCS-MP2/CBS high-level calculations of cation- $\pi$ -cation interaction energies for the most relevant interactions identified.

## II. COMPUTATIONAL METHODS

### Identification and selection of cation- $\pi$ -cation interactions

In this section, we detail our strategy for identifying and ranking cation- $\pi$ -cation interactions in proteins. To conduct the analysis of cation- $\pi$ -cation interactions was performed over a representative list of proteins of Cluster90

database. Given the big amount of cation- $\pi$ -cation interactions complexes found and characterized in this database which consists of 20.759 non-redundant protein structures taken from the PDB, criterions accurate was used. A distance cutoff of  $d_1+d_2 < 7,5 \text{ \AA}$  and angle cutoff:  $\alpha_1+\alpha_2 < 30^\circ$ . The letter is defined between the normal of the ring and a vector between the centre of the ring and the positively charged site was considered. In case of Lys and Arg, the distance is taken to N and C atoms respectively. In case of Trp, both rings are considered, and in the case of His, the two nitrogens of the five membered ring were considered. The results were filtered in order to remove duplicities in each interaction due to symmetry considerations within the pdb.

In order to do that, an analysis of geometries was carried out looking up the distribution of distances and angles found in the interactions. Based on this criterion we selected 51 interactions for energetic analysis. Were classified the trimers systems in accordance with the type of aromatic (Phe, Tyr and Trp) and cationic (Lys, Arg and His) aminocids, as well as ions ( $\text{Na}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$ ,  $\text{Li}^+$ ,  $\text{Cu}^{2+}$ ). Among all the structures analyzed, the representative structure of cation- $\pi$ -cation interactions multiple is 1A22 (Human growth hormone), that is a special case with three consecutive cation- $\pi$ -cation interactions K379-W386-R411, R411-F425-R413, R413-Y422-K415.

Calculation of conservation and accessibility

We have evaluated the conservation of cation- $\pi$ -cation interactions in each protein with the aid of the BLAST (Basic Local Alignment Search Tool). That compares the query sequence against a large number of protein sequences, which search database non-redundant protein sequences (nr) using algorithm BLASTp (protein-protein BLAST)<sup>11</sup>. To investigate the conservation of interaction in proteins, was considered sequence of proteins above >80% of identity. For example, in an alignment of 20 sequences that all share >90% pairwise identify conservation may be interesting. In contrast, if the pairwise identify is below 30% then lower conservation scores will be informative.

The computations of solvent accessibility and protonation of residues involved in cation- $\pi$ -cation interaction were calculated by NACCESS and ProPka programs, that calculates the accessible area of a molecule from PDB. The pattern was categorized into buried, partially buried and exposed based on ranges of accessible surface area.

### Construction of the model Systems

From the representative pdb coordinates was built simplified model systems of each amino acid aromatic and cationic on which we have carried out theoretical calculations to study cation- $\pi$ -cation interactions. Each trimer then was reduced to a system that could be studied computationally. Lys and Arg were represented as ammonium and guanidinium ions, and Phe, Tyr, and Trp were represented as benzene, phenol, and indole, respectively. The geometries of the structures of such monomers were optimized at *MP2/aug-cc-pVDZ* level using the Gaussian 09 suite of programs. Thereafter was performed the alignment the structures, overlapping the optimized structures of the monomers with the structures of pdb reference.

### Calculation of interaction energy

To evaluate the non-additive contribution to three bodies interactions in cation- $\pi$ -cation by comparing the contributions, we conducted the calculations the interaction energies of both systems as models and pdb's, apply different basis sets incorporating polarization and diffuse functions, a total of 6 basis sets were also tested to examine the convergence of the interaction energy with basis set size, we have used the basis sets 6-31G, 6-31+G(d), 6-31++G(d,p), 6-311++G(d,p) and also the correlation consistent basis sets aug-cc-pVDZ and aug-cc-pVTZ, the last two were considered as the reference to calculate the errors on the interaction energies obtained with the other basis sets. The calculations were developed using the ab initio level SCS/MP2, the new method improves upon MP2, thereby rectifying many of its problems, indicates significant robustness and suggests it as a valuable quantum chemical method of general use.<sup>12</sup>

$$E_{SCS/MP2} = E_{HF} + 1/3(E_{corr(\alpha-\alpha)} + E_{corr(\beta-\beta)}) + 6/5E_{corr(\alpha-\beta)}$$

In principle, it could be possible to extrapolate the energies of the systems (and hence the interaction energy) to the complete basis set limit (CBS) using the Dunning DZ-TZ basis set series, will combine the calculations MP2/cc-pVDZ and MP2/cc-pVTZ level by formula Helgaker.<sup>13-15</sup> Such a formula allows obtain CBS  $E_{XY}^\infty$  energy, combining the calculated energies cc-pVXZ level according to the expression.

$$E_{XY}^\infty = \frac{E_x^{corr} X^3 - E_x^{corr} Y^3}{X^3 - Y^3}$$

To obtain more accurate energetics, were applied and basis set superposition errors (BSSE) were subtracted from the computed BDEs using the correction counterpoise (CP), that is calculations of monomers using the base set trimer. Counterpoise corrections for basis set superposition error (BSSE) were included in the interaction energy calculations, using the Gaussian 09 software.

$$E_\infty^{tot} = \frac{3\alpha}{3\alpha-2\alpha} E_3^{HF} - \frac{2\alpha}{3\alpha-2\alpha} E_2^{HF} + \frac{3\beta}{3\beta-2\beta} E_3^{corr} - \frac{2\beta}{3\beta-2\beta} E_2^{corr}$$

The magnitude of the different components of the interaction energy was examined by means of the Symmetry-adapted perturbation theory (SAPT) as implemented in the SAPT2006 program. In particular, we have used the SAPT2 computational scheme, which should provide an interaction energy  $U_{tot}^{SAPT2}$  equivalent to the supermolecular MP2 one. Within the SAPT2 framework the interaction energy is decomposed as shown in eqn, where the first four terms stand for the electrostatic  $U_{ele}$ , induction  $U_{ind}$ , exchange  $U_{exch}$  and dispersion  $U_{disp}$  components. In addition, the SAPT expansion yields two terms corresponding to the coupling between exchange and induction  $U_{exch-ind}$  and between exchange and dispersion  $U_{exch-disp}$ . Finally, the last term  $\delta HF$  accounts for a collection of higher order induction and exchange-induction terms.<sup>16</sup>

$$U_{tot}^{SAPT2} = U_{ele} + U_{ind} + U_{exch} + U_{disp} + U_{exch-ind} + U_{exch-disp} + \delta HF$$



### III. RESULTS AND DISCUSSION

Using the above criteria, we scanned a larger dataset of representative protein crystal structures, taken from the Protein Data Bank (PDB). Were identified in the *Cluster90* database 20.759 non-redundant protein structures. The results point out to 2.328 structures which have one or more cation- $\pi$ -cation interaction corresponding 11.3%. In total, 2.898 interactions were identified.

#### Structural analysis

- Among the cation- $\pi$ -cation interactions identified on the PDB, the frequency of aromatic amino acids found (Tyr > Phe > Trp) is similar to the frequency observed for 2-body cation- $\pi$  complexes.

- Because Trp amino acids are less frequent in proteins than Tyr or Phe, the comparatively large number of cation- $\pi$ -cation interactions identified suggest a preference for Trp over Tyr or Phe.

#### Energetic analysis

The non-additive contribution (3-body) is significant (up to 7 kcal/mol). Thus, cation- $\pi$ -cation interactions are expected to be badly described by standard additive FFs. In a gas phase environment, cation- $\pi$ -cation interactions are strongly repulsive, in the range 20-35 kcal/mol. Further work is needed to understand the strength of cation- $\pi$ -cation interactions in a protein environment, where the cation-cation repulsion will be strongly screened.

### IV. CONCLUSION

Our results demonstrates that cation- $\pi$ -cation interactions are indeed common in proteins and one every ten proteins contain at least one cation- $\pi$ -cation interaction. Indicate that non-additive three-body contributions are significant in this case, implying that cation- $\pi$ -cation interactions constitute an even more challenging case expected to be dramatically misrepresented by standard additive force fields. In vacuum, the interactions identified are strongly repulsive. Further work is being carried out in order to understand the strength of cation- $\pi$ -cation interactions in a protein environment, where the cation-cation repulsion will be strongly screened.

### REFERENCES

1. Ma, J. C. & Dougherty, D. A. *Chem. Rev.* **97** (1997) 1303–1324
2. A. Subha Mahadevi and G. Narahari Sastry. *Chem. Rev.* **113** (2013) 2100.
3. E. Cubero, F. J. Luque, M. Orozco. *Proc. Natl. Acad. Sci. U.S.A.* **95** (1998) 5976.
4. H. Minoux, C. Chipot. *J. Am. Chem. Soc.* **121** (1999) 10366.
5. I. Soteras, C. Curutchet, A. Bidon-Chanal, F. Dehez, J. G. Ángyán, M. Orozco, C. Chipot, F. J. Luque. *J. Chem. Theory Comput.* **3** (2007) 1901.
6. Cubero, Luque & Orozco PNAS 95, 5976 (1998)
7. Minoux & Chipot, JACS 121, 10366 (1999)
8. Soteras et al. JCTC 3, 1901 (2007)
9. François Dehez, János G. Ángyán, Ignacio Soteras, F. Javier Luque, Klaus Schulten, and Christophe Chipot; Modeling Induction Phenomena in Intermolecular Interactions with an Ab Initio Force Field *J. Chem. Theory Comput.* 2007, 3, 1914-1926.
10. Ignacio Soteras, Modesto Orozco and F. Javier Luque; Induction effects in metal cation–benzene complexes. *Phys. Chem. Chem. Phys.*, 2008, 10, 2616–2624.
11. Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers and David J. Lipman; Basic Local Alignment Search Tool. *J. Mol. Biol.* 1990, 215, 403-410.
12. Grimme, 2003. *The Journal of Chemical Physics* 118 (9095).
13. ASGER HALKIER, TRYGVE HELGAKER, POUL JORGENSEN, WIM KLOPPER, HENRIK KOCH, JEPPE OLSEN, ANGELA K. WILSON. Basis-set convergence in correlated calculations on Ne, N2 and H2O. *Chemical Physics Letters*, 286 (1998) 243-252.
14. D. G. Truhlar. (1998), *Chemical Physics Letters* . **294** (1998) 45-48.
15. T. Helgaker, Trygve; et al (1997). *The Journal of Chemical Physics*. 106 (9639).
16. Jerzierski, Moszynski, Szalewicz *Chem. Rev.* (1994), 94, 1887.

# Heating Bulk Ions in DEMO with ICRF waves

Dani Gallart  
Barcelona Supercomputing Center  
daniel.gallart@bsc.es  
Mervi Mantsinen

ICREA/ Barcelona Supercomputing  
Center  
mervi.mantsinen@bsc.es  
Yevgen Kazakov

Laboratory for Plasma Physics LPP-  
ERM/KMS

**Abstract**-Ion cyclotron resonance frequency heating (ICRF) is one of the auxiliary heating schemes presently envisaged for ITER and DEMO. In this paper we analyse the potential of ICRF waves to heat the fuel ions in DEMO. Our analysis is carried out for the EU DEMO design<sup>1</sup> ( $B = 6.8$  T,  $I = 18.6$  MA,  $R = 9.25$  m,  $a = 2.64$  m) optimized for a maximum pulse length of 2.3 hrs using the ICRF modelling codes PION and TORIC [2, 3]. We focus on second harmonic heating for tritium and fundamental minority heating of <sup>3</sup>He with a few percent of <sup>3</sup>He in a 50%:50% D-T plasma. The dependence of the ICRF characteristics and the ICRF-accelerated ions on the ICRF and plasma parameters is investigated, giving special attention to the DEMO design point at a plasma temperature of 30 keV and an electron density of  $1.2 \cdot 10^{20} \text{m}^{-3}$ .

## I. INTRODUCTION

The plasma is a gas made of electrons and ions and, therefore, is a really good conductor. When plasma is heated at a temperature of the order of 10 keV - 100.000.000 °C fusion reactions begin to occur delivering an important quantity of energy. In order to produce fusion reactions, plasma needs to be confined inside special devices. For the magnetic confinement method this is achieved via the application of a magnetic field that restricts the trajectory of the plasma particles due to the Lorentz force.

### A. Ion Cyclotron Resonance Frequency (ICRF) Heating

For a tokamak fusion reactor, heating fuel ions to thermonuclear temperatures is of vital importance. Ion cyclotron resonance frequency (ICRF) heating is an auxiliary mechanism for heating the plasma in a fusion reactor, i.e. a tokamak. ICRF is based on launching electromagnetic waves from the low-field side (outer side of the tokamak). Wave-particle resonance occurs when the parallel Doppler shifted frequency of the wave is equal to an exact harmonic of the cyclotron frequency of the particle, i.e.  $\omega = k_{\parallel}v_{\parallel} + n\omega_{ci}$ , where  $\omega$  is the wave frequency,  $\omega_{ci} = q_i B(r)/(A_i m_p)$  ( $q_i$  is the charge,  $B(r)$  the background magnetic field,  $A_i$  the mass number and  $m_p$  the proton mass) is the ion cyclotron frequency and  $n = 1$  for the fundamental and  $n > 2$  for higher harmonics. When resonance occurs, ions start damping the wave by absorbing its energy. This effect modifies the distribution function of ions which develops a tail in the high energy region. The fast ions produced by the energy absorption from the electromagnetic waves play an important role in heating the bulk plasma. Therefore, it is crucial to know how the energy of the wave is distributed among ions and electrons, and how the fast ions produced deliver their energy to the other particles, ions and electrons. The critical energy is the energy at which ions transfer energy equally to background electrons and ions, for energies larger than  $E_c$  collisions with background electrons

predominate while for energies lower than  $E_c$  bulk ion heating is obtained.

$$E_c = 14.8AT_e \left[ \sum_j \frac{n_j Z_j^2}{n_e A_j} \right]^{\frac{2}{3}} \quad (1)$$

## II. DEMONSTRATION POWER PLANT (DEMO)

The DEMONstration power plant is a proposed nuclear fusion power plant that is expected to be built after the experimental reactor ITER. ITER's main purpose is to confirm the feasibility of nuclear fusion as a source of net electrical energy while DEMO will be the first fusion reactor to produce electrical energy.

TABLE I  
DEMO, ITER, AND JET PARAMETERS

Parameter	DEMO <sup>1</sup>	ITER	JET
Major radius $R_o$ (m)	9.25	6.2	2.96
Minor radius $r$ (m)	2.64	2	1.25-2.10
Tor. Magnetic field $B$ (T)	6.8	5.3	3.45
Plasma current $I_p$ (MA)	18.6	15	4.8
Safety fac. $q0$ and $q95$	1.1, 3	1.0, 3.5	1.0, 5.0
Elongation $\kappa$	1.52	1.7	1.68
Triangularity $\delta$	0.33	0.33	0.4

The Joint European Torus (JET) is the fusion reactor located in Culham, UK. It is nowadays the biggest fusion reactor in the world.

## III. ANALYSIS OF BULK ION HEATING IN DEMO

We concentrate our studies on the second harmonic ( $n=2$ ) ICRF heating of tritium with and without <sup>3</sup>He in a 50%:50% D-T plasma.

### A. ICRF scenario

The scenario is for a standard midplane launch with thermal plasma (there is no ICRF+NBI interaction) and the toroidal mode number  $N = Rk_{\parallel}$  is fixed. The basic parameters are shown in the following table (Table II).

<sup>1</sup>We are grateful to Dr. T. Franke and Dr. R. Wenninger (PPPT, Garching) for DEMO parameters. They are for the EU DEMO design optimised for a maximum pulse length of 2.3 hrs (with CD 2.7hrs).

Although there are many cases studied, there are two of them that are analyzed deeper, they correspond to  $T = 30$  keV

and for densities  $n_e = 10^{20}$  and  $n_e = 1.2 \cdot 10^{20}$ . The case with  $n_e = 1.2 \cdot 10^{20}$  and  $T = 30 \text{ keV}$  is the DEMO design point.

TABLE II  
ICRF PARAMETERS AT DEMO AND ITER

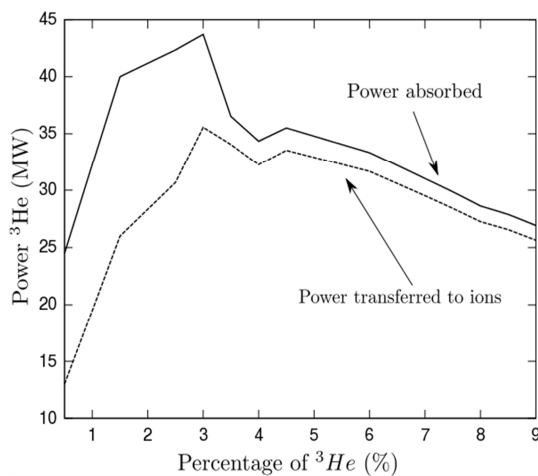
Parameter	DEMO
Toroidal magnetic field $B$ (T)	ICRF frequency (MHz)
ICRF frequency $f$ (MHz)	66, 70, 74
Resonance location ( $r_{res}/a$ )	0.2 (LFS), -0.05, -0.2 (HFS)
ICRF Power (MW)	100
Toroidal mode number	50

### B. Minority heating with $^3\text{He}$

In minority heating typically the plasma contains a few percent in concentration of an ion species which interacts with the ICRF waves. In this case, the frequency of the wave has been set at the fundamental harmonic of  $^3\text{He}$ ,  $\omega = \omega_{^3\text{He}} = 2\omega_T$ .

The analysis has been carried out for a scan in the concentration of  $^3\text{He}$  (see Fig. 1), the total electron density  $n_e$ , temperature  $T$ , and wave frequencies (central and off-axis).

Fig. 1. Power absorbed by  $^3\text{He}$  and transferred to bulk ions (MW) for each percentage of  $^3\text{He}$ . ICRF:  $N = 50$ ,  $f = 70 \text{ MHz}$ ,  $T = 30 \text{ keV}$ ,  $P = 100 \text{ MW}$ .



For increasing concentration of  $^3\text{He}$  until 3-5% the power absorbed by  $^3\text{He}$  increases substantially while after a concentration of 5% the polarization of the wave starts to become less favorable and, therefore, the absorption power decreases, the power not absorbed by  $^3\text{He}$  goes directly to the electrons. In this case, for the DEMO design point, the power that actually remains in the bulk ions population is almost the same between 3-5%.

### C. Second harmonic tritium heating scenario

The second harmonic tritium heating scenario is a 50%:50% D-T plasma heated by ICRF waves with a frequency equal to the second harmonic of T, i.e.  $\omega = 2\omega_T$ . In this case fast ions become more energetic than in the minority heating and the critical energy  $E_c$  is easier to be reached by fast ions. The results achieved are presented in Table III, here the efficiency ( $\eta$ ) stands for the ratio between the power transferred and the

power absorbed by  $^3\text{He}$  and T respectively. Different frequencies produce different locations for the resonance region (see ion cyclotron formula in section I. A).

TABLE III  
EFFICIENCY FOR DIFFERENT SCENARIOS WITH AN ICRF OUTPUT POWER OF 100 MW AT THE DEMO DESIGN POINT

Composition	66 MHz	70 MHz	74 MHz
$^3\text{He}$ 3%	55.84	35.56	26.65
$^3\text{He}$ 4%	54.80	32.25	24.37
$^3\text{He}$ 5%	48.10	31.90	23.50
$^3\text{He}$ 6%	45.20	31.76	22.54
T	43.00	26.65	15.51

## IV. CONCLUSION

We have simulated ICRF heating for DEMO plasmas. Simulations show that for minority heating a concentration of 3-5% of  $^3\text{He}$  matches the highest efficiency while by placing the resonance region slightly off-axis (at 0.2a) ICRF heating is considerably enhanced in both scenarios avoiding a substantial fraction of direct electron damping. Notice that although minority heating shows a better efficiency the plasma dilution decreases the fusion reaction yield.

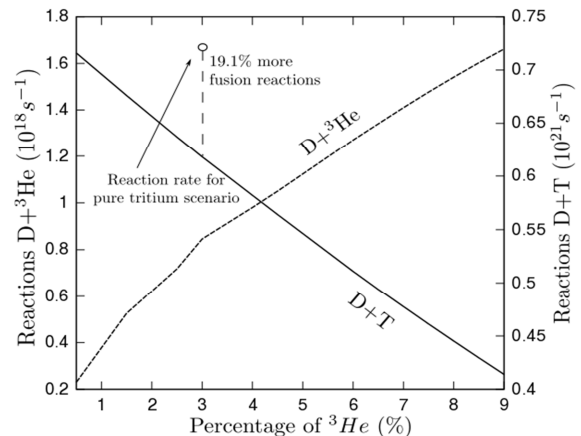


Fig. 2. Comparison of fusion reaction rate between  $^3\text{He}$  and T scenarios.

## REFERENCES

- [1] T. H. Stix, *Fast wave heating of a two component plasma*, Nuclear Fusion **15** (1975) 737.
- [2] L.-G. Eriksson, T. Hellsten and U. Willén, Nuclear Fusion **33** (1993) 1037
- [3] M. Brambilla and R. Bilato, Nuclear Fusion **46** (2006) S387.
- [4] L.-G. Eriksson and P. Helander, Physics of plasmas **1** (1994) 308

# Rayleigh wave ellipticity measurements in the Iberian Peninsula and Morocco

Clara Gómez-García<sup>1</sup>, Antonio Villaseñor<sup>1</sup>  
<sup>1</sup>Institute of Earth Sciences “Jaume Almera” (CSIC)  
cgomez@ictja.csic.es, antonio@ictja.csic.es

**Abstract-** We combine Rayleigh wave ellipticity, or H/V (horizontal to vertical) amplitude ratios, measurements obtained using teleseismic earthquake recordings and ambient noise cross correlations to provide improved constraints on the crustal models across the Iberian Peninsula and Morocco. To obtain the H/V ratios, we use more than 250 shallow ( $h < 40$  km) teleseismic events with magnitudes  $M_w > 6.0$  recorded at more than 450 seismic stations. We also use all the multicomponent ambient noise cross correlations computed for each station pair for the time period 2010 to 2012. Periods between 20 and 100 seconds are investigated. We observe a good agreement between the uppermost geological features of the crust and the obtained Rayleigh H/V ratios, with low values in major mountain ranges and high ratios in sedimentary basins. Combination of Rayleigh H/V ratio measurements from both earthquakes and ambient noise data with phase velocities and other types of seismic data will help to better constrain the Earth's structure at different crustal levels.

## I. INTRODUCTION

Rayleigh waves are distinguished among other things by an elliptic trajectory of particle motion in the vertical-radial plane. Rayleigh wave ellipticity, defined as the ratio of horizontal to vertical spectral amplitudes (H/V ratio from now on for simplicity), has substantial shallow earth sensitivity since it depends on the local crustal structure.

Traditionally, ambient noise combined with the H/V spectral ratio have usually been used in microzonation studies, to characterize site response, predict ground motion or invert for S velocity (e.g., Ref. [4] and [9]). However, the interpretation of the H/V ratios in these studies depends on the noise composition. Measurements from ambient noise predominantly consist on fundamental mode Rayleigh waves for low frequency microseisms ( $< 1$  Hz) but for higher frequency microtremors ( $> 1$  Hz) there is no consensus. The microtremor H/V ratio has no clear physical origin because it depends on the contents of noise signals and the relative amplitudes among different wave types (see Ref. [2] and [3] for a review).

Studies of Rayleigh H/V ratios from earthquake signals for long periods ( $> 20$  sec) are less common (e.g., Ref. [5], [6] and [10]) because it has been reported considerable variability in the observations due to the influence of small scale heterogeneity [5]. In spite of the large scatter observed in Rayleigh H/V ratios for long periods, Reference [10] described stable estimates after collecting statistics for numerous earthquakes.

Recently, a new approach has been provided to study the crustal structure using Rayleigh wave ellipticity for earthquake

signals or ambient noise cross correlations [6][7][8][11]. As Rayleigh wave ellipticity provides independent information from phase velocity data, a joint inversion will improve the inversions based only on phase velocity. In this work, we perform a Rayleigh wave H/V ratio study using teleseismic earthquake recordings and multicomponent ambient noise cross correlations with the aim of obtaining improved constraints on the uppermost crustal structure in the Iberian Peninsula and Morocco.

## II. DATA ANALYSIS

We compute Rayleigh H/V ratios for more than 250 teleseismic events between 2010 and 2012 by more than 450 stations from several seismic networks deployed across the Iberian Peninsula and Morocco.

Focusing on the study of teleseismic signals, we use shallow earthquakes ( $h \leq 40$  km) to minimize overtone contamination and large enough (magnitudes  $M_w > 6.0$ ) to ensure the excitation of low-noise, long period, mantle waves [5]. To compute the H/V ratios at each station location we first rotate the three component recordings, for each earthquake and station, from East-West, North-South and vertical to transverse, radial and vertical components. Then, we remove the instrument response and apply a band-pass filter in the frequency band 0.02 to 20 Hz. Differences in the sampling rates for different instruments are taken into account for the analysis. As we have a large collection of records, we compute the H/V ratios using a frequency-time analysis (FTAN, [1]) applied in an automated way.

## III. RESULTS AND CONCLUSIONS

Rayleigh H/V ratios are related to the elastic properties of materials and velocity gradients in the upper crust. Zones such as sedimentary basins are associated to high H/V ratios whereas low H/V measurements are related to major mountain ranges. In Fig. 1 we show the Rayleigh H/V ratios estimated for more than 450 sites in the Iberian Peninsula and Morocco for 30 seconds of period. We observe high H/V ratios in sedimentary basins (Cenozoic and Foreland basins) as Guadalquivir and Ebro Basins and low H/V ratios in Pyrenees, Iberian Massif and High Atlas. Although the standard deviation for each average H/V ratio is, in general, very small, we would need to consider a larger collection of teleseismic signals to make the results statistically significant [10].

The combination of Rayleigh H/V ratio measurements using both earthquakes and ambient noise data with phase velocity and other types of seismic data will provide us with a more realistic interpretation of the upper crust structure across the

Iberian Peninsula and Morocco and a better understanding of the relationship between geologic features at different crustal levels. Moreover, the inclusion of short-period Rayleigh wave H/V ratio measurements based on noise cross correlations in the H/V ratios obtained from earthquakes enables to resolve the structure in the uppermost 1 km since they are more sensitive to near-surface structure.

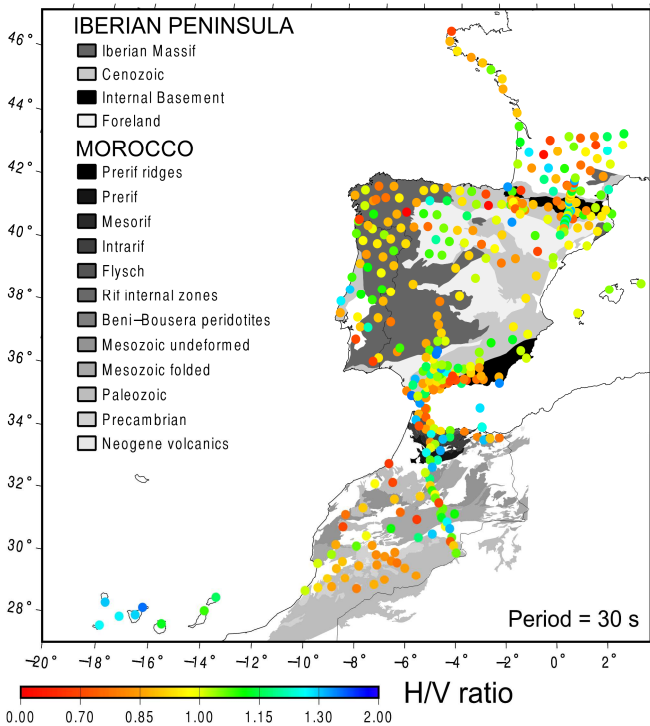


Fig. 1. 30 s Rayleigh wave H/V ratio across the Iberian Peninsula and Morocco.

#### ACKNOWLEDGMENT

We are funded by REPSOL CO-DOS project and the data used in this research were a contribution of Team Consolider-Ingenio 2010 TOPO-IBERIA (CSD2006-00041).

#### REFERENCES

[1] G.D. Bensen, M.H. Ritzwoller, M.P. Barmin, A.L. Levshin, F. Lin, M.P. Moschetti, N.M. Shapiro, and Y. Yang, "Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements," *Geophys. J. Int.*, 169, 1239–1260, 2007. doi: 10.1111/j.1365-246X.2007.03374.x

[2] S. Bonnefoy-Claudet, C. Cornou, J. Kristek, M. Ohrnberger, M. Whatheler, P.Y. Bard, D. Fäh, P. Moczo and F. Cotton, "Simulation of seismic ambient vibrations: I. H/V and array techniques on canonical models," *Proceedings of the 13<sup>th</sup> World Conference on Earthquake Engineering*, Vancouver, Canada. Paper # 1130, 2004.

[3] S. Bonnefoy-Claudet, F. Cotton and P.Y. Bard, "The nature of the seismic noise wave field and its implication for site effects studies: A literature review," *Earth Sci. Rev.*, 79(3–4), 205–227, 2006.

[4] D. Fäh, F. Kind, and D. Giardini, "A theoretical investigation of average H/V ratios", *Geophys. J. Int.*, 145, 535–549, 2001. doi: 10.1046/j.0956-540x.2001.01406.x

[5] A.M.G. Ferreira and J.H. Woodhouse, "Observations of long period Rayleigh wave ellipticity", *Geophys. J. Int.*, 169, 161–169, 2007. doi: 10.1111/j.1365-246X.2006.03276.x

[6] F. C. Lin, B. Schmandt and V.C. Tsai, "Joint inversion of Rayleigh wave phase velocity and ellipticity using USArray: Constraining

velocity and density structure in the upper crust," *Geophys. Res. Lett.*, 39, L12303, 2012. doi: 10.1029/2012GL052196

[7] F.C. Lin, V.C. Tsai and B. Schmandt, "3-D crustal structure of the western United States: application of Rayleigh-wave ellipticity extracted from noise cross-correlations," *Geophys. J. Int.*, 198, 656–670, 2014. doi: 10.1093/gji/ggu160

[8] F. Lin and B. Schmandt, "Upper crustal azimuthal anisotropy across the contiguous U.S. determined by Rayleigh wave ellipticity," *Geophys. Res. Lett.*, 41, 8301–8307, 2014. doi: 10.1002/2014GL062362

[9] Y. Nakamura, "A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface," *Q. Rep. Railw. Tech. Res. Inst.*, 30, 25–30, 1989.

[10] T. Tanimoto and L. Rivera, "The ZH ratio method for long-period seismic data: sensitivity kernels and observational techniques," *Geophys. J. Int.*, 172, 187–198, 2008. doi: 10.1111/j.1365-246X.2007.03609.x

[11] T. Tanimoto, T. Yano and T. Hakamata, "An approach to improve Rayleigh-wave ellipticity estimates from seismic noise: application to the Los Angeles Basin," *Geophys. J. Int.*, 193, 407–420, 2013. doi: 10.1093/gji/ggs123

# State-of-the-Art Climate Predictions for Energy Climate Services

Verónica Torralba<sup>1</sup>, Melanie Davis<sup>1</sup> and Francisco Doblas-Reyes<sup>1,2,3</sup>, Nube González-Reviriego<sup>1</sup>

(1) Institut Català de Ciències del Clima (IC3), Barcelona, Spain

(2) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

(3) Barcelona Supercomputing Center (BSC), Barcelona, Spain

veronica.torralba@ic3.cat

**Abstract-** Seasonal predictions of 10-m wind speed can be used by the wind energy sector in a number of decision making processes. Two different techniques of post-processing are applied in order to correct the unavoidable systematic errors present in all forecast systems. Besides an assessment of the impact of these corrections on the quality of the probabilistic forecast system is provided.

Key words: wind energy, seasonal forecasts, bias-correction, forecast verification

## A. INTRODUCTION

The need of accurate forecasts becomes increasingly important in the wind energy sector due to the intermittent nature of wind power generation, and the need to match supply with demand at all times [1].

Climate predictions of time scales from a month to decades in the future, and that are tailored to the wind energy sector represent the cutting edge in climate sciences to forecast wind power generation. The seasonal prediction addresses a long list of challenges to produce climate information that responds to the expectations of the users [2]. At these time scales, current energy practices use a deterministic approach based on retrospective climatology, but seasonal predictions have recently been shown to provide additional value.

In particular probabilistic climate predictions of near surface winds can allow end users to take calculated, precautionary action with a potential cost savings to their operations. Electricity system operators can use these predictions to adapt energy supply availability from wind farms, and allow the electric network to conveniently adapt demand and resources [3].

For this reason our main goal is to inform users, with greater accuracy than their current approach, what will be the most likely range of wind speed of the upcoming years. This study analyses the ECMWF S4 seasonal forecast system for wind speed to assess the quality of these predictions and its properties.

## B. DATA AND METHODS

The forecasts of 10-m wind speed from the ECMWF S4 model on winter (December, January and February) with a start date of the 1<sup>st</sup> of November are used, because in this season the wind speed variability is higher. The 10-m wind

speed data from ERA-Interim have been also used as reference dataset.

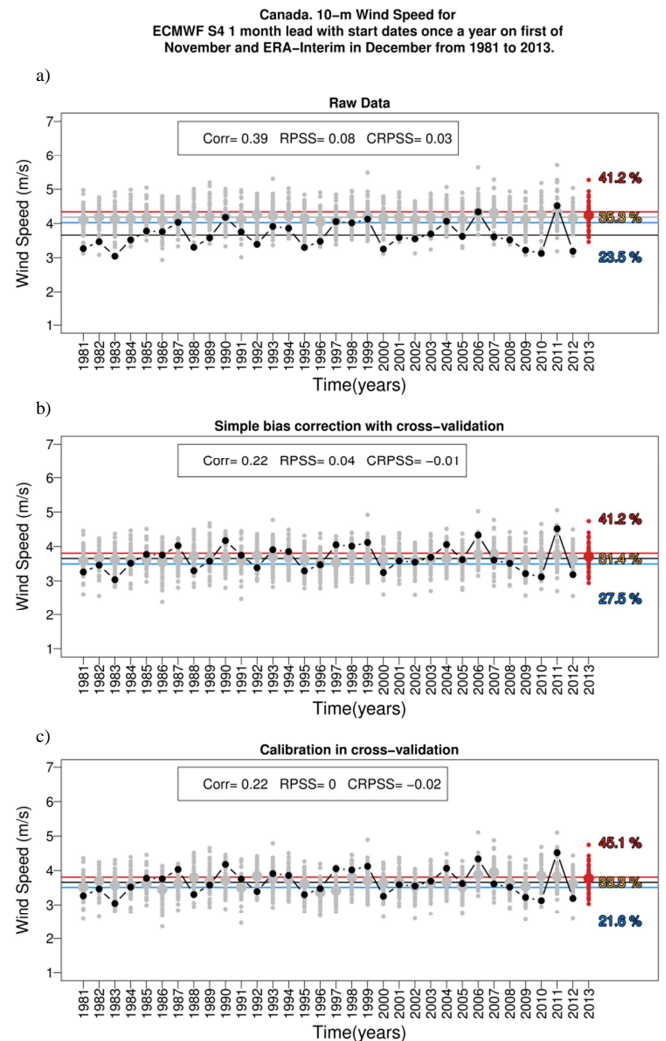


Figure 1. Time series of 10-m wind speed i: a) Raw data, b) Simple Bias corrected and c) Calibrated, in the period 1981-2013. The ensemble members of the hindcasts are represented as small grey dots and the ensemble mean is represented with a large grey dot for each start date. The grey horizontal line shows the mean of the hindcast in whole period. The black dots represent the 10-m wind speed values of ERA-Interim. The black horizontal line shows the mean of the reference in whole period. The blue and red horizontal lines show the lower and upper terciles, respectively. The ensemble members of the forecast year are represented as small red dots and the ensemble mean is represented with a large red dot. The percentages indicates the number of members in each category, which are limited by the terciles.

As with every variable predicted in a coupled model forecast system, the prediction of wind speed is affected by biases. For probabilistic forecasts, this defect consists of their lack of sufficient (probabilistic) reliability: they generally are under-dispersive [4]. To overcome this, two different techniques for the post-processing of ensemble forecasts are considered: a simple bias correction and a calibration method. Both methods use the “one-year out” cross-validated mode, and they provide corrected forecasts with improved statistical properties.

### C. RESULTS

The impact of these bias corrections on the quality of the ECMWF S4 predictions of near surface wind speed during winter is explored. To offer a comprehensive picture of the post-processing effect on the forecast quality of the system, it is necessary to use different scoring measures as skill scores (Figure 1) or rank histograms (Figure 2). As illustrated, a key region for the wind energy sector in Canada has been analysed.

Figure 1 shows that the skill scores decrease when post-processing techniques are applied. All operations performed on a forecast will increase its uncertainty. Even correcting the bias in the mean implies estimating the mean from the hindcasts, which is an estimate with its own uncertainty and is thus propagated into the forecasts

The rank histograms show if the ensemble members and the verifying observation come from the same probability distribution, in which case the forecasts are statistically consistent and the rank histogram should be flat or uniform. These histograms display the cumulative frequency in the y-axis, which provides quantitative information as to whether the deviations from reliable behavior are systematic or merely random. For the raw data (Figure 2a) a bias seems to be present as the rank histogram shows overpopulated lower ranks. The bias corrected and the calibrated rank histograms (Figures 2b and 2c, respectively) are more homogeneously populated, therefore the reliability of the ensemble improves when post-processing is applied

### D. CONCLUSIONS

This study reveals that the different techniques to correct climate predictions produce a statistically consistent ensemble. However, the operations performed decrease their skill, which would correspond to an increase in the uncertainty. Therefore, even though the bias correction is fundamental for climate services, this comes at a price in terms of forecast quality.

### REFERENCES

- [1] Füss, R., Mahringer, S., & Prokopczuk, M. (2013). Electricity derivatives pricing with forward-looking information. University of St. Gallen, School of Finance Research Paper, (2013/17).
- [2] Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., & Rodrigues, L. R. (2013). Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, 4(4), 245-268.
- [3] Weisser, D., & Foxon, T. J. (2003). Implications of seasonal and diurnal variations of wind velocity for power output estimation of a turbine: a case study of Grenada. *International Journal of Energy Research*, 27(13), 1165-1179.
- [4] Pinson, P. (2012). Adaptive calibration of (u, v) wind ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138(666), 1273-1284.

10m Wind Speed for ECMWF S4 1 month lead with start dates once a year on first of November and ERA-Interim in DJF from 1981 to 2012 in Canada.

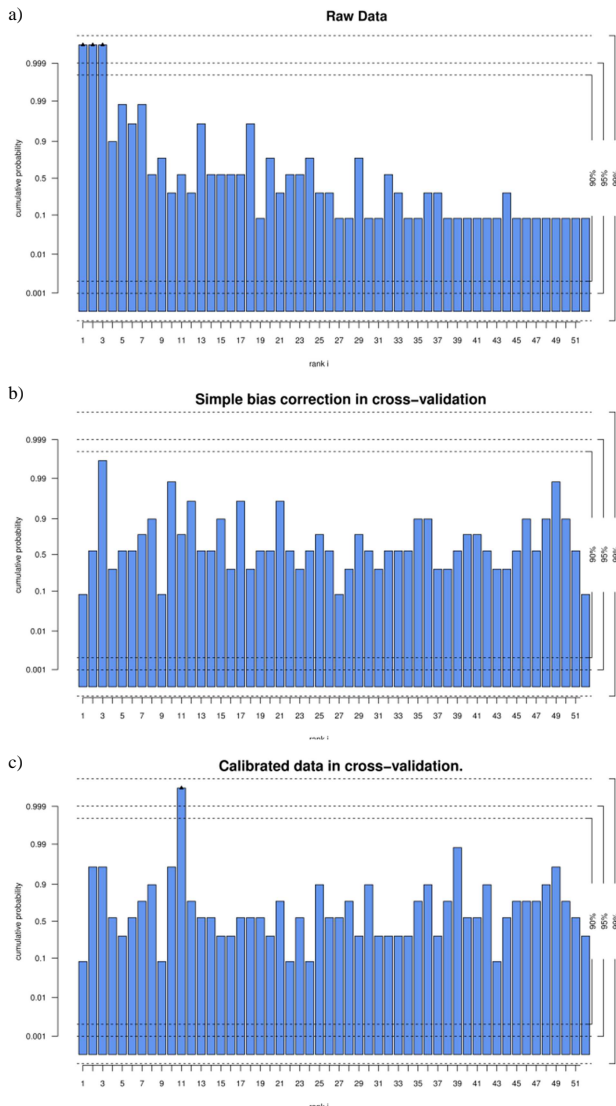


Figure 2. Rank histograms of 10-m wind speeds forecasts a) Raw data, b) Simple Bias corrected and c) Calibrated, in the period 1981-2012. The x-axis represents the rank and the y-axis the cumulative probabilities, which shows if the deviations from the reliable behaviour are systematic or random. The intervals on the right of the plot indicate central 90, 95 and 99 % simultaneous confidence intervals.

# Parallel programming issues and what the compiler can do to help

Sara Royuela\*, Xavier Martorell†

Barcelona Supercomputing Center\*, Universitat Politècnica de Catalunya†  
{sara.royuela, xavier.martorell}@bsc.es

**Abstract-** *Twenty-first century parallel programming models are becoming real complex due to the diversity of architectures they need to target (Multi- and Many-cores, GPUs, FPGAs, etc.). What if we could use one programming model to rule them all, one programming model to find them, one programming model to bring them all and in the darkness bind them, in the land of MareNostrum where the Applications lie. OmpSs programming model is an attempt to do so, by means of compiler directives.*

*Compilers are essential tools to exploit applications and the architectures they run on. In this sense, compiler analysis and optimization techniques have been widely studied, in order to produce better performing and less consuming codes.*

*In this paper we present two uses of several analyses we have implemented in the Mercurium[3] source-to-source compiler: a) the first use is to help users with correctness hints regarding the usage of the OpenMP and OmpSs tasks; b) the second use is to be able to execute OpenMP in embedded systems, with very little memory, thanks to calculating the Task Dependency Graph of the application at compile time. We also present the next steps of our work: a) extending range analysis for analyzing OpenMP and OmpSs recursive applications, and b) modeling applications using OmpSs and future OpenMP4.1 tasks priorities feature.*

**Keywords-** *Compiler, Static Analysis, Task Dependency Graph, OpenMP, OmpSs, Embedded System*

## I. INTRODUCTION

Static and dynamic analysis and optimization techniques are widely used in order to enhance performance and power consumption. Dynamic techniques benefit from a non-restricted knowledge of the application (addresses and values of all variables are known), but they require the execution of the program along with the instrumentation library. This means adding overhead, as well as having results tied to a specific execution. Other data-sets or architectures may change the results of the analysis. On the contrary, static techniques have limited information (values of the variables and pointer aliasing situations), but have the benefit of being effortless from the point of view of the programmer, as well as being valid for any input data. Neither of the options is perfect nor valid for everything, and they can indeed be combined for better results.

## II. CORRECTNESS

Although programming models such as OpenMP and OmpSs are tantalizing due to its simplicity and scalability, they also bring forth difficulties when it comes to fully exploit their capabilities. The compiler can be crucial to anticipate bugs that may be very hard to find at runtime. We focus on the OpenMP and OmpSs tasking models to define a set of situations that may cause: a) a runtime failure, b) a loss of performance, or c) a non-deterministic result. We present different cases the compiler is able to detect and the hints it

provides to the programmer. To understand these cases some background in the memory model of OpenMP [1] is needed.

A. Automatic storage variables as shared. Automatic storage variables are allocated and deallocated automatically when the program flow enters and leaves the enclosing code block. When the compiler detects that such a variable may be accessed after its scope has been exited, it proposes the following solutions: a) changing the data-sharing attribute from shared to private, for basic data types; b) adding a `taskwait`, for arrays and structures.

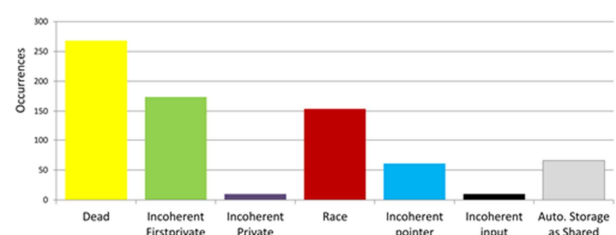
B. Data-race situations. Data-races occur when two or more threads access shared data and at least one of the accesses is a write. When the compiler detects such a situation, it proposes the following solutions: a) protecting the accesses with a critical or an atomic construct, b) adding a `taskwait` between the uses.

C. Incoherent data-sharing. The data-sharing attribute of a variable must be coherent with its usage. We check three situations, in the following order: a) variables defined within a task and never used in that task, but used after the synchronization of the task, should be shared; b) private variables in a task should be defined before being read, otherwise they must be `firstprivate`; c) `firstprivate` variables in a task should not be defined before being read, otherwise they must be `private`.

D. Incoherent dependencies. Task dependencies are used to impose an order in the execution of the tasks. Dependable objects must be coherent with the usage of those objects in the task. We check three situations: a) objects accessed via pointer should specify the dependency in the accessed storage, instead of the pointer; b) `input` dependences should be read and never written within the task; c) `output` dependences should be written within the task, and should not be read before being written.

We have tested this work with over 70 students and 5 benchmarks on different courses, and the results are shown in Figure 1. Most errors captured by the compiler are related with the default data-sharing attributes because users forget to explicitly change them. The less common errors are those caused by users explicitly defining incorrect data-sharing attributes. Dependences errors are not common because only one of the five benchmarks includes task dependences.

FIGURE 1  
Occurrences of each Correctness Mistake



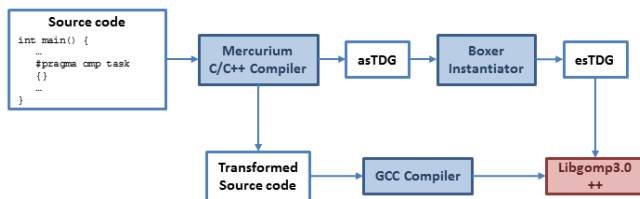


We also have compared our results with those of the Oracle Solaris Studio 12.3 (OCS12.3) compiler for the three first cases, because it does not implement support for task dependences. Since OCS12.3 does not provide hints about performance, Mercurium suggests more accurate solutions in case A. OCS12.3 may report wrong messages for case B, which can be solved by enclosing the analyzed task in a parallel construct. Finally, OCS12.3 does not consider declaring a variable as private when its initial value is not read, so it may result in a loss of performance in case C.

## V. STATIC TASK DEPENDENCY GRAPH

Tasks dependencies impose an order in the execution of the tasks. OpenMP and OmpSs runtimes build a Task Dependency Graph (TDG) ensuring that order. Building the graph at runtime is feasible for HPC systems, where large amounts of memory with a reasonable performance penalty are available. Nonetheless, it may be impossible for embedded systems, where memory can be very limited. The Programming Models and the Real Time teams at BSC have united their forces to build a tool-chain able to derive a TDG statically. The tool-chain is shown in Figure 2.

FIGURE 2  
Offline Tool-chain for an Static TDG Creation



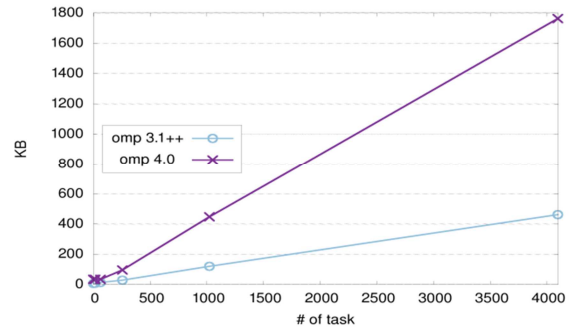
Generating the TDG statically requires a series of steps. First of all, the Mercurium compiler generates a Parallel Control Flow Graph [4] (PCFG) that extends the common CFG with parallelism information. Then induction variables analysis is executed to discover the evolution of each loop. Finally, range analysis [5] extended with support for parallelism is used to calculate the value of the variables at each point of the program. All this information is used to generate an *augmented static TDG* (asTDG) containing: a) one node per each task/taskwait/barrier construct, b) all possible synchronization edges among these nodes, along with predicates defining the conditions for the edges to exist, and c) information about all control flow statements (conditionals and loops) involved in the execution of the previously mentioned constructs.

After that, the Boxer Instantiator expands the asTDG obtaining the complete TDG that will be executed at runtime. This is performed in two steps: 1) expansion of the control flow structures (i.e. decide the number of iterations of each loop statement, and the branches taken in each conditional statement); 2) check of the dependency predicates to decide the task instances that have actual dependences. This process results in an *expanded static TDG* (esTDG).

Finally, the code generated by Mercurium is passed through the GCC back-end compiler to generate the binary that will run in our lightweight libgomp. The tests of our tool-chain are

shown in Figure 3. The memory consumption of the runtime using our statically computed TDG scales (in terms of memory consumption) much better than the plain libgomp supporting tasks dependencies, while increasing the number of tasks.

FIGURE 3  
MEMORY USAGE (IN KB) OF DIFFERENT RTLs



## VI. MODELING TASK PRIORITIES

OpenMP4.1 and OmpSs include support for defining tasks priorities, thus promoting some tasks over others when all of them are ready to be executed. By using our static tool-chain to compute the TDG, we intend to model the applications, and extract patterns in which the priorities of the tasks can be decided at compile time. This is part of our ongoing work and we are looking for applications to exploit this functionality.

## ACKNOWLEDGMENT

We acknowledge the Mercurium group, always helping and giving priceless opinions and contributions to this work. We also thank the Real Time group which brought up fresh ideas and push forward this collaboration.

## LIST OF PUBLICATIONS

- (appeared) R. Ferrer, S. Royuela, D. Caballero, A. Duran, X. Martorell, E. Ayguadé. "Mercurium: Design Decisions for a S2S Compiler". *Cetus Users and Compiler Infrastructure Workshop, PACT*, 2011.
- (appeared) S. Royuela, A. Duran, C. Liao, D.J. Quinlan. "Auto-scoping for OpenMP tasks". *IWOMP*, 2012.
- (appeared) S. Royuela, A. Duran, X. Martorell. "Compiler automatic discovery of ompss task dependencies". *LCPC*, 2013.
- (accepted) D. Caballero, S. Royuela, R. Ferrer, A. Duran, X. Martorell. "Optimizing Overlapped Memory Accesses in User-directed Vectorization". *ICS*, 2015.
- (accepted) S. Royuela, R. Ferrer, D. Caballero, X. Martorell. "Compiler Analysis for OpenMP Tasks Correctness". *CF*, 2015.
- (submitted) R. Vargas, S. Royuela, MA Serrano, E. Quiñones. "A Lightweight OpenMP Run-time for Embedded Systems". *ICCAD*, 2015.

## REFERENCES

- OpenMP ARB. "OpenMP API, v.4.0", July 2013, <http://www.openmp.org/mp-documents/OpenMP4.0.0.pdf>.
- S. Royuela, R. Ferrer, D. Caballero, X. Martorell, "Compiler Analysis for OpenMP Tasks Correctness", *CF15*.
- BSC. "The Mercurium compiler", <http://pm.bsc.es/mcxx>.
- S. Royuela, A. Duran, and X. Martorell, "Compiler Analysis and its application to OmpSs", *Master's Thesis at UPC*, 2012.
- R. Ernani, V. H. Sperle, F. M. Quintão, "A fast and low-overhead technique to secure programs against integer overflows", *CGO13*.

# Automatic Query Driven Data Modelling in Cassandra

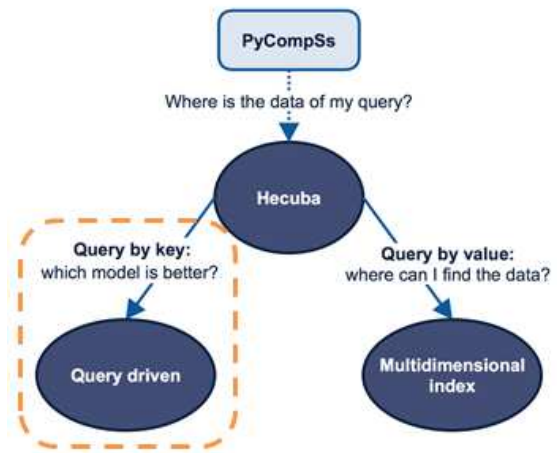
Roger Hernandez, Yolanda Becerra, Jordi Torres, Eduard Ayguadé  
Barcelona Supercomputing Center  
roger.hernandez@bsc.es

**Abstract**-Non-relational databases have recently been the preferred choice when it comes to dealing with Big Data challenges, but their performance is very sensitive to the chosen data organisations. We have seen differences of over 70 times in response time for the same query on different models. This brings users the need to be fully conscious of the queries they intend to serve in order to design their data model. The common practice then, is to replicate data into different models designed to fit different query requirements. In this scenario, the user is in charge of the code implementation required to keep consistency between the different data replicas. Manually replicating data in such high layers of the database results in a lot of squandered storage due to the underlying system replication mechanisms that are formerly designed for availability and reliability ends. We propose and design a mechanism and a prototype to provide users with transparent management, where queries are matched with a well-performing model option. Additionally, we propose to do so by transforming the replication mechanism into a heterogeneous replication one, in order to avoid squandering disk space while keeping the availability and reliability features. The result is a system where, regardless of the query or model the user specifies, response time will always be that of an affine query.

## DESCRIPTION AND PREVIOUS RESULTS

The leap into the Big Data world can be too challenging our out of the field for some information technologies users. This, among other concerns motivated the apparition of Hecuba, a set of tools and interfaces developed in our research group, which aims to facilitate programmers an efficient and easy interaction with non-relational databases. Currently all implementations are made on Apache Cassandra as an example database although it is easy to port this implementation to any non-relational key-value data store.

In the work presented here, we focus on the Query Driven Data Modelling part of Hecuba, a concept for which we have also developed a prototype which automatically decides which data model should be queried by taking advantage of the differences in the topology of the data stored in the Cassandra models and replicas, and overall increase performance by aiming at the affine query-model relationships. Currently, all this management falls on the user responsibility; therefore he would manually require building all the different replicas. This implies further hassle such as assuming the extra redundancy, maintaining the consistency between all the models that share common data when it had to be inserted or changed, or being mindful about performance issues or extra load time when having to modify database data since it would involve to do one write per data replica instead of a single write to the



database. With our suggestion, all this hassle would disappear becoming yet another part of Hecuba that could benefit end-users by removing their need to specialise on the details of the database behaviour.

In our previous publication [2] we could confirm that there existed affinities between queries and models, with great performance differences when combining their possibilities, and the results were so motivating this gave an obvious track opening towards the development of a Query Driven Data Modelling prototype, which seeks to transparently apply the query assignments of the client towards the database aiming to get the best performance it can offer between the available models.

Since we already presented performance results in our previous work, this publication focuses on the design, development and functional analysis of the prototype, finally confirming that the possibilities on performance increases we saw were possible, transparently become a reality with the Query Driven Data Modelling prototype.

## PAST PUBLICATIONS

- [1] C. Cugnasco, R. Hernandez, Y. Becerra, J. Torres and E. Ayguadé, "Aeneas: a tool to enable applications to effectively use non-relational databases" *Procedia Computer Science ICCS 2013*
- [2] R. Hernandez, C. Cugnasco, Y. Becerra, J. Torres and E. Ayguadé, "Experiences of using Cassandra for molecular dynamics simulations" *Euromicro International Conference on Parallel Distributed and Network-based Processing PDP 2015*
- [3] R. Hernandez, C. Cugnasco, Y. Becerra, J. Torres and E. Ayguadé "Automatic Query Driven Data Modelling in Cassandra" *Procedia Computer Science ICCS 2015*

# Modelling the Contact Propagation of Nosocomial Infection in Emergency Departments

Cecilia Jaramillo<sup>1</sup>, Manel Taboada<sup>3</sup>, Francisco Epelde<sup>2</sup>, Dolores Rexachs<sup>1</sup>, and Emilio Luque<sup>1</sup>

<sup>1</sup>Department of Computer Architecture & Operating Systems

Universitat Autònoma de Barcelona, Bellaterra, 08193, Barcelona, Spain

<sup>2</sup>Medicine Department, Hospital Universitari Parc Taulí

Universitat Autònoma de Barcelona, Sabadell, 08208, Barcelona, Spain

<sup>3</sup>Tomas Cerda Computer Science School

Universitat Autònoma de Barcelona, Sant Cugat, 08714, Barcelona, Spain

cjaramillo@caos.uab.es, dolores.rexachs@uab.es, emilio.luque@uab.es, fepelde@tauli.cat, manel.taboada@eug.e

**Abstract**-The nosocomial infection is a special kind of infection that is caused by microorganisms acquired inside a hospital. In the daily care process of an emergency department, the interactions between patients and sanitary staff create the environment for the transmission of such microorganisms. Rates of morbidity and mortality due to nosocomial infections are important indicators of the quality of hospital work. In this research, we use Agent Based Modeling and Simulation techniques to build a model of Methicillin-resistant *Staphylococcus Aureus* propagation based on an Emergency Department Simulator which has been tested and validated previously. The model obtained will allow us to build a contact propagation simulator that enables the construction of virtual environments with the aim of analyzing how the prevention policies affect the rate of propagation of nosocomial infection

## INTRODUCTION

The nosocomial infection is a kind of infection that is caused by microorganisms acquired inside a health care environment [1]. It is the most common type of complication affecting hospitalized patients. Inside a health care environment we can find several microorganism that can be causative of a nosocomial infection, but our work has focused in the propagation of the Methicillin-resistant *Staphylococcus Aureus* (MRSA), one of the most common and dangerous microorganisms in this environment. The presence of these bacteria could mean serious health problems for a patient. It is a common cause of skin, wound and blood stream infections as it may be responsible for a greater hospital length of stay, expensive treatments and an increased mortality [2]. These bacteria live in the skin of some patients and could be transmitted to another patient by physical contact through the interaction between patients, healthcare staff and environment. The most common transmission vias are: healthcare staff's hands, contaminated medical equipment and objects in the hospital room environment. We use Agent Based Modeling and Simulation techniques to build the model and the simulator of Methicillin-resistant *Staphylococcus Aureus* contact propagation in emergency departments. The simulator allows us to build virtual scenarios to understand the phenomenon of the propagation and the potential impact of the implementation of prevention policies on propagation

## Conceptual Model of MRSA Transmission

The model proposed in our research has the advantage of being developed based on a previous emergency department model and previous emergency simulator, both of which have been developed as part of previous research works [3][13] carried out with the collaboration of healthcare staff at the Emergency Department of Hospital Universitari Parc Taulí. This previous model used ABMS techniques to define the full attention process of an emergency department. In our research, we make use of an Agent Based Model and Simulation (ABMS) to create a contact transmission model of MRSA inside an emergency department. For this purpose, we defined all actors involved in the emergency department process and their specific function and behavior. Every person who has a role in the emergency department is defined as an active agent in our model. The environment objects and equipment are representing as passive agents.

**Transmission Forms:** There are two forms of contact transmission, direct transmission and indirect transmission.

1) **Direct transmission:** When MRSA bacteria are transmitted from an active agent (transmission vector) to another active agent (susceptible agent). For instance, when an infected/contaminated patient is touched by a member of healthcare staff without a physical barrier.

2) **Indirect transmission:** When an active agent (transmission vector) touches medical equipment or objects in the hospital environment and MRSA bacteria is transmitted to the object, later, a susceptible agent (patient or healthcare staff) has contact with the same object and acquires the microorganism.

TABLE I  
POSSIBLE WAYS OF TRANSMISSION

Agent 1	Agent 2
Carrier healthcare staff	Susceptible patient
Carrier healthcare staff	Passive agent
Infected/colonized patient	Passive agent
Infected/colonized patient	Healthcare staff
Passive agent	Passive agent
Passive agent	Healthcare staff
Passive agent	Susceptible patient

*Agents and Behaviors:* In our research, we make use of an Agent Based Model and Simulation (ABMS) to create a contact propagation model of MRSA inside an emergency department. For this purpose, we defined all actors involved in the emergency department process and their specific function and behavior.

- 1) *Active Agents:* Every person who has a role in the emergency department is defined as an active agent in our model: patients, doctors, nurses, admission staff, laboratory technicians, auxiliary personnel and cleaning staff (Fig. 1). For the purpose of our propagation model we divided active agents in two sets: Patient and Healthcare Staff. It is necessary to make this classification because each group can take a different infectious status. We classified patients, according their infectious status in non-colonized, colonized or infected. All non-colonized patients are susceptible to acquiring MRSA, but some of them could be more susceptible than others, such as patients with open wounds, external devices among other reasons. Colonized patients have MRSA bacteria but they do not show symptoms, and infected patients have MRSA bacteria and show symptoms. Healthcare Staff has three possible infect status: non-carrier, carrier and colonized. The carrier status is a temporary colonized.

In an attempt to control MRSA transmission rate, some concrete actions are performed in ED such as hand washing, hand disinfection and the use of isolation material, all of which are called Prevention Policies. Only the Healthcare Staff agents can be perform these prevention policies.

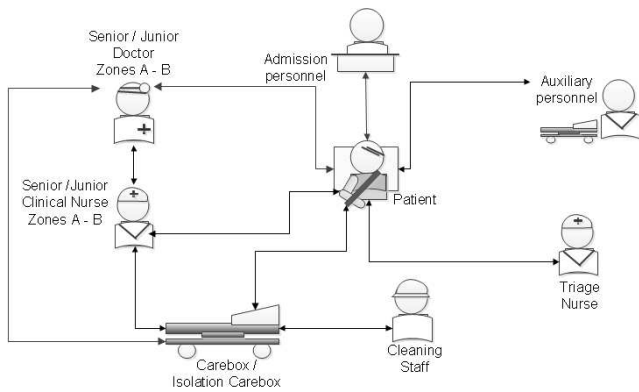


Fig. 1. General Diagram of Physical Contact.

- 2) *Passive Agents:* It is very important to take into account the environmental objects that healthcare staff use in the process of attention patient because the MRSA bacteria can be transmitted when an active agent touch a contaminated surface. For our purpose, we created the passive agent “Carebox” in order to

represent all the environmental objects and equipment that can be in contact with the active agents. We use the infected variable to reflect whether a carebox is contaminated with MRSA or not. The possible value for infected variable are: non-contaminated or contaminated. When an MRSA patient is assigned to a carebox, the infected variable takes the value contaminated, and it will remain at such a value until a disinfection process has been carried out by cleaning staff.

#### TRANSMISSION MODEL

Frequent interaction between patients and healthcare staff is the principal way to MRSA transmission. Whenever an interaction between agents is given, we consider an MRSA transmission is likely. However, there is a wide variety of factors that will determine if the transmission has a direct impact on the infectious status of the susceptible agent and change it from non-carrier or non-colonized, to carrier, colonized, infected or whether in contrast, the bacteria are removed and the infectious status of the susceptible agent will not change. Our model considers in each interaction the susceptibility of the agent who is at risk of acquiring MRSA. If the agent at risk is a patient, the analysis is based on the health state of the patient (predisposition, infected status), but if the agent at risk is a member of the healthcare staff, the analysis takes in account other values such as accomplishment and effectiveness of prevention policies (hands wash, disinfection hands, etc.). The model takes into account the parts of the overall attention process in which the agents interact with each other, because the contact propagation can take place in these moments. Whenever the healthcare agent finishes an attention task, they may or may not apply the prevention policies: hand washing, hand disinfection or using isolated material. The action that is executed can be effective or not.

#### ACKNOWLEDGMENT

This research has been supported by: MINECO (MICINN) Spain under contract TIN2011-24384. Ecuador government (SENESCYT) under contract 2013-AR7L335.

#### REFERENCES

- [1] J.S. Garner, W.R. Jarvis, T.G. Emori, T.C. Horan, and J.M. Hughes, “Definitions for nosocomial infections”, American journal of infection control, vol. 16(3), pp 128–140, 1988.
- [2] N. Graves et al., “Effect of HealthcareAcquired Infection on Length of Hospital Stay and Cost,” Infection Control and Hospital Epidemiology, vol. 28, no. 3, pp. 280–292, 2007.
- [3] Eduardo Cabrera, Manel Taboada, Ma Luisa Iglesias, Francisco Epelde, and Emilio Luque. Simulation optimization for healthcare emergency departments. Procedia Computer Science, 9:1464–1473, 2012.
- [4] Zhengchun Liu, Eduardo Cabrera, Dolores Rexachs, and Emilio Luque. A generalized agent-based model to simulate emergency departments. The Sixth International Conference on Advances in System Simulation (SIMUL) IARIA 2014, 2014.

# MACC: Mercurium ACcelerator Model

Guray Ozen, Eduard Ayguade, Jesus Labarta  
Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain  
Universitat Politècnica de Catalunya (UPC-BarcelonaTECH)  
{name.surname}@bsc.es

**Abstract-** GPU Offloading is emergent programming model. OpenMP includes in its latest 4.0 specification the accelerator model. In this paper we present a newly implementation of this specification while generating "native" GPU code in the OmpSs programming model developed at the Barcelona Supercomputing Center. Focused on targeting NVIDIA GPUs, our work is based on an OpenMP 4.0 implementation in the Mercurium source-to-source compiler infrastructure referred MACC[1] (Mercurium ACcelerator compiler). Finally, the paper is also discussing challenges of code generation for GPUs.

## I. INTRODUCTION

The use of accelerators has been gaining popularity in the last years due to their higher peak performance and performance per Watt ratio when compared to homogeneous architectures based on multicores. However, the heterogeneity they introduce (in terms of computing devices and address spaces) makes programming a difficult task even for expert programmers.

The OmpSs[2] proposal has been evolving during the last decade to lower the programmability wall raised by multi/many-cores, demonstrating a task-based data flow approach in which offloading tasks to different number and kinds of devices, as well as managing the coherence of data in multiple address spaces, is delegated to the runtime system. This paper presents the latest implementation of OmpSs which includes partial support for the accelerator model in OpenMP 4.0 specification. We adopted those functionalities that are necessary to specify computational kernels in a more productive way.

## II. MACC: MERCURIUM ACCELERATOR COMPILER

A new compilation phase (MACC is abbreviation for "Mercurium Accelerator Compiler".) has been included in the Mercurium[3] compiler supporting the OpenMP 4.0 accelerator model with the OmpSs runtime. MACC takes care of kernel configuration, loop scheduling and appropriate use of the memory hierarchy for those tasks whose *device* is set to *acc* in the *target* clause. Some of the OpenMP 4.0 directives for accelerator *target data*, *target update* directives and *map* clause) are simply ignored because we delegate their functionality to the runtime system. Others have been extended to better map with the OmpSs model or to provide additional functionalities.

### A. Kernel Creation

MACC offloads all code inside of *target* directive into GPU. Basically it generates kernel code and when code reaches that point, it invokes device kernel. Since MACC is based OmpSs, it behaves *target* directive as a *task*. It is novel

approach of MACC and as well as according to OpenMP 4.1 drafts this specification will bring that approach soon. By this approach, MACC could put together advantages of *task* and *target* at the same time. The main advantages of that usage different from standard OpenMP 4.0 are device-to-device transfer became available and MACC could schedule multi-gpu task.

### B. Kernel Thread Hierarchy Generation

When generating kernel code MACC needs to decide: 1) the dimensionality of the resources hierarchy (one-, two- or three-dimension teams and threads) and 2) the size in each dimension (number of teams and threads). In order to support the organization of the threads in multiple dimensions MACC allows the nesting of *parallel for* directives inside a *target* region.

### C. Loop Directives and Data Clauses

MACC performs a cyclic mapping of loop iterations and tries to eliminate redundant "one-iteration" loops and simplifies increment expressions for induction variables in order to improve kernel execution time. MACC also makes use of shared memory for threads in a team based on the specification of *private* and *firstprivate* data structures in the *distribute* directive, so that each team allocates a private copy in its own shared memory.

However, for very large private arrays this is not possible to apply. For these cases we have implemented tree new clauses (*dist\_private*, *dist\_firstprivate* and *dist\_lastprivate*); with these clauses and the chunk size provided in the *dist\_schedule(static,chunk\_size)* clause in the *distribute* directive the compiler just allocates a portion of the arrays to each team and performs the copies.

### D. Reduction

Reduction is a very important of GPU parallelism as it was in the CPU side. When generating reduction code, MACC uses different reduction algorithms depending upon primitive type in order to obtain best speedup. It uses `__shfl` intrinsic (by this instruction it can obtain significant speedup) and doing reduction by using shared memory for thread level reductions. As for Thread Block Level Reduction, it uses atomic operations of GPUs and calling second kernel approaches.

## III. PERFORMANCE EVALUATION

### A. MACC Reduction on Jacobi

Jacobi is a simple iterative program to get an approximate solution of a linear system  $A*x=b$ . The structure of the code is shown in Figure-1, with two different annotations that correspond to two versions.

OpenACC(non-optimised)	MACC (auto Optimized)
<pre>while (...) { #acc kernels\   copyin(u) copyout(uold)   for (...)     // &lt;..computation..&gt;  #acc kernels \   copyout(u) copyin(uold)\   reduction(+:err)   for (...)     // &lt;..computation..&gt; }</pre>	<pre>while (...) { #omp target device(acc) #omp task in(u) out(uold) #omp parallel for   for (...)     // &lt;..computation..&gt;  #omp target device(acc) #omp task out(u) in(uold) #omp parallel for \   reduction(+:err)   for (...)     // &lt;..computation..&gt; }</pre>

Figure-1

Figure-2 shows benchmark between 3 versions. There are two OpenACC(HMPP) versions baseline and optimized. Optimized means when programmers use explicitly data directives, OpenACC keeps data into device. Since MACC behaves code will be offloaded as a task, programmers doesn't need to indicate any extra directives.

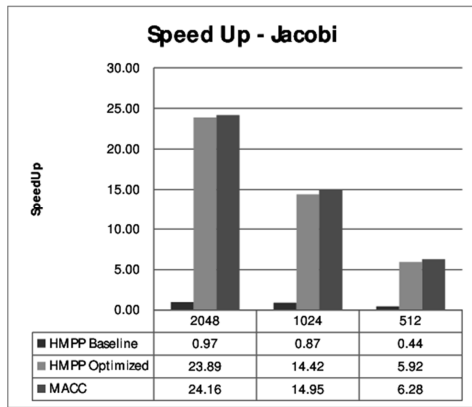


Figure-2

**B. MACC Shared Memory performance on DG-kernel**  
 DG is a kernel version of a climate benchmark developed by National Center for Atmospheric Research dg-kernel. The code consist single target. Data directionalities and characteristics are indicated at figure-3. By using new MACC's clauses, partly using shared memory is enabled. Additionally array B could be taken back from shared memory. This usage corresponds opt-1 at figure-4.

```
#omp target device(acc)
#omp task in(A[0:SMALL],C[0:HUGE]) inout(B[0:HUGE])
out(D[0:BIG])
#omp teams distribute parallel for first_private(A)
  dist_first_private([CHUNK]C)\
  dist_first_last_private([CHUNK]B)
for(...)
<<..Computation..>>
```

Figure-3

Besides, when MACC build kernel hierarchy equals to iteration size, it removes redundant for iteration in order to avoid thread-divergence. At figure-4 this optimization denotes opt-2.

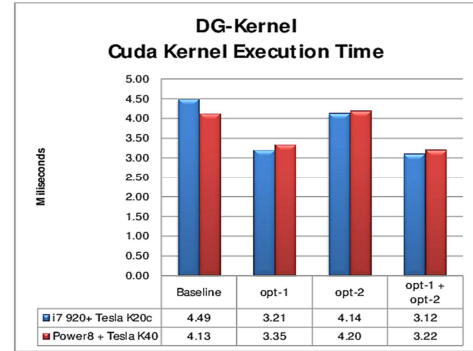


Figure-4

**C. MACC multi-GPU scheduling on NAS-CG**  
 The last code we have selected for the experimental evaluation in this paper is NAS CG. The main computational part of the application contains several loops that can be made tasks and offloaded to a device or executed on the host. To execute this loop we want to use the two GPUs available in the node. We could enable multi gpu by dividing most expensive iteration into chunks, since nanox can schedule them automatically. Figure-5 shows comparison graph.

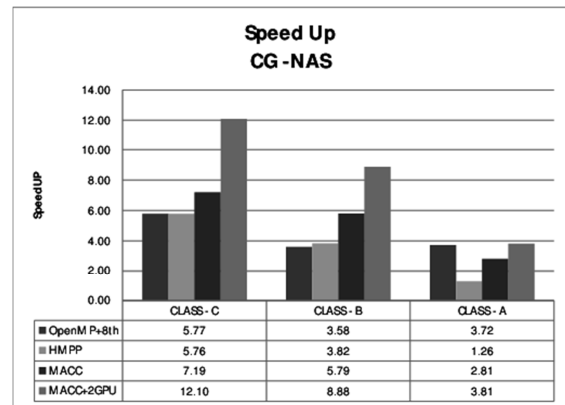


Figure-5

## REFERENCES

- [1] Guray Ozen, Eduard Ayguadé, Jesús Labarta: On the Roles of the Programmer, the Compiler and the Runtime System When Programming Accelerators in OpenMP. IWOMP 2014:
- [2] The OmpSs programming model. <http://pm.bsc.es>
- [3] Mercurium Compiler. <http://pm.bsc.es/mcxx>

# A 3D-1D cardiac-vascular computational feedbacked model

Alfonso Santiago, Mariano Vázquez, Guillaume Houzeaux, Jazmin Aguado-Sierra, Mariña López, Matías Rivero (1)  
 Pablo J. Blanco, Sansuke Watanabe (2)  
 (1) Barcelona Supercomputing Center, Barcelona - Spain  
 (2) Laboratório Nacional de Computação Científica, Petrópolis -Brasil.  
 alfonso.santiago@bsc.es

**Abstract-** A first version of a cardiovascular coupled model is presented. The modelling tools are Alya, the BSC tool for biomechanical simulations; and ADAN55, the LNCC model of the arterial blood flow. The former solves the heartbeat by Finite Element Method (FEM). The latter is a 1D model of the arterial blood flow in a 55-branched geometry. Previous generations of the involved models had fixed boundary conditions. With this novel tool a feedbacked cardiac-vascular model is obtained, offering more physiological boundary conditions and providing further insight in the cardiac-systemic interactions.

## I. INTRODUCTION

The used models were BSCs Alya, and LNCCs ADAN55. Alya is a 3D, multiscale, multiphysics, HPC code, that allows modelling and solving the heartbeat electromechanical problem using the Finite Element Method (FEM). The ADAN55 is a reduced 55-artery version of the Anatomically Detailed Arterial Network (ADAN) model, which takes into account more than 1500 arteries from an average human body, featuring a physiologically consistent systemic impedance at the aortic root, among other characteristics. Both codes are coupled by a black-box decomposition approach previously proposed in [7]. At each time step the Fluid Structure Interaction (FSI) problem is solved in the ventricle, computing the output flow through the aorta which gives a pressure response back. In previous generations of the models, each one had predefined boundary conditions: for Alya a pressure value was imposed on the endocardium, and an experimentally measured flow contour was imposed in the aortic root of the ADAN55 model. With this novel coupling a more physiological response to the blood ejection is obtained, as the systemic impedance is a complex result of the entire 1D arterial model which changes with the flow generated by the heart model. In the same way, using the heart model output flow connected to the ADAN55 input, offers the possibility to obtain a physiological and feedbacked stimuli that varies with changes in the arterial system.

## II. METHODOLOGY

### A. Governing equations

The cardiovascular system can be decomposed as a coupled electromechanical system: the propagation of the action potential induces the mechanical deformation of the solid (myocardium), which reduces the inner cavity inducing the pumping action against the fluid (blood) through the arteries. In this way, three problems have to be solved: the propagation of the electric potential, the deformation of the solid due to the electrophysiologic stimuli and outer forces, and the fluid dynamics inside the ventricle and the arteries.

The governing equations for the electrophysiology potential model are described in [1]:

$$(1) \quad C_m \frac{\partial \phi}{\partial t} = \frac{\partial}{\partial x_i} \left( \frac{D_{ij}}{S_v} \frac{\partial \phi}{\partial x_j} \right) + I_{ion}$$

where  $D_{ij}$  is the diffusion tensor,  $C_m$  and  $S_v$  are the membrane capacitance, and the surface-to-volume ratio respectively. The expression of  $I_{ion}$  is what determines the complexity of the model. In this case we use the Fitzhugh-Nagumo model [2] due to its simpleness. The mechanical deformation governing equations are described in [3]. In the expression of the momentum balance of the solids:

$$(2) \quad \rho_o \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial P_{iJ}}{\partial X_J} + \rho B_i$$

where  $\rho$  is the initial density of the body,  $B_i$  are the body forces and  $P_{iJ}$  the first Piola-Kirchoff stress tensor. A modified version of the Holzapfel and Odgen [4] model is used for the energy function which defines the constitutive model. The coupling between the electrophysiology and the solid model is done through the Hunter and Nash model [5].

Inside the cavity, the fluid is modeled by the incompressible Navier-Stokes, as explained in [6].

The governing equations in the arterial network, and the way they are solved, are described in [7]. The condensed 1D Navier-Stokes equations in compliant vessels comprise momentum and mass conservation as:

$$(3) \quad \frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left( \alpha_m \frac{Q^2}{A} \right) + \frac{A}{\rho} \frac{\partial P}{\partial x} + \frac{2\pi R}{\rho} \tau_o = 0$$

$$(4) \quad \frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0$$

$$(5) \quad P = P_0 + \frac{Eh}{R_0} \left( \sqrt{\frac{A}{A_0}} - 1 \right) + \frac{Kh}{R_0} \frac{1}{2\sqrt{A_0 A}} \frac{\partial A}{\partial t}$$

where  $A$  is the luminal area,  $R$  is the radius,  $Q$  is the flow rate,  $P$  is the mean pressure,  $\rho$  is the density,  $\alpha_m$  is the momentum correction factor,  $\tau_o$  accounts for the viscous effects,  $h$  is the wall thickness and  $E$  and  $K$  are the material parameters that characterize the elastic and viscoelastic material responses. (3) and (4) are the momentum and continuity balance, respectively. (5) has the FSI terms for the 1D formulation.

### B. Numerical test and results

A ventricle is simulated using an ellipsoidal cavity made of active cardiac tissue, with fibers arranged vertically, with a short non-active tubular part playing the role of the aorta. The

activation potential is started at the bottom. The arterial network is also simplified, including 55 segments, called ADAN55. Figure 3 shows the flowrate and a scaled (1%) pressure in the coupling point. Figure 1 and 2 shows the ventricle-like geometry and part of the ADAN55 mesh.

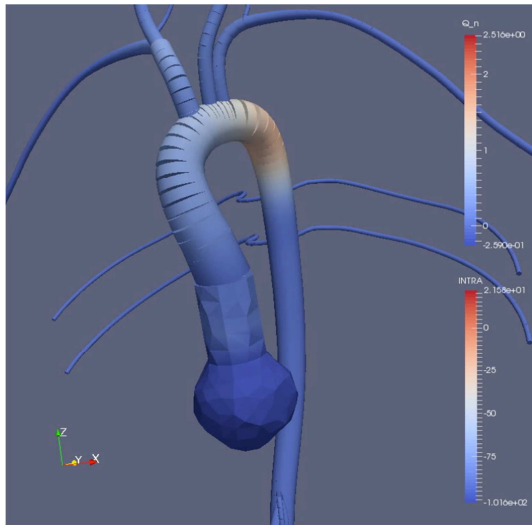


Fig.1 Voltage in the solid and flow in the arterial network.

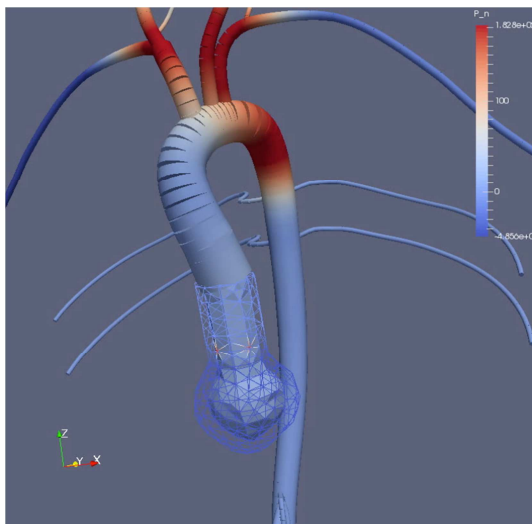


Fig 2.Pressure in the arterial network.

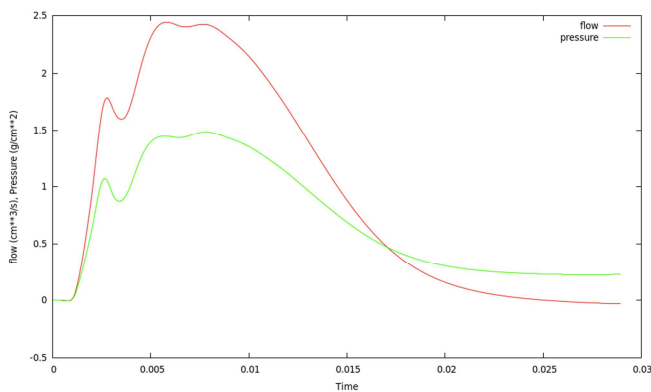


Figure3: Pressure (1% scaling) and flow in the coupling point.

### III. CONCLUSIONS

In this work we show preliminary results of a coupled simulation tool that widens the possibilities in cardiovascular modeling. A complete simulator of the cardiovascular system will allow bioengineers and medical researchers to study the response of the full model to changes in any of the components, extending the predictive and descriptive capabilities of the standalone versions of the models, and providing further insight in the cardiac-systemic interactions.

### ACKNOWLEDGMENT

The Ph.D. studies of A. S. are partially funded by the EU Brazil Cloud Connect Project, funded by the European Commission under the Cooperation Programme, Framework Programme Seven (FP7).

### REFERENCES

- [1] Vázquez, M., Arís, R., Houzeaux, G., Aubry, R., & Villar, P. (2011). A massively parallel computational electrophysiology model of the heart. *International Journal for Numerical Methods in Biomedical Engineering*. doi:10.1002/cnm
- [2] FitzHugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, (1948).
- [3] Lafortune, P., Arís, R., Vázquez, M., & Houzeaux, G. (2012). Coupled electromechanical model of the heart: Parallel finite element formulation. *International Journal for Numerical Methods in Biomedical Engineering*, 28, 72–86. doi:10.1002/cnm
- [4] G. A. Holzapfel and R. W. Ogden, "Constitutive modelling of passive myocardium: a structurally based framework for material characterization," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1902, pp. 3445–3475, 2009.
- [5] Nash, M., & Hunter, P. (2000). Computational mechanics of the heart. *Journal of Elasticity and the Physical Science of*, 113–141.
- [6] Houzeaux, G., Vázquez, M., Aubry, R., & Cela, J. M. (2009). A massively parallel fractional step solver for incompressible flows. *Journal of Computational Physics*, 228(17), 6316–6332.
- [7] Blanco, P. J., Leiva, J. S., & Buscaglia, G. C. (2013). A black-box decomposition approach for coupling heterogeneous components in hemodynamics simulations. *International Journal for Numerical Methods in Biomedical Engineering*, 29(November 2012), 408–427.



# Evaluation of Modelling Systems in High Resolution to Assess the Air Pollutant Impacts on Human Health

Sergio Natan González-Rocha, Jose M. Baldasano  
 Barcelona Supercomputing Centers (BSC – CNS) – Earth Sciences  
 sergio.gonzalez@bsc.es

**Abstract-** Nowadays the modelling of systems in high resolution is being used for air quality and other forecasting applications, where a spatial area is related with different interrelated variables that could be displayed on a map. This area is usually represented by global domains (hundred to thousands of square km); when smaller regions need to be represented, a high resolution modelling system can be used, these systems goes from one square km to dozen of square km, health is one of these issues where this kind of resolution can be used. In Europe, Asia, North America, South America and other countries, health problems related with the air pollution and climate change is a concern for individuals and world organizations like the WHO; today studies show the relation between morbidity and mortality rates, air pollution and effects on human health; these modelling systems in high resolution help us to simulate scenarios and propose solutions to this problematic. So the objective of this work is to evaluate the system performance WRF – CMAQ and CALIOPE on high resolution (4 km x 4 km) to determine air pollutant impacts of PM<sub>10</sub>, PM<sub>2.5</sub>, Ozone, NO<sub>2</sub> and SO<sub>2</sub> on population, using BenMAP for assess impact on health. The methodology suggested is the time series analysis of two years of hospital admissions, morbidity and mortality rates and the air quality forecasting of the cities selected, previously modelled in WRF, CMAQ and CALIOPE; after that, the Response Functions (DRF/ERF) to determine the impacts on health and the BenMAP software will be used. It is expecting find the scenarios that could decrease the mortality and morbidity rates in diseases like lung cancer; chronic respiratory obstructive disease, asthma, and the acute respiratory diseases in adults and children under ten years old.

Key words: Air quality, Modelling, Health impact

## I. INTRODUCTION

Today Air quality (AQ) is an environmental issue that has become a concern for the effects on health, [1][2][3]. AQ studies show results on increased morbidity and mortality in Europe, USA, Asia, Latin America and Caribbean; respiratory diseases, cardiovascular diseases, nervous system and lung cancer are some examples, [4]; This is no longer just an issue of big cities such as Mexico City, New York, Hong Kong or Sweden, this is also a problem in small and medium cities, [5][6][7][8][9][10][11].

## II. MODELLING SYSTEMS

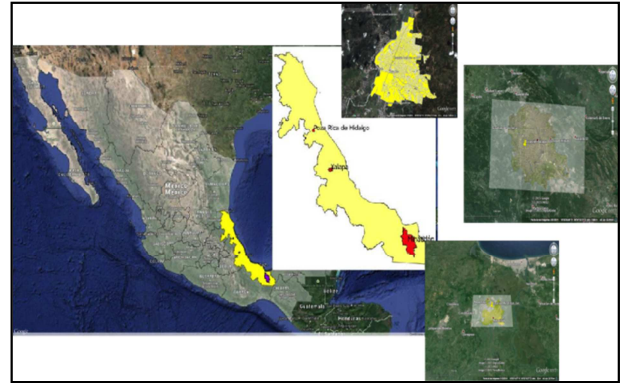


Fig. 1 Representation of two domains, a National domain and a high resolution domain at Veracruz cities, Mexico.

Reference [16] mentions that the air quality modelling (AQM) plays an important role in the policies and strategies for the air pollution control in many countries. There are different systems that allows estimated these values of AQ, some of these are: AERMOD, CALPUFF and CMAQ for example used by the US-EPA, CALINE 4 in UK, WRF- Chem in Mexico and CALIOPE Air Quality Forecast model (AQFM) in Spain. Tools coupled GIS are also options that provide support for this type of projects, [17][18][19].

## III. HEALTH IMPACT ASSESSMENT (HIA)

The health impact assessment is realized by a methodology that permit evaluate the human health impacts from different causes, nowadays the air pollution is one of these causes as can be seen in the Fig. 2. There are different ways to assess, the short-term studies and the long-term studies. In this project the short-term has been selected, by adapting the methodology proposed by the "Instituto Nacional de Ecología y Cambio Climático" INECC from Mexico, to assess the mortality and morbidity population rates related to PM<sub>10</sub>, PM<sub>2.5</sub>, Ozone and temporal series analysis, [20].

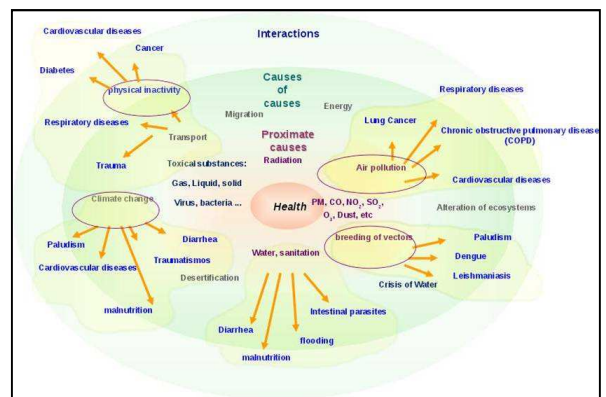


Fig. 2 Human health impacts from different variables and factors.

#### IV. EXPECTED RESULTS

1. Estimate the exposure to air pollution in population and the effects on human health in the cities selected.
2. Find the rates of mortality and morbidity that could be decreased with this assessment.
3. Find some causes or etiologies that produce the number of deaths or morbidity.
4. Find a model or models with the best fit.
5. Determine the spatial scope of air pollutants.
6. Determine the benefits obtained in the simulation scenarios.

#### ACKNOWLEDGMENT

The description in this poster, is a first advance of a research project related to human health and air pollution; Thanks to the “Consejo Nacional de Ciencia y Tecnología” (CONACyT), the Universidad Veracruzana in Mexico, and the Barcelona Supercomputing Center – Centro Nacional de supercomputación (BSC – CNS) in Barcelona, Spain, for the valued support to this postdoctoral fellowship; thanks to the SEDEMA and SSA in Veracruz, Mexico for the information and data of air quality and health.

#### REFERENCES

- [1] Portal de la Organización Mundial de la Salud, (2005), Guías de consultado el 15 de mayo de 2014, disponible en [http://whqlibdoc.who.int/hq/2006/WHO\\_SDE\\_PHE\\_OEH\\_06.02\\_spa.pdf?ua=1](http://whqlibdoc.who.int/hq/2006/WHO_SDE_PHE_OEH_06.02_spa.pdf?ua=1)
- [2] Portal de la OMS (2013); Nota informativa 266, consultada el fecha disponible en <http://www.who.int/mediacentre/factsheets/fs266/es/>
- [3] Portal de la Organización Mundial de la Salud Consultado el 15 de mayo de 2014, disponible en [http://www.who.int/topics/environmental\\_health/es/](http://www.who.int/topics/environmental_health/es/)
- [4] Portal de la OMS (2011); Nota informativa 313, consultada el 15 de mayo de 2014, disponible en <http://www.who.int/mediacentre/factsheets/fs313/es/>
- [5] Baldasano J.M., E. Valera and P. Jiménez, Air Quality Data from Large Cities, 2013, *The Science of the Total Environment*, 307, 1-3: 141-165
- [6] Zhen Cheng, Jingkun Jiang, Oscar Fajardo, Shuxiao Wang, Jiming Hao, Characteristics and health impacts of particulate matter pollution in China (2001–2011), *Atmospheric Environment*, Volume 65, February 2013, Pages 186-194, ISSN 1352-2310, <http://dx.doi.org/10.1016/j.atmosenv.2012.10.022>, <http://www.sciencedirect.com/science/article/pii/S1352231012009867>
- [7] T.W. Smith, C.J. Axon, R.C. Darton, The impact on human health of car-related air pollution in the UK, 1995–2005, *Atmospheric Environment*, Volume 77, October 2013, Pages 260-266, ISSN 1352-2310, <http://dx.doi.org/10.1016/j.atmosenv.2013.05.016>, (<http://www.sciencedirect.com/science/article/pii/S1352231013003592>)
- [8] E. Fridell, et al., A modelling study of the impact on air quality and health due to the emissions from E85 and petrol fuelled cars in Sweden, *Atmospheric Environment*, Volume 82, January 2014, Pages 1-8, ISSN 1352-2310, <http://dx.doi.org/10.1016/j.atmosenv.2013.10.002>, (<http://www.sciencedirect.com/science/article/pii/S1352231013007553>)
- [9] I. J. Simpson, et al., Air quality in the Industrial Heartland of Alberta, Canada and potential impacts on human health, *Atmospheric Environment*, Volume 81, December 2013, Pages 702-709, ISSN 1352-2310, <http://dx.doi.org/10.1016/j.atmosenv.2013.09.017>, (<http://www.sciencedirect.com/science/article/pii/S135223101300705X>)
- [10] R.V. Zelm, et al., European characterization factors for human health damage of PM10 and ozone in life cycle impact assessment, *Atmospheric Environment*, Volume 42, Issue 3, January 2008, Pages 441-453, ISSN 1352-2310, <http://dx.doi.org/10.1016/j.atmosenv.2007.09.072>, (<http://www.sciencedirect.com/science/article/pii/S1352231007008667>)
- [11] C. Chio, et al., Health risk assessment for residents exposed to atmospheric diesel exhaust particles in southern region of Taiwan, *Atmospheric Environment*, Volume 85, March 2014, Pages 64-72, ISSN 1352-2310, <http://dx.doi.org/10.1016/j.atmosenv.2013.11.072>, (<http://www.sciencedirect.com/science/article/pii/S1352231013009229>)
- [12] I. Aguilera, et al., Evaluation of the CALIOPE air quality forecasting system for epidemiological research: The example of NO<sub>2</sub> in the province of Girona (Spain), *Atmospheric Environment* Volume 72, February 2013, pages 134 – 141, ISSN 1352-2310, <http://dx.doi.org/10.1016/j.atmosenv.2013.02.035>
- [13] X. Basagaña et al., Soil dust aerosols and wind as predictors of seasonal Meningitis incidence in Niger, *Environmental Health Perspectives* 2014; Vol. 122, Number 7, pages 679 – 686; <http://dx.doi.org/10.1289/ehp.1306640>
- [14] X. Basagaña et al., Short-term effects of particulate matter constituents on daily hospitalizations and mortality in five south-European cities: Results from the MED-PARTICLES project, *Environmental International*, 75, 2015, pages 151 – 158, <http://dx.doi.org/10.1016/j.envint.2014.11.011>
- [15] Y. Akita et al., Large Scale Air Pollution Estimation Method Combining Land Use Regression and Chemical Transport Modeling in a Geostatistical Framework, *Environmental Science & Technology*, 2014, Vol. 48, pages 4452-4459, [http://www.bsc.es/sites/default/files/public/earth\\_science/2014-envscitech-akita-largescaleairpollutionestimationmethodcombininglurandctm.pdf](http://www.bsc.es/sites/default/files/public/earth_science/2014-envscitech-akita-largescaleairpollutionestimationmethodcombininglurandctm.pdf), [dx.doi.org/10.1021/es405390e](http://dx.doi.org/10.1021/es405390e)
- [16] S. Gulia, S.M. Shiva Nagendra, M. Khare, I. Khanna; Urban air quality management – A review, *Atmospheric Pollution Research*, 2015, Vol. 6, pages 286 – 304, doi: 10.5094/APR.2015.033
- [17] Elbir, T.; A GIS based decision support system for estimation, visualization and analysis of air pollution for large Turkish cities, *Atmospheric Environment*, 2004, 38, pages 4509 – 4517.
- [18] Elbir, T., Mangir, N., Sirmsir, S., Eren, T., Ozdemir, S., Development of a GIS-based decision support system for urban air quality management in the city of Istanbul, *Atmospheric Environment*, 2010, 44. pages 441 – 454.
- [19] Gulliver, J., Briggs, D., 2011. STEMS-Air: A simple GIS-bases air pollution dispersion model for city-wide exposure assessment, *Science of the Total Environment*, 2011, 409, pages 2419-2429
- [20] Instituto Nacional de Ecología y Cambio Climático, Guía para evaluar los impactos en la salud por la instrumentación de medidas de control de la contaminación atmosférica, 2012, 1a edición, ISBN 978-607-824-635-9, México.

# Binding Free Energy and Ligand Orientation Calculations using A Monte Carlo Method with Markov State Analysis

Daniel Lecina-Casas<sup>1,\*</sup>, Ryoji Takahashi<sup>1</sup>, and Victor Guallar<sup>1,2</sup>

1- Barcelona Supercomputing Center, c/Jordi Girona 29, 08034 Barcelona, Spain

2- Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain  
daniel.lecina@bsc.es

**Abstract** - Computing binding free energies has great implications in drug design. Using PELE technique, it has been shown that one can get quick and accurate estimations by means of a Markov state model<sup>3</sup>. We improved our methodology to compute faster binding relative free energy differences, mainly by analysis reducing the sampled region. This possibility opens a way in all-atom drug lead optimization by efficiently scoring a list of potential candidates in terms of binding affinities (approximately in 24hours), while still modeling accurately the protein-drug induced fit. Furthermore, we added information of the ligand orientation allowing us to obtain a better insight of the entrance mechanism. First, we show benchmark results - a series of benzamidine-like inhibitors in trypsin. Then, we apply our method to a more realistic scenario: the binding to a glucocorticoid receptor, and we show the performance for a new benchmark with a larger range of binding free energies (~14 kcal/mol). Simulations are obtained with our new in-house code PELE++, an improvement over the technique presented in references [1,2], (paper in preparation).

## I. METHODS

We first run simulations using PELE++. This allows us to obtain more data in a given computational time compared to regular MD.

Pele++ combines a stochastic approach with protein structure prediction algorithms. After every iteration, a step is accepted based on the Metropolis criterion. Pele++ can be used for free in our webserver<sup>2</sup> in <https://pele.bsc.es>.

We obtain the centroid position and quaternion of rotation from a reference structure for each snapshot.

For the analysis, we build a Markov State Model. To do so we can use for example EMMA (<https://simtk.org/home/emma>), or MSMBUILDER (<https://simtk.org/home/msmbuilder>).

Quantitative binding free energy difference is obtained by means of:

$$\Delta G = -k_B T \ln(V_b/V_o) - \Delta W$$

To get a better mesoscopic insight, cluster-cluster analysis (e.g. PCCA+) is used.

We use the quaternion information to get quantitative results of the entrance and orientation in the binding site.

Fig. 1. PELE scheme. First, perturbation is performed by means of ligand perturbation followed by protein perturbation (ANM). Afterwards, the structure is refined by means of a side chain prediction and minimization.

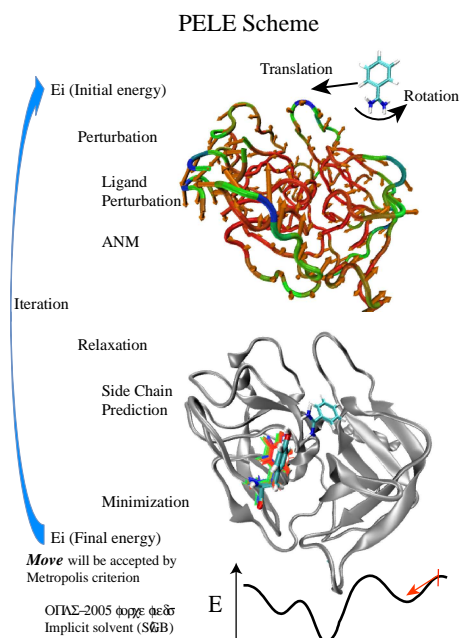


Fig. 1. PELE scheme. First, perturbation is performed by means of ligand perturbation followed by protein perturbation (ANM). Afterwards, the structure is refined by means of a side chain prediction and minimization.

## II. RESULTS

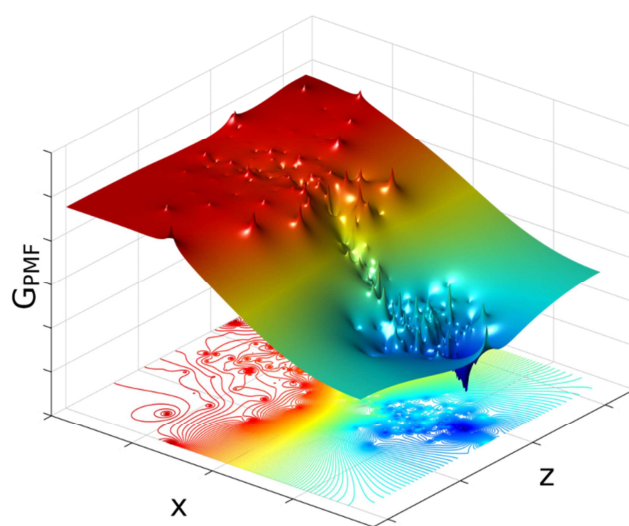


Fig. 2. Potential of mean field for a hormone receptor.

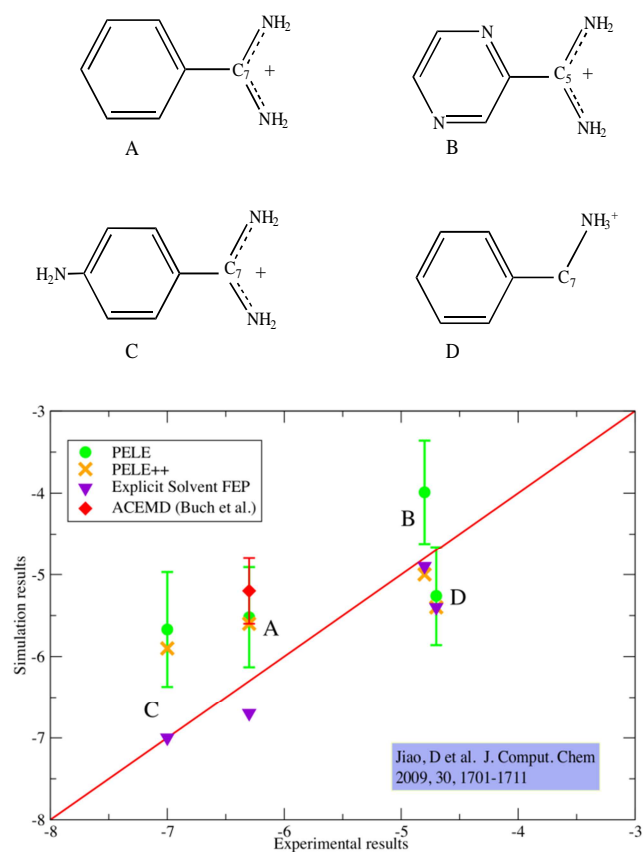


Fig. 2. Different trypsin inhibitors with their predicted and experimental binding free energies.

### III. CONCLUSIONS AND FUTURE WORK

We are able to compute satisfactorily differences in binding free energy. Besides, we gained a better insight into the binding mechanism by means of the orientation analysis.

But some questions arose: What approach should be used if the ligand is completely flexible? How do we analyze the rotatable bonds? How much sampling do we need?

### REFERENCES

- [1] Borrelli, K. W., Vitalis, A., Alcantara, R. & Guallar, V. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput.* 1, 1304-1311 (2005).
- [2] Madadkar-Sobhani, A. & Guallar, V. PELE web server: atomistic study of biom. *Nucl. Acids Res.* 41, W322-W328 (2013).
- [3] Takahashi, R., Gil, V. A. & Guallar, V. Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations. *Journal of Chemical Theory and Computation* 10, 282-288, doi:10.1021/ct400678g (2014).

# A parallel coupled algorithm for the solution of deformable two-body contact problem

M. Rivero<sup>1</sup>, J.C. Cajas<sup>1</sup>, G. Houzeaux<sup>1</sup>, M. Vazquez<sup>1,2</sup>

<sup>1</sup>Barcelona Supercomputing Center, Spain

<sup>2</sup>IIIA/CSIC, Spain

matias.rivero@bsc.es

**Abstract-** *This work presents a parallel iterative method for numerical solving frictionless contact problem for two elastic bodies. Each iterative step consists of a Dirichlet problem for the one body and a Neumann problem for the other in order to enforce the contact boundary conditions.*

## I. INTRODUCTION

Contact problems deal with the deformation of separate bodies which interact when they come in touch. They take an important place in computational mechanics. Usually, they are formulated as constrained minimization problems which may be solved using optimization techniques such as penalty method, Lagrange multipliers, augmented Lagrangian method, etc [1]. When using an implicit scheme, this approach requires the construction of the so-called contact elements, which incorporate the contact constraints in the global weak form. In most cases, the use of contact elements requires the update of the mesh graph in a fixed number of time steps. On the other hand, some of the optimization techniques increase the number of degrees of freedom in the problem, by introducing Lagrange multipliers as unknowns [2].

## II. MODEL

Domain decomposition techniques provide a powerful tool for the numerical approximation of partial differential equations [3]. Based on these techniques, we introduce an HPC-based algorithm for the numerical solution of a nonlinear contact problem between two deformable bodies. In this algorithm there are no contact elements involved and there is no need to increase the degrees of freedom of the problem. The boundary data transfer at the contact zone and the parallel implementation are the main aspect of this algorithm. We perform a two-way coupling, where in each step we solve a boundary value problem for one body using the boundary condition at the contact zone imposed by the second body. This boundary condition will be Dirichlet or Neumann. The algorithm is implemented in HPC machines and runs using several processors.

## ACKNOWLEDGMENT

This work has been done with the support of the grant SEV-2011-0067 and fellowship BES-2012-052278 of Severo Ochoa Programme, awarded by the Spanish Government.

## REFERENCES

- [1] "Computational Contact and Impact Mechanics". Tod A. Laursen. Springer, 2002.
- [2] "Computational Contact Mechanics". Peter Wriggers. Springer, 2006.
- [3] "Domain Decomposition Methods – Algorithms and Theory". A. Toselli, O. Widlund. Springer, 2005.
- [4] "A Domain Decomposition Algorithm for Contact Problems: Analysis and Implementation". J. Haslinger, R. Kucera, T. Sassi. *Math. Model. Nat. Phenom.* Vol. 4, No. 1, 2009, pp. 123-146
- [5] "A Dirichlet-Neumann Type Algorithm for Contact Problems with Friction". R. Krause, B. Wohlmuth. *Computing and Visualization in Science* 5 (3), 139-148

# Resource Management for Software Defined Data Centers for Heterogeneous Infrastructures

Marcelo Carneiro do Amaral

David Carrera

mamaral@bsc.es

**Abstract-** *Software Defined Data Center (SDDC) provides more resource management flexibility since everything is defined as a software, including the network as Software Defined Network (SDN). Typically, cloud providers overlook the network, which is configured in static way. SDN can help to meet applications goals with dynamic network configuration and provide best-efforts for QoS. Additionally, SDDC might benefit by instead of be composed by heavy Virtual Machines, use light-weight OS Containers. Despite the advantages of SDDC and OS Containers, it brings more complexity for resource provisioning. The goal of this project is to optimize the management of container based workloads deployed on Software-defined Data Centers enabled with heterogeneous network fabrics through the use of network-aware placement algorithms that are driven by performance models.*

## INTRODUCTION

The widespread adoption of cloud is driving the way in which computing solutions are delivered, allowing easier on-demand scaling of resources and pay-as-you-go [1] approaches. While the core of cloud computing shares some ideas with previous approaches like utility and grid computing, its growth and the maturity of technologies like virtualization make it much more efficient in terms of cost, maintenance, and energy consumption [2]. The cloud not only provides automated, large-scale resource management, but it also embodies a different notion of software design. It has the ability to collect, transport, process, store, and access data from anywhere [3], and empowers the use of service-oriented architectures, which are becoming the de facto approach used to increase agility and introduce best practices and patterns.

System virtualization plays a key role in cloud computing platforms, but it is not an entirely new paradigm, some of its foundational research dates back to the early 1970s [4]. The core of system virtualization relies on providing a software layer that appears equivalent to a physical machine. These virtual environments are called VMs (Virtual Machines) and provide the same inputs, outputs, and behavior that would be expected from physical hardware, and can emulate an entire OS (Operating System) [5].

Although virtual machines have plenty of benefits, they also suffer from an inherent overhead introduced by the use of multiple operating system stacks. A possible solution that addresses this kind of overhead and provides a more lightweight virtualization has emerged over the past few years: OS-level virtualization, more commonly called containers [6]. Containers rely on the idea of wrapping, limiting, isolating, controlling, and accounting the resource usage of a set of processes without the need to simulate an entire OS, what results in better performance than VMs and very fast boot times. The design of containers encompasses the idea of processes grouping. This approach leads to a new per-process model for applications, the so-called Microservice Architecture. In such a model, each container represents an independent and small application component that runs its

own (few) processes and communicates with other components using a lightweight mechanism [7]. One of the advantages of splitting the application into small components is to efficiently manage each component based on different requirements.

Containers address some problems related to the complexity of Virtual Machines. But this is not the only problem that Clouds facilities need to face: they have overlooked for years the importance of taking into consideration network properties when making workload placement decisions. Anything beyond providing isolation between workloads is usually ignored and therefore deployments are not network-aware, what results in usual problems of performance, application interference, and performance variation [8] due to poor placement decisions. Therefore, workload management is a task that requires the use of control knobs not only for workloads but also for network fabrics.

Fortunately, network management has been making progress over the last years through the development of the Software-defined Networks (SDN) paradigm, which allow for more efficient, flexible and controllable network environments. In particular, SDNs provides very relevant mechanisms for network aware management of workloads, such as resource limits [9], flow control [10] and network slicing techniques [11]. However, since SDN is very new, the integration with cloud management technologies is still on its early stages of development. The success of SDNs in the field of networks has been empowering the growing momentum of more general "Software Defined Environments" or "Software Defined Everything" environments, in which every component of the Data Center is managed from a centralized software logic. The result is the paradigm of data center infrastructures defined by software (SDDC - Software Defined Data Center), including the network (SDN), storage (SDS) and compute (SDC) infrastructure. More specifically, SDDC allow for the underlying hardware to be utilized as generalized pools of compute, network, and storage [12, 13] resources. There are many aspects of SDDC that boost its resource management, such as the ability to programmatically create, move, delete, snapshot, and restore an entire data center composed of software-defined compute, storage, and network.

In conclusion, the result of the intersection between a strong emergence of SDDCs and the growing momentum of container-based workload deployments in the Cloud is a range of new opportunities for developing new Data Center optimization strategies that need to be studied. And this is the research space that this PhD project plans to cover.

## PROPOSAL

The thesis statement of this PhD proposal is the following:  
**It is possible to optimize the management of container**

**based workloads deployed on Software-defined Data Centers enabled with heterogeneous network fabrics through the use of network-aware placement algorithms that are driven by performance models.**

For that purpose, the thesis will address three major research challenges further described as following:

- Research Challenge 1: Develop novel performance models for workloads running in containers. Although containers are getting an increasing attention by all major Cloud providers, it remains as an still unexplored space. In particular, this challenge will be split in the following specific objectives:
  - Study of novel designs for deploying applications based on the container per-process models.
  - Build performance models for containers to analyze scalability, overhead and delays associated to their use.
  - Perform a detailed comparative study between deployments based on the use of bare-metal machines, containers and virtual machines.
  - Evaluate and quantify the impact of containers on different representative workloads.
- Research Challenge 2: Develop resource provisioning strategies for workloads run in containers. The strategies will be driven by high-level objectives, expressed in the form of SLAs. In particular, this challenge will be split in the following specific objectives:
  - Design novel SLA-driven provisioning strategies based on user profiles and data center conditions.
  - Build an automated provisioning engine for vertical and horizontal scaling of container-based workloads.
  - Propose machine learning mechanisms to support the provision decision maker based on previously developed performance models.
- Research Challenge 3: Develop network-aware placement algorithms that leverage the previously developed performance models and provisioning strategies. The algorithms will take into consideration network properties (feature-wise and performance-wise) and workload characteristics to optimize the operation of SDDCs. As this is an NP-hard problem, heuristics will be used to provide approximate solutions to the optimization problem. In particular, this challenge will be split in the following specific objectives:
  - Develop heuristics for addressing the problem of making network-aware placement decisions on SDDCs.
  - Build an automated network-aware placement engine, supporting different approaches for provisioning in container-enabled SDDCs.

REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529-551, April 1955.

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: Towards a cloud definition," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 1, pp. 50–55, Dec. 2008. [Online]. Available: <http://doi.acm.org/10.1145/1496091.1496100>

[2] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IIT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599 – 616, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167739X0800197>

[3] T. Erl, R. Puttini, and Z. Mahmood, *Cloud Computing: Concepts, Technology & Architecture*, 1st ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2013.

[4] R. P. Goldberg, "Architecture of virtual machines," in *Proceedings of the Workshop on Virtual Computer Systems*. New York, NY, USA: ACM, 1973, pp. 74–112. [Online]. Available: <http://doi.acm.org/10.1145/800122.803950>

[5] M. Pearce, S. Zeadally, and R. Hunt, "Virtualization: Issues, security threats, and solutions," *ACM Comput. Surv.*, vol. 45, no. 2, pp. 17:1–17:39, Mar. 2013. [Online]. Available: <http://doi.acm.org/10.1145/2431211.2431216>

[6] R. Dua, A. R. Raja, and D. Kakadia, "Virtualization vs containerization to support paas," in *Proceedings of the 2014 IEEE International Conference on Cloud Engineering*, ser. IC2E '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 610–614. [Online]. Available: <http://dx.doi.org/10.1109/IC2E.2014.41>

[7] J. Lewis and M. Fowler, "Microservices," 2014, accessed in: 21-January-2015. [Online]. Available: <http://martinfowler.com/articles/microservices.html>

[8] A. Gupta and D. Milojicic, "Evaluation of hpc applications on cloud," in *Open Cirrus Summit (OCS)*, 2011 Sixth, Oct 2011, pp. 22–26. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2287093>

[9] B. Raghavan, K. Vishwanath, S. Ramabhadran, K. Yocum, and A. C. Snoeren, "Cloud control with distributed rate limiting," *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 4, pp. 337–348, Aug. 2007. [Online]. Available: <http://doi.acm.org/10.1145/1282427.1282419>

[10] R. Sherwood, M. Chan, A. Covington, G. Gibb, M. Flajslik, N. Handigol, T.-Y. Huang, P. Kazemian, M. Kobayashi, J. Naous, S. Seetharaman, D. Underhill, T. Yabe, K.-K. Yap, Y. Yiakoumis, H. Zeng, G. Appenzeller, R. Johari, N. McKeown, and G. Parulkar, "Carving research slices out of your production networks with openflow," *SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 1, pp. 129–130, Jan. 2010. [Online]. Available: <http://doi.acm.org/10.1145/1672308.1672333>

[11] M. Motiwala, M. Elmore, N. Feamster, and S. Vempala, "Path splicing," *SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 4, pp. 27–38, Aug. 2008. [Online]. Available: <http://doi.acm.org/10.1145/1402946.1402963>

[12] D. Ziembecki and M. Tulloch, *Integrated Cloud Platform*. Microsoft Press, 2014, accessed in: 28-January-2015. [Online]. Available: [http://download.microsoft.com/download/3/B/2/3B27DCBA-A35C-4A0B-87A7-98B956AE98BB/Microsoft System Center Integrated Cloud PlatformPDF.pdf](http://download.microsoft.com/download/3/B/2/3B27DCBA-A35C-4A0B-87A7-98B956AE98BB/Microsoft%20System%20Center%20Integrated%20Cloud%20Platform.pdf)

[13] E. W. D. Rozier, P. Zhou, and D. Divine, "Building intelligence for software defined data centers: Modeling usage patterns," in *Proceedings of the 6th International Systems and Storage Conference*, ser. SYSTOR '13. New York, NY, USA: ACM, 2013, pp. 20:1–20:10. [Online]. Available: <http://doi.acm.org/10.1145/2485732>.

# A framework for multidimensional indexes on distributed and highly-available data stores

Cesare Cugnasco  
Barcelona Supercomputing Center  
cesare.cugnasco@bsc.es

Advisor: Yolanda Becerra  
UPC-Department of Computer Architecture  
[yolandab@ac.upc.edu](mailto:yolandab@ac.upc.edu)

**Abstract**-No-relational databases are nowadays a common solution when dealing with a huge data set and massive query workload. These systems have been redesigned from scratch in order to achieve scalability and availability at the cost of providing only a reduce set of low-level functionality, thus forcing the client application to implement complex logic. As a solution, our research group developed Hecuba, a set of tools and interfaces, which aims to facilitate developers with an efficient and painless interaction with non-relational technologies.

This paper presents a part of Hecuba related to a particular missing feature: multidimensional indexing. Our work focuses on the design of architectures and the algorithms for providing multidimensional indexing on a distributed database without compromising scalability and availability.

## I. INTRODUCTION

No relational distributed databases have grown in importance in the last years, as they are the best solution to the exponential increase of the quantity of information that nowadays database system have to manage. These systems' architecture focuses on scalability and availability, so that doubling the servers doubles the performance of the overall system and at the same time, the loss of a server does not result in a system failure.

NoSQL databases have largely proved to have achieved these goals. However, they still lack many desirable functionalities and thus many applications that would profit from their potentiality, are still forced to use older and less powerful solutions.

With Hecuba [Fig. 1] our research group aims to ease the aisle of using these technologies by proving a set of tools helping the developing application on the top of distributed database.

This paper is about a particular module of Hecuba, the one in charge of create, query and management multidimensional indexes on the top of a key-value database. Multidimensional indexes are the most effective way for selecting elements from a collection when we want to limit the results to only ones that meet a particular series of constraints set by two or more characteristics (dimensions).

Research on multidimensional indexes and distributed databases have been historically divided not only by their technical background, but also by their general goal. Research on indexing algorithms focuses on how to implement new kinds of query and how to reduce the bare number of I/O requests needed to serve them. Differently, research on distributed databases, focuses on how to achieve high scalability and robust availability on systems built on multiple servers.

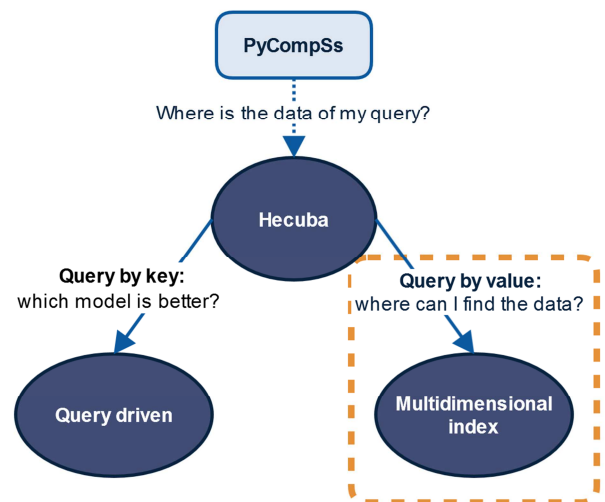


Figure 6: The architecture of Hecuba

## II. CHALLENGES

What makes the meeting of these two worlds complex, is that many of the features, such as **locks**, **isolation** or strong **consistency** - that indexing algorithms need to work correctly - have been "sacrificed" in modern distributed systems, as they were huge obstacles to achieve the desired goals of scalability and availability.

New concepts, such as **eventual consistency**, **census**, **immutability** and **versioning**, have been proposed as partial replacements of those operations, but they all came with costs and limitations.

As a result, any application built on these data systems, needs careful engineering in order to decide whether to use, one technique rather than another, to decide how to deal with boundary cases and, most importantly, when and in which cases, to pay a higher performance cost.

At the same time, the research on multidimensional indexes has continued to produce new algorithms and ad-hoc optimizations to deal with particular datasets or query requirements. Unfortunately, these optimizations are usually inefficient for other applications. For example, some algorithms can work better with a lower number of dimensions while others behave best with a high number of correlated dimensions. As a result, a one-size-fits-all approach, derived from the implementation of a single algorithm, is not sufficient to deal with all possible applications.



Even though NoSQL databases are profusely used in many and complex scenarios - from transactional to streaming and batch reporting - a distributed, and highly available database system, able to deal with arbitrary dimensional data by implementing different indexing algorithms, still does not exist to our knowledge.

### III. APPLICATIONS

Multidimensional indexes have plenty of applications in nowadays research and industry, even though their usage is limited by the scalability of the indexing system. Our work aims to overcome these limits thus enabling much wider and massive applications.

#### HPC

Molecular dynamic simulations are typical applications in High-Performance Computing (HPC) research, as long as they are used for wide cases from theoretical physics, biochemistry and biophysics. These simulations require a massive quantity of computational resources and thus they are executed in Supercomputers. Scientists use these simulations in order to validate theories and study particular physical behavior.

The typical workflow consists in: first the design of the experiment (simulation), then to reserve a time spot in a supercomputer, to execute the simulation and, once it completes, to copy the resulting trajectory files on a local computer in order to analyze the results. If instead the trajectories were directly written in a distributed multidimensional database while the simulation is running, there would be multiple benefits:

1. Researchers would be able to visualize partial results during the execution of a simulation so that they may be able to correct or to interrupt a simulation in case of incorrect behavior.
2. The database can be used as a central repository for all the simulations so that different old studies can be easily used in new research
3. The database could support complex queries such as finding similar images or trajectories across a vast database of simulation, providing a powerful research tool for researchers.

#### Low latency complex query on multiple attributes.

One of the most common mistakes approaching Cassandra as long as others NoSQL databases, is to try to issue a query that implies multiple conditions on the dataset. Indeed, the database instantaneously reports that it does not support this kind of query. Instead, the same request issued on RDBMSs would perform correctly, even though requiring a considerable amount of time.

As a workaround, in NoSQL databases it is common practice to create materialized views in order to filter by attributes, but still, it can help only with a limited subset of queries. Differently, a multidimensional index would be the perfect solution for implementing this kind of requests and, at the same, time achieving the desired constraints of availability and low latency.

#### Image and trajectory similarity research

In order to search for similar elements between images or trajectories, the existing algorithms work applying transformations on the input data so it can be described using a limited set of descriptor parameters. For example, a trajectory can be discretized using the Fourier transformation and describing the trajectory with the resulting parameters.

In such a way, a similarity query can be converted into a multidimensional range one where the boundaries are set on the descriptor parameter values.

Even though the complex part of similarity research consists in the choice of which algorithm and descriptor parameters are more effective for each specific case, the underlying system is always a multidimensional database.

#### Data warehousing

On Line Analytics Processing systems (OLAP), have historically used multidimensional indexes in order to support complex and reporting queries. In any way, for their architecture, they work with a two exclusive phases workflow. In the Extraction, Transformation and Loading phase (ETL), the data is read from an external source - usually from logs or transactional databases - and then it is elaborated and finally loaded into a multidimensional index. Only in the second phase, once the ETL phase concludes, it is possible to issue queries on the system. This approach forces the OLAP system to be unavailable to queries for a considerable amount of time. For many applications this can be unbearable, such as for fraud detection systems working in trade markets or anomaly detection systems for the network of large Internet Service Provider.

#### ACKNOWLEDGMENT

The research leading to these results has received the support of the grant SEV-2011-00067 of Severo Ochoa Program, awarded by the Spanish Government

#### PAST PUBLICATIONS

- [1] C. Cugnasco, R. Hernandez, Y. Becerra, J. Torres and E. Ayguadé, "Aeneas: a tool to enable applications to effectively use non-relational databases" *Procedia Computer Science ICCS 2013*
- [2] R. Hernandez, C. Cugnasco, Y. Becerra, J. Torres and E. Ayguadé, "Experiences of using Cassandra for molecular dynamics simulations" *Euromicro International Conference on Distributed and Network-based Processing PDP 2015*

# Mathematical Representation of the Hardware Round-Robin Scheduler Analytical Model for Single-ISA Heterogeneous Architectures

Daniel Nemirovsky<sup>†‡</sup>, Nikola Markovic<sup>†‡</sup>, Osman Unsal<sup>†</sup>, Mateo Valero<sup>†‡</sup>, Adrian Cristal<sup>†‡†</sup>

Computer Architecture for Parallel Paradigms, Barcelona Supercomputing Center (BSC), Barcelona, Spain <sup>†</sup>  
Department of Computer Architecture, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain <sup>‡</sup>  
Artificial Intelligence Research Institute - Spanish National Research Council (IIIA-CSIC), Barcelona, Spain <sup>†</sup>  
(*daniel.nemirovsky,nikola.markovic,osman.unsal,mateo.valero,adrian.cristal*)@bsc.es

## INTRODUCTION

Introduction Over the past few years, hardware chip manufacturers have been shifting strategies to overcome the well-known Power Wall which imposes the practical limitations to increasing CPU performance by raising the processor's running frequency. The most recently applied schemes to exploit high single and multi-threaded performance at lower energy costs have been to increase the number of cores per chip as well as their diversity leading to what is commonly referred to as heterogeneous multi or many core architectures. In spite of the visible gains the architectures offer, heterogeneity poses significant challenges to the OS, one of the most critical being thread scheduling[1]. In this paper we provide contributions consisting of 1) a Hardware Round-Robin Scheduling policy motivated by [2] and derivation of a mathematical model for analyzing it, and 2) an analysis and comparison of the performance results obtained through our mathematical model with ones gathered via a simulator and real physical system. HRRS policy and mathematical model In order to effectively enhance the scheduling efficiency on a heterogeneous system within a realistic scope of implementation, we have sought a scheduling scheme that leaves the OS unmodified. To achieve this, we have provided the OS with an abstracted view of the heterogeneous physical hardware as being instead composed of four identical (homogeneous) logical cores. As a consequence, the OS scheduler will now map software threads onto the logical cores which enables the OS scheduling policies and implementation to be left unaltered. Meanwhile, as the OS scheduler maps threads to the logical cores, which may happen at every software-quantum or while handling interrupts, the Hardware Round-Robin Scheduler (HRRS), which is triggered into action at every hardware-quantum, maps the logical cores to the physical cores. In essence, the HRRS remaps the software threads that are abstractly assigned to a logical core by the OS to the physical cores of the underlying hardware which actually execute the threads.

$$T = \frac{W}{(1-C)*B + (N-2)*S + (1-C*\frac{B}{S})*S} * N * Q \quad (1)$$

Equation (1) represents the basic mathematical model for execution time of all workloads on all cores where N

represents the total number of cores and hardware threads, Q is the scheduling quantum, W represents the workload per thread (equal per thread),  $B = \text{CPI\_large}/Q$ ,  $S = \text{CPI\_small}/Q$ , and  $C < 1$  represents the fraction of the scheduling quantum spent in context switching. The overhead for migrating a workload from one core to another has three components: 1) a penalty for storing and restoring the architecture state (i.e., the registers which are at most a few kilobytes of state), 2) an overhead attributed to the time it takes to drain a core's pipeline prior to migration, and 3) migration overhead due to cache compulsory misses and sharing effects. The latter is by far the largest and most important component, being at least two order of magnitude larger in terms of latency cost than the first two components combined. In order to estimate the migration overhead we have compared the overhead on a real machine (Intel i7 quad-core Haswell micro-architecture) by applying the method described by Li C. et al. [3], with our simulation results which were obtained using the Sniper simulator [4] and SPEC2006 benchmark suite. Evaluation summary In order to be able to exploit time varying workload execution behavior while keeping migration overhead small, we set the time slice granularity (i.e., hardware scheduling quantum) to be 1ms, compared to the typical 4ms quantum used by the OS software scheduler. Based on evaluations from previous work as well as our own experimental results from running on a physical machine, we have assumed a fixed 3,000 cycle penalty for storing and restoring the architecture state. Our experiments were carried out on a real machine as well as with the Sniper simulator and used different scheduling quanta show that migration overheads ranges from 3,000 to 30,000 cycles for a workload less than 500KB and up to around 300,000 cycles for workloads greater than 500KB. Our simulations show average performance benefit of 11 percent for HRRS compared to the OS software scheduler. After including these migration overheads into our mathematical model we compare the model's predicted performance of the HRRS scheme with results extracted using the simulator and got an error rate of less than 3 percent.

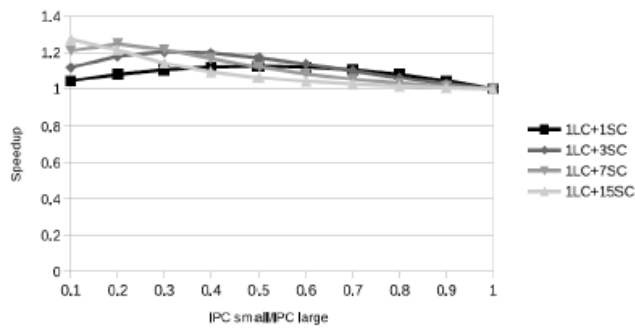


Figure 1. Analytical model comparing ideal HRS performance with an ideal Linux OS scheduler on different heterogeneous system configurations

#### REFERENCES

- [1] K. Van Craeynest et al. Scheduling heterogeneous multi-cores through performance impact estimation (pie). In Proc. ISCA, 2012.
- [2] N. Markovic et al. Thread lock section-aware scheduling on asymmetric single-isa multi-core. In IEEE CAL, 2014.
- [3] S. Li et al. Quantifying the cost of context switch. In Proc. Workshop Exp.Comput. Sci., 2007.
- [4] T. Carlson et al. Sniper: Exploring the level of abstraction for scalable and accurate parallel multi-core simulation. In Proc. SC, 2011.

# Theoretical study on the activation mechanism of AMP-kinase by means of Molecular Dynamics Simulations

C. Estarellas,<sup>1</sup> G. Leal Moraes,<sup>2</sup> A. Castro,<sup>3</sup> J. Lameira,<sup>2</sup> A. Bidon-Chanal,<sup>1</sup> F. J. Luque<sup>1</sup>

<sup>1</sup>Departament of Physical Chemistry, Faculty of Pharmacy, Campus de l'Alimentació de Torribera, Universitat de Barcelona, Av. Prat de la Riba, 171, 08921 Santa Coloma de Gramenet, Spain, <sup>2</sup>Department of Chemistry, Universidade Federal do Pará, Belem, Brasil, <sup>3</sup>Institute of Medicinal Chemistry, CSIC, Madrid, Spain  
cestarellas@ub.edu

**Abstract**-Mammalian AMP-activated protein kinase (AMPK) is a Ser108/Thr132 heterotrimeric enzyme complex (one catalytic subunit  $\alpha$  and two regulatory subunits  $\beta$  and  $\gamma$ ) with a key role as sensor in the cellular energy homeostasis. This function confers AMPK a major role in numerous metabolic disorders, such as type 2 diabetes, obesity and cancer, and explains the progressive interest as a therapeutic target. AMPK is regulated by several mechanisms including indirect and direct activators, which show clear specificity by a particular subunit. We have carried out a series of molecular dynamic simulations of the apo and holo forms of AMPK to gain insight into the mechanism of AMPK activation.

## I. INTRODUCTION

Mammalian AMP-activated protein kinase (AMPK) is a Ser108/Thr132 protein kinase with a key role as sensor in the cellular energy homeostasis [1]. This function confers AMPK a major role in numerous metabolic disorders, such as type 2 diabetes, obesity and cancer, and explains the progressive interest as a therapeutic target. AMPK is a heterotrimeric enzyme complex composed by a catalytic  $\alpha$ -subunit and two regulatory subunits known as  $\beta$  and  $\gamma$ . It is regulated by several mechanisms, including indirect activators such as metformin, rosiglitazone and resveratrol, and direct activators, such as compound A-769662 [2] (Figure 1A). Some of these activators show a clear specificity for a particular AMPK subunit, which provide a possible way to design new compounds that could activate AMPK in specific tissues, minimizing the putative negative effects of AMPK activation at whole body level [3].

The X-ray structure of AMPK bound to the direct activator A-769662 has been recently reported (PDB entry: 4CFF) [2]. We have carried out a series of molecular dynamic simulations of the apo and holo forms of AMPK, with the phosphorylated Ser108 (pSer108), to gain insight into the mechanism of AMPK activation. The results indicate that the holo structure is less flexible than the apo form, most likely due to the strong interactions formed between pSer108 and Lys29/Lys31 upon binding of the activator. Furthermore, the presence of the activator leads to a significant alteration in the shape and size of the ATP-binding pocket. Moreover, this effect is linked to the dynamical fluctuations observed between the N-terminus of the catalytic  $\alpha$ -subunit and the regulatory domain of the  $\beta$ -subunit. On the basis of these results, we speculate that these structural and dynamical features may provide the molecular basis for the synergy found experimentally between the binding of activator and phosphorylation of pSer108 in regulating the activity of AMPK.

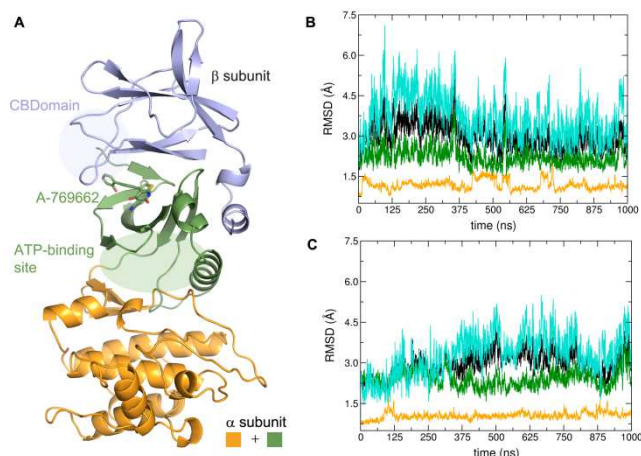


Fig. 1. A) 3D-structure of AMPK. B) RMSD of apo protein. C) RMSD of phosphorylated holo protein (complex with A769662 activator)

## II. RESULTS

### A. Conformational flexibility

MD simulations show that the complex between the phosphorylated Ser108 (pSer108) kinase domain of AMPK (pAMPK) complexed with the activator A-769662 presents an overall reduction in the conformational flexibility compared to the apo form, which is especially significant in the N-lobe region (Figure 1).

### B. Essential dynamics

The apo form shows two regions with different dynamic behavior. One region formed by the N-terminal domain of the kinase and the Carbohydrate Binding Domain (CBM) located in the  $\beta$ -subunit, which show much higher flexibility than the rest of the structure. Remarkably, in the apo form the ATP-binding loop presents a reduced flexibility (marked with a red arrow in Figure 2, top). In contrast, the dynamical behavior of the holo form involves a correlated movement of the N-terminal domain of the kinase, the  $\beta$ -subunit and the flexible loop located in the upper part of the ATP-binding site (Figure 2, bottom).

This remarkable difference in the dynamics of the apo and holo forms might be related to the presence of the activator. Thus, we speculate that it would act like a glue, by releasing the water molecules present in the binding pocket and filling the space between the N-terminal domain of the  $\alpha$  subunit and the CBM of the  $\beta$  subunit, and therefore, transferring the movement of these regions to the ATP-binding site.

It can be expected that the increased flexibility of the upper loop in the ATP-binding site should favor the population of

conformations suitable for accommodating ATP, which in turn should lead to an increase in the activity of AMPK.

relevant for keeping the activator tightly bound to the pocket at the N-terminus domain.

## CONCLUSIONS

On the basis of these findings, it can be speculated that binding of A769662 and Ser108 phosphorylation act synergistically to enhance the binding of ATP.

At present, we hypothesize that this effect implies a sizable rearrangement of the binding pocket residues that would shape the ATP-binding site, specifically leading to a conformational change in the mouth of the binding pocket that would alter the balance between open and closed states. In turn, this effect should favour the binding of ATP, and the concomitant increase in enzymatic activity.

Future studies will extend these simulations to the complexes with bound ATP in order to ascertain the impact on the ATP arrangement in the binding site and the association/dissociation events.

## ACKNOWLEDGMENT

We thank Spanish Ministerio de Innovación y Ciencia (SAF2014-57094-R), Generalitat de Catalunya (2014SGR1189), ICREA Acadèmia and Xarxa de Recerca en Química Teòrica i Computacional (XRQTC) for financial support, and the Barcelona Supercomputation Center for computational resources. C.E. thanks to Spanish Ministerio de Economía y Competitividad for the postdoctoral fellowship (FPD-2013-15572).

## REFERENCES

- [1] D. Carling, F.V. Mayer, M.J. Sanders, S.J. Gamblin. "AMP-activated protein kinase: nature's energy sensor," *Nat. Chem. Biol.* 2011, 7, 512–518.
- [2] B. Xiao, M.J. Sanders, D. Carmena, N.J. Bright, L.F. Haire, E. Underwood, B.R. Patel, R.B. Heath, P.A. Walker, S. Hallen, F. Giordanetto, S.R. Martin, D. Carling, S.J. Gamblin. "Structural basis of AMPK regulation by small molecule activators," *Nat. Commun.*, 2013, 4, 3017.
- [3] J.W. Scott, N. Ling, S.M. Issa, T.A. Dite, M.T. O'Brien, Z.P. Chen, S. Galic, C.G. Langendorf, G.R. Steinberg, B.E. Kemp, J.S. Oakhill. "Small Molecule Drug A-769662 and AMP Synergistically Activate Naive AMPK Independent of Upstream Kinase Signaling," *Chem. Biol.* 2014, 21, 619–627.

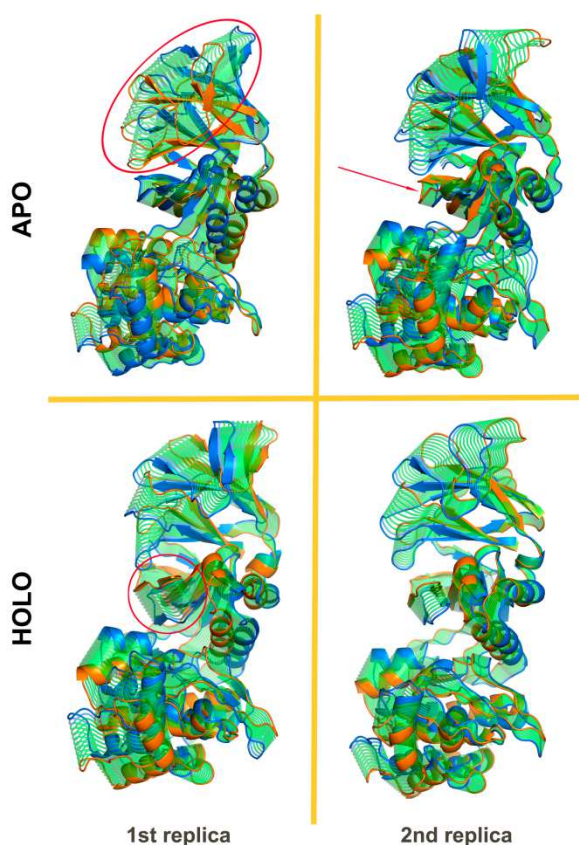


Fig. 2. Representation of the first essential mode, which contributes between 20-30% for apo and holo forms.

### C. Distances

A key molecular determinant of this structural alteration is the residue Arg83, which is highly flexible in the *apo* form, being able to form salt bridge interactions with either Asp88 or Asp136. The *apo* form shows frequent exchange between those salt bridge interactions along the MD simulation in comparison to the *holo* form, where the presence of the activator shifts the interaction pattern toward the contact with Asp88.

The activator is tightly bound in the binding pocket. Most of the contacts comprise van der Waals interactions with hydrophobic residues, but the binding is also assisted by the formation of hydrogen bonds between the amide NH unit and the carboxylate group of the Asp88 and between the hydroxy unit of the phenyl ring with the carbonylic oxygen of Gly19. In addition, the binding of A769662 is further assisted by the contact between the phosphate group of pSer108 and Lys29. However the conformational flexibility of the phosphate group allows the formation of complementary interactions with either the hydroxyl group attached to the fused bicycle of the activator or with Lys31. Overall, this interaction seems

# Influence of Temperature on the Topological Features of Inner Cavities in Cytoglobin

Constantí Seira, F. Javier Luque, Axel Bidon-Chanal

Department of Chemical Physics, Faculty of Pharmacy, University of Barcelona, Avgda. Prat de la Riba 171, 08921 Santa Coloma de Gramanet  
abidonchanalb@ub.edu

## Abstract

*Cytoglobin (Cygb) is a novel member of the globin family in man, but there is no clear evidence about its biological function. Cygb exhibits a highly complex ligand rebinding kinetics, which agrees with the structural plasticity of the inner cavities and tunnels found in the protein matrix. In this work we have examined the effect of temperature on the topological features of Cygb. To this end, the structural and dynamical properties of human Cygb are compared with those determined for the Antarctic fish *Chaenocephalus aceratus*. The results support a distinct temperature-dependence of the topological features in the two proteins, suggesting different adaptations to cold and warm environments.*

## I. INTRODUCTION

Cytoglobin (Cygb) is a fourth member of the globin family discovered in man [1]. Human Cygb consists of 190 amino acids, showing extensions of about 20 amino acids at both C- and N-termini with respect to standard globins. The amino acid sequence fits well into the conserved globin fold pattern, covering helices from A to H, and key residues such as the proximal (F8) and distal (E7) histidines, and phenylalanine CD1 at the CD corner. Along with neuroglobin, Cygb is a bis-histidyl hexacoordinated globin. In the absence of exogenous ligands, the sixth heme iron coordination site is occupied by the distal HisE7 residue in both ferric and ferrous forms. Binding of exogenous ligands is possible only with concomitant displacement of the distal HisE7, a regulatory mechanism that has been suggested to require a substantial conformational change.

The crystal structure of bis-histidyl hexacoordinated Cygb shows an extended apolar protein matrix cavity, which differs from the topology of inner cavities found in other hemoglobins [2], thus raising questions about the functional role of Cygb in cells. Besides a role as scavenger of nitric oxide or reactive oxygen species, oxygen storage and sensor protein [3, 4], a potential implication in cancer as a tumour suppressor gene has been suggested [5].

The internal volume and the topological nature of inner cavities are largely dependent on the coordination state of Cygb [6]. In particular, a substantial change in both protein dynamics and inner cavities is observed upon transition from the CO liganded to the pentacoordinated and bis-histidyl hexacoordinated species. This agrees with the complex kinetic pattern derived from carbon monoxide rebinding assays, which suggests the involvement of distinct reaction intermediates, reflecting rebinding from temporary docking sites, second order recombination, and formation (and dissociation) of a bis-histidyl heme hexacoordinated reaction

intermediate. These findings are also consistent with the expected versatility of the molecular activity of this protein.

In this work our interest is focused on the temperature dependence of the topological features of inner cavities of Cygb. Specifically, our aim is to ascertain the impact of temperature on the protein dynamics and the number and nature of inner cavities. To this end, we want to determine the structural, dynamical and topological properties of human Cygb with the protein in the Antarctic fish *Chaenocephalus aceratus*. The results obtained from molecular simulations will be complemented with data derived from flash photolysis experiments to discuss the impact of temperature on the ligand migration properties of these two proteins.

## II. METHODS

The structural model of *C. aceratus* Cygb (caCygb) was built up using SWISSMODEL taking advantage of the X-ray structures human Cygb (hCygb) and the significant sequence similarity between the two proteins (around 55% sequence identity). Then, the protein structures were refined using the protocol outlined in previous studies [6], and molecular dynamics simulations were carried out for both hCygb and caCygb at three temperatures (10, 25 and 40 degrees). For hCygb simulations the X-ray structures 3AG0 and 4B3W were used as templates in order to account for the conformational change of residue Trp151. Production runs were extended up to 1.5  $\mu$ s using Amber12 and the parm99SBildn force field. Essential dynamics was used to analyse the dynamical behaviour of the protein. The topology of inner cavities were examined by using the MDpocket program. The analysis was performed for a series of 5000 snapshots taken in 250 ns windows during the last half of the trajectory.

## III. RESULTS

Previous studies performed for hCygb at 298 K revealed that the oxygenated protein contains a big cavity. In the simulation started from the X-ray structure 4B3W there was no clear passage from the protein interior to the bulk solvent. However, the trajectory started from X-ray structure 3AG0 showed a well-defined path that connects the primary docking site to the bulk solvent, passing through residues in loop AB and the final segment of helix G. The analysis also suggested the existence of a secondary path that involves the migration through distinct pocket sites and the exit via a passage between helix A and loop EF. Remarkably, the integrity of

these paths is lost when simulations at 283 and 313 K are examined (data not shown). In contrast to the preceding findings, simulations performed for caCygb reveal a well-shaped path through the protein matrix at 283 K, which is nevertheless disrupted when the temperature is raised at 298 and 313 K (Fig. 1).

Comparison between the residues that shape the inner tunnel in the two proteins suggest a seemingly key difference, which is the mutation of Met30 in hCygb to Ser in caCygb. This change leads to the formation of a stable hydrogen-bond interaction between the Ser hydroxyl group and the indole NH unit of Trp151. This interaction fixes the orientation of Trp151, which in turn favors the formation of the tunnel in the interior of caCygb, especially at low temperatures. In contrast, the presence of Met instead of Ser in hCygb leads to an increased fluctuation of Trp151, thus contributing to blur the tunnel.

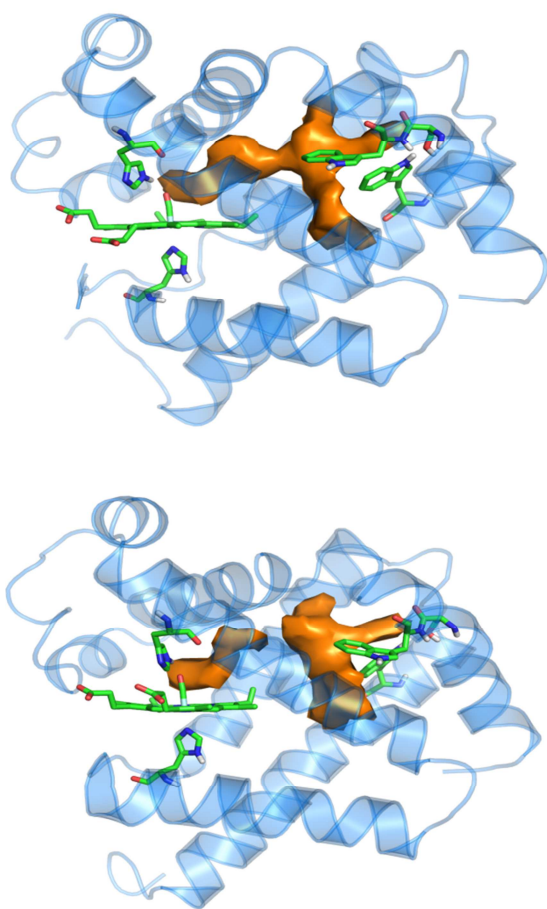


Fig. 1. Representation of the inner tunnel found in caCygb (corresponding to a frequency level of 50% as determined from MDpocket calculations; shown in orange) at (top) 283 and (bottom) 298 K. The protein backbone is shown in blue and the oxygenated haem is shown as green sticks, together with the distal and proximal histidine residues.

An additional factor appears to be the distinct flexibility of the residues around the tunnel. Several analyses performed for different layers of residues suggest that the residues shaping the tunnel are less flexible in caCygb than in hCygb. Thus, it seems that they are pre-configured through van der Waals and hydrogen-bonded interactions to properly shape the tunnel at

low temperature in caCygb, whereas the increased flexibility afforded by temperature compromises the integrity of the tunnel. In contrast, the enhanced flexibility in hCygb seems to favor the formation of interactions between Trp151 and Ile114 or alternatively Ser111, thus favoring a proper arrangement for the formation of the tunnel at higher temperature.

#### IV. CONCLUSION

The results supports a different temperature sensitivity of both caCygb and hCygb. The former contains an inner tunnel at 283 K, whereas in the latter it appears at 298 K. In the two cases the tunnel is blurred upon increase of the temperature to 313 K. Comparison of the two proteins points out that the Met30→Ser mutation might be a key factor in shaping the features of the tunnel in the interior of the proteins. Thus, the balance between the flexibility of residues in layers around the tunnel and the stability of specific interactions between residues seems to be critical for preserving the nature of the inner tunnel. Future studies will then attempt to determine the integrity of the topological features of inner cavities by introducing the corresponding mutation in the two proteins.

#### ACKNOWLEDGMENT

Financial support from the Spanish Ministerio de Innovación y Economía (Project SAF2011-27642) and the Generalitat de Catalunya (2014SGR1189) is acknowledged. We thank the Barcelona Supercomputer Center for access to supercomputation resources.

#### REFERENCES

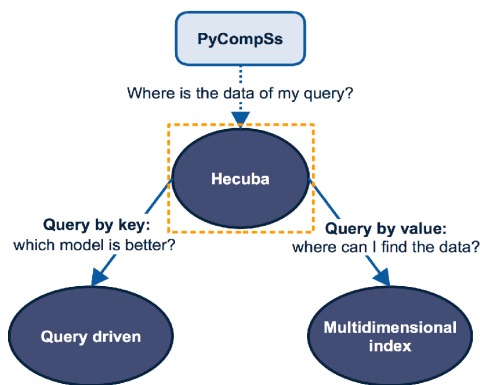
- [1] N. Kawada, D.B. Kristensen, K. Asahina, K. Nakatani, Y. Minamiyama, et al., "Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells," *J. Biol. Chem.* 2001, 276, 25318–25323.
- [2] D. DeSanctis S. Dewilde, A. Pesce, L. Moens, P. Ascenzi, et al., "Crystal structure of cytoglobin: The fourth globin type discovered in man displays heme hexacoordination," *J. Mol. Biol.* 2004, 336, 917–927.
- [3] T. Hankeln, B. Ebner, C. Fuchs, F. Gerlach, M. Kaberkamp, et al., "Neuroglobin and cytoglobin in search of their role in vertebrate globin family," *J. Inorg. Biochem.* 2005, 99, 110–119.
- [4] M. Schmidt, F. Gerlach, A. Avivi, T. Laufs, S. Wystub, et al., "Cytoglobin is a respiratory protein in connective tissue and neurons, which is up-regulated by hypoxia," *J. Biol. Chem.* 2004, 279, 8063–8069.
- [5] N. Shivapurkar, V. Stastny, N. Okumura, L. Girard, Y. Xie, et al., "Cytoglobin, the newest member of the globin family, functions as a tumour suppressor gene," *Cancer Res.* 2008, 78, 7448–7456.
- [6] M. Gabba, S. Abbruzzetti, F. Spyrikis, F. Forti, S. Bruno, et al., "CO rebinding kinetics and molecular dynamics simulations highlight dynamic regulation of internal cavities in human cytoglobin," *PLoS ONE*, 2013, 8, e49770.

# Hecuba: NoSql made easy

Guillem Alomar, Yolanda Becerra, Jordi Torres  
Barcelona Supercomputing Center  
guillem.alomar@bsc.es

**Abstract**—Non-relational databases are nowadays a common solution when dealing with huge data set and massive query work load. These systems have been redesigned from scratch in order to achieve scalability and availability at the cost of providing only a reduce set of low-level functionality, thus forcing the client application to take care of complex logics. As a solution, our research group developed Hecuba, a set of tools and interfaces, which aims to facilitate programmers with an efficient and easy interaction with non-relational technologies.

Keywords— Cassandra, PyCompSs, HPC.



## I. INTRODUCTION

Hecuba is a set of tools that facilitates programmers an efficient and intuitive interaction with non-relational databases. We have added to Hecuba the implementation of the interface necessary to provide PyCOMPSs with a Storage Backend suitable to support Big Data applications, which in our case currently is Cassandra.

Cassandra implements a non-centralized architecture, based on peer-to-peer communication, in which all nodes of the cluster are able to receive and serve queries. Each node of the cluster is assigned a token. A partitioner function uses this token to decide how data is distributed among the nodes in the shape of a ring. Data in Cassandra is stored on tables by rows, which are identified by a key chosen by the user. Cassandra stores data by rows, and one node is responsible for hosting a specific row. The target node is chosen based on the key of the row and of the token of each node by the partitioner algorithm. In order to guarantee data availability, Cassandra can be configured to keep several replicas for all data.

The mapping of a Python dictionary on a data model in Cassandra is straightforward as both consist on values indexed by keys. We have decided to map each class containing one or more Persistent Dictionary on a Cassandra table. How to implement a class backed up by Hecuba: It is just necessary to indicate that the class is a subclass of StorageObj, which is a class implemented inside Hecuba, and that contains all the methods that are necessary to access and manipulate data

backed up by Cassandra. Thus, the code of the application is mostly independent of the data backend used. Notice that programmers can add their own methods to the class definition.

## II. INTERFACE IMPLEMENTATION

Following we describe the implementation of the interface implemented by Hecuba as part of the StorageObj class. The methods exported to programmers allow to perform the following tasks:

- Make an object persistent: the implementation of this method creates the Cassandra table that it is necessary to hold the object and populates it with the data that the object had in memory.
- Object instantiation: if the constructor receives a parameter, the constructor binds the instantiated object with the corresponding Cassandra table, translating all the following accesses to that object into accesses to that table. On the contrary, if the constructor does not receive a parameter, all data associated to that new object instance will be kept in memory until the makePersistent method is executed.
- Query/Update data: if the object is in memory, the implementations of these methods are translated into the usual methods of a Python dictionary; if the object is backed by a Cassandra table, then the implementations of these methods consist on queries performed on the Cassandra table.
- Data iteration: the implementation of the keys method of a Persistent Dictionary returns a block of keys that will be the input parameter of a task. In order to enhance data locality, we decided to create the blocks of keys based on their node location. We have implemented two different iterators: one of them to create the different blocks of keys and the other one to traverse all the keys in a block (Figure 1).

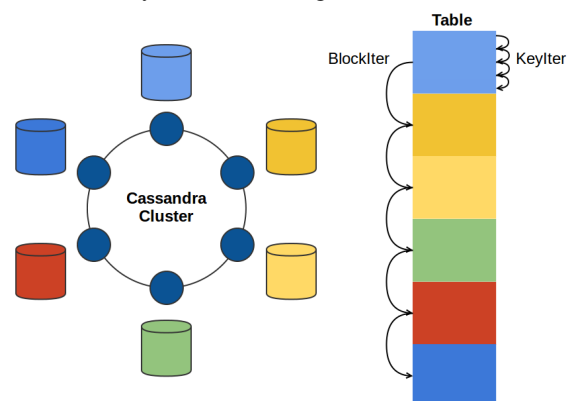


Figure 1: Representation of Hecuba Iterators.



- Delete a persistent object: this operation is implemented by the delete persistent method of the StorageObj and it just deletes from the database the table that was backing the object.

### III. DATA LOCALITY

PyCompSs not only enables an easy way to define which functions will be parallelized, but also allows Hecuba to exploit data locality.

PyCompSs has at its disposal the location of all blocks of data that Hecuba has stored in Cassandra. Whenever a task needs to be run in a node, PyCompSs will decide to which node this task will be sent, depending on the location of the stored data that the task needs.

PyCompSs has a score function, which gives a higher priority to nodes with the most information needed by the task, and in case that this first node is not available for any reason, the task will be sent to the following node with higher priority. As data in Cassandra is replicated in many nodes, a task that needs more than one block of data will be sent to the node with the most needed blocks available (Figure 2).

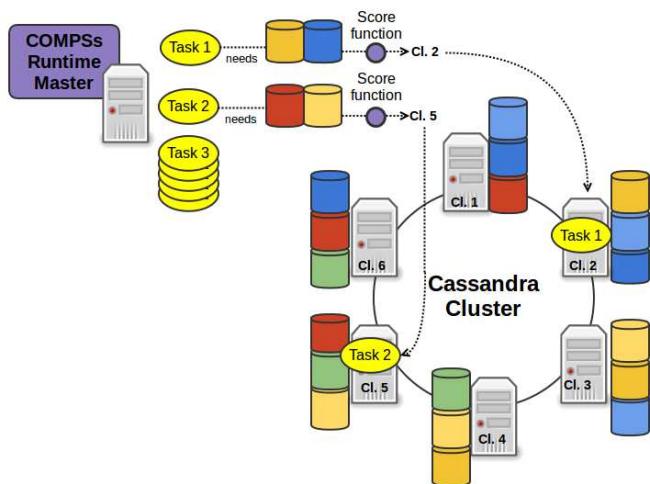


Figure 2: Representation of how COMPSs exploits data locality in an examples with tasks needing 2 blocks of data in a Cassandra Cluster with data replicated in 3 nodes.

### ACKNOWLEDGMENTS

The research leading to these results has received the support of the grant SEV-2011-00067 of Severo Ochoa Program, awarded by the Spanish Government.

# Scaling Irregular Array-type Reductions in OmpSs

Jan Ciesko<sup>1</sup>, Sergi Mateo<sup>1</sup>, Xavier Teruel<sup>1,2</sup>, Vicenç Beltran<sup>1</sup>,  
Xavier Martorell<sup>1,2</sup>, Rosa M. Badia<sup>1,2</sup> and Jesús Labarta<sup>1,2</sup>  
<sup>1</sup>Barcelona Supercomputing Center  
<sup>2</sup>Universitat Politècnica de Catalunya  
{jan.ciesko, sergi.mateo, xavier.teruel, vicenc.beltran,  
xavier.martorell, rosa.m.badia, jesus.labarta}@bsc.es

**Abstract** – Array-type reductions represent a frequently occurring algorithmic pattern in many scientific applications. A special case occurs if array elements are accessed in a non-linear, often random manner, which makes their concurrent and scalable execution difficult. In this work we present a new approach that consists of language- and runtime support to facilitate programming and delivers high scalability on modern shared-memory systems for such irregular array-type reductions. A reference implementation in OmpSs, a task-parallel programming model, shows promising results with speed-ups up to 15x on the Intel Xeon processor.

## I. INTRODUCTION

Irregular array-type reductions, also referred to as scatter-update, represent memory updates over an array type. The non-atomic operation as well their dynamic memory access pattern make their concurrent execution non-trivial and require careful handling to achieve scalability and correctness. Fig. 1 shows a scalar, regular and irregular array type reduction over *target* where in case of an irregular array-type reduction, the update positions depend on indexes generated by a function *f*. It becomes obvious that algorithms containing an irregular

```
loop construct over i          loop construct over i
{                               {
  target = op(target, RHS);      target[i] = op(target[i], RHS);
}                               }

loop construct over i
{
  j = f(i);
  target[j] = op(target[j], RHS);
}
```

Figure 7: Different types of reductions require different parallelization techniques

array-type reduction are cache inefficient due to distant memory accesses and consequently execution performance is bound to the speed of the memory subsystem. Further in order

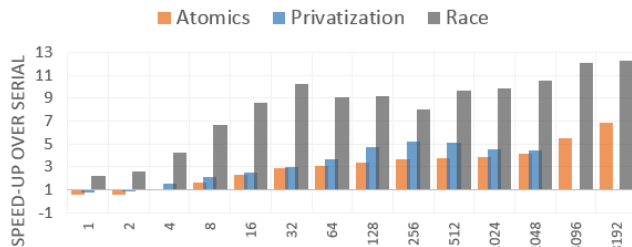


Figure 8: Application scalability of different approaches compared to a plain implementation with data races showing achievable performance on a single MareNostrum node

to avoid a race condition where multiple threads perform an update of a single memory location at the same time, accesses either need to be synchronized (via thread synchronization or memory barriers such as atomics), ordered [1] or redirected [2].

Access redirection to a thread-private copy of the reduction target is a common approach that eliminates the need for access synchronization. While this works well for scalar types, it becomes expensive for arrays and even useless for large data sets.

Figure 2 shows the performance impact of atomics and array privatization in the *RandomAccess* [3] kernel benchmark over serial execution running with 16 threads and different problem

```
for(j = 0; j < num_tasks; j++) {
  INT_TYPE seed = ran[j];
  #pragma omp task concurrent (Table[0:N])
  {
    for( INT_TYPE i = 0; i < N_block; ++i ) {
      seed = LCG_MUL64 * seed + LCG_ADD64;
      INT_TYPE pos = seed >> (bitSize - logTableSize);
      #pragma omp atomic
      Table[pos] ^= seed;
    }
  }
  #pragma omp taskwait
}
```

Figure 9: *RandomAccess* implemented with atomics, showing the compiler generated instruction that triggers a memory barrier

sizes. Its source code is shown in Figure 3. Consequently a new approach is needed that improves cache efficiency, reduces lock contention, eliminates memory barriers and is applicable on large input data sets at the same time. It turns out that by redirecting accesses to an array of thread-local linear buffers to temporarily store memory updates of a certain memory region of the reduction array and to flush the buffers when they are full is a simple yet efficient technique to meet the above requirement. We present this approach in more detail in the next chapter.

## II. RUNTIME SUPPORT

To support irregular array-type reductions in OmpSs, we developed a new approach called Privatization with In-lined Block-ordered Reductions - *PIBOR*. In this approach all memory accesses to the original reduction array are redirected to a thread-private buffer. While this is comparable to regular privatization, the buffer is filled linearly, is limited to a pre-set size and additionally stores the memory address along the data of each access. Once the buffer is filled up, the owning thread reduces the buffer to global memory.

Typically writing out data to global memory requires to perform a global lock over the entire data structure which serializes execution. We prevent this by assigning buffers to discrete memory regions of the reduction array. In this case accesses to the original reduction array in a certain region are stored in the correspond buffer. In case the buffer runs full, the owning thread tries to acquire a lock that protects only the particular memory region of the global array. Buffers corresponding to different regions can now be reduced in parallel and by increasing the number of regions, the effect of lock contention over a single region can be efficiently mitigated. A schematic overview of an application that runs  $N$  tasks on  $N$  threads and performs a reduction over an array divided into  $M$  locations is shown in Figure 4.

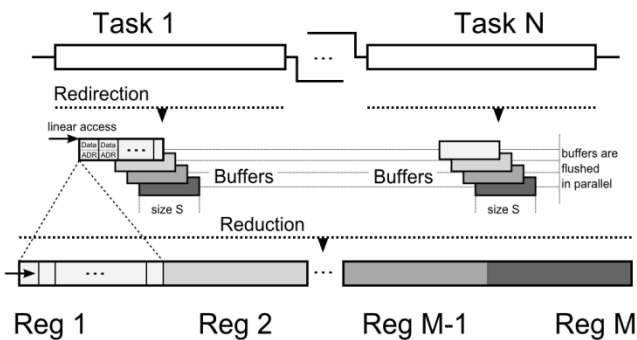


Figure 10: A schematic view of PIBOR showing thread-local buffers and their corresponding array regions

Since buffers correspond to different regions, each memory access needs to be inspected in order to determine its correct buffer. We do so by applying a hash function on the address of the accessed element. The entire process is shown in Figure 5.

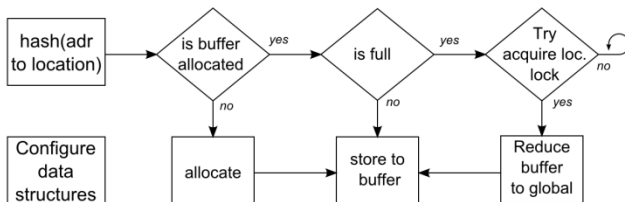


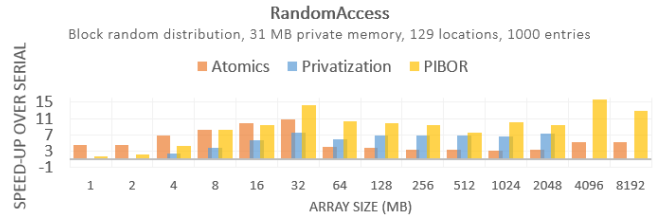
Figure 5: Execution diagram showing the process of privatization and reduction to global data

### III. LANGUAGE SUPPORT

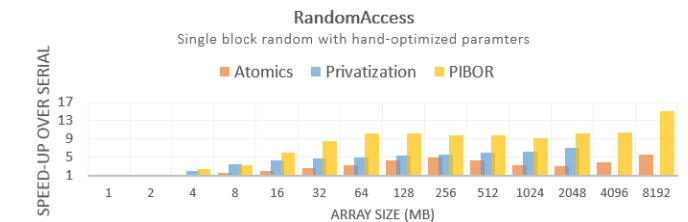
High programmability while maintaining execution transparency is a key requirement for modern programming models. Since PIBOR is conceptually related to privatization, an approach often found in declarative programming models such as *OpenMP* [4], its introduction puts minimal effort on front-end compilers, current specifications and user understanding. The following shows language support for array reductions in *OmpSs* and its compiler generated code.

```
#pragma omp task reduction (array[0:N])
{
  array[pos1] ^= RHS;      T * __tmp;
  array[pos2] ^= RHS;      RT.request( & array[posn], __tmp)
  ...                      * __tmp = RHS;
  ...
}
```

## IV. CASE STUDIES



We evaluate the presented approach using *RandomAccess* on a single *MareNostrum* 16-way SMP node. *RandomAccess* is a kernel benchmark that allows to simulate different access patterns. In particular we looked at three representative scenarios: uniform random distribution, block random distribution and single block random distribution where all accesses are restricted to a single memory region. Performance



results are shown in Figure 6 and 7.

Figure 6: PIBOR requiring only 31 MB to achieve speed-ups of up to 15x on a SMP node. Block random distribution favors atomic updates for small (cacheable) problem sizes due to reduced impact of memory barrier and no cache-line invalidations

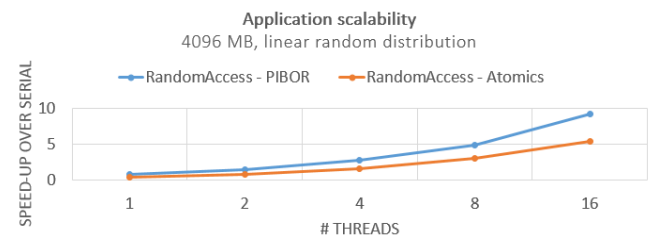


Figure 7: Performance scalability of PIBOR compared to atomics

## V. CONCLUSION AND FUTURE WORK

The presented approach scales by redirecting previously random memory accesses of a region into a linear buffer. Since each buffer corresponds to a memory region of the reduction array, buffers can be flushed in parallel. Further work is directed towards automated tuning of location granularity and buffer sizes and experiments on different processors including *Xeon Phi* and *Power8*.

### ACKNOWLEDGMENT

I would like to thank all my coauthors for their invaluable insights and their patience when exposed to my ideas during countless meetings.

### REFERENCES

[1] H. Yu and L. Rauchwerger, Adaptive Reduction Parallelization, 14<sup>th</sup> ACM Intl. Conf. on Supercomputing, 2000.

[2] H. Han and C.-W. Tseng, A Comparison of Parallelization Techniques for Irregular Reductions, 15th IEEE Int'l. Parallel and Distributed Processing Symp. (IPDPS'2001)

[3] RandomAccess, HPC Challenge,  
<http://icl.cs.utk.edu/projectsfiles/hpcc/RandomAccess>, 2015

[4] OpenMP, Application Programming Interface, [www.openmp.org](http://www.openmp.org), 2015

# DLP Acceleration on General Purpose Cores

Milovan Duric, Oscar Palomar, Osman Unsal, Adrian Cristal, Mateo Valero  
Barcelona Supercomputing Center  
{first}.{last}@bsc.es

**Abstract-** High-performance and power-efficient multimedia computing drives the design of modern and increasingly utilized mobile devices. State-of-the-art low power processors already utilize chip multiprocessors (CMP) that add dedicated DLP accelerators for emerging multimedia applications and 3D games. Such heterogeneous processors deliver desired performance and efficiency at the cost of extra hardware specialized accelerators. In this paper, we propose dynamically-tuned vector execution (DVX) by morphing one or more available cores in a CMP into a DLP accelerator. DVX improves performance and power efficiency of the CMP, without additional costs for dedicated accelerators.

## I. INTRODUCTION

The performance targets of mobile processors are largely driven by the requirements of real time multimedia applications and physical simulations from 3D games. The abundant data level parallelism (DLP) in these applications offers potential for power-efficient acceleration. In this paper we propose dynamically-tuned vector execution (DVX) that exploits DLP on lightweight mobile processors. DVX performs operations over configurable large vector operands, by enabling the resources of general purpose cores to execute vector instructions like classic vector processors. As opposed to classic vector processors that use a fixed number of vector lanes, DVX tunes the size of its execution substrate to a particular workload phase. For this feature, DVX adds lightweight threads of vector instructions controlled by hardware. The threads scale the execution of vector instructions over multiple cores with minimal performance overheads.

## II. DVX ON GENERAL PURPOSE CORES

DVX morphs the general purpose cores to execute vector instructions that perform compute, register and memory operations over configurable large vector operands. Atomic Instruction Blocks (AIB) with the vector compute instructions allocate the compute resources of one or more cores to perform computation over the vector operands. One instance of such vector AIB is fetched and decoded once, but repeatedly executed multiple times to compute the large operands. The vector operands are accessible only outside the AIB, as the block input or output operands. The vector compute instruction inside the AIB process the operands slice-by-slice in a pipeline, as it is shown in Figure I. By processing large operands in slices (sub-vectors), each new execution of the vector AIB requires only the available

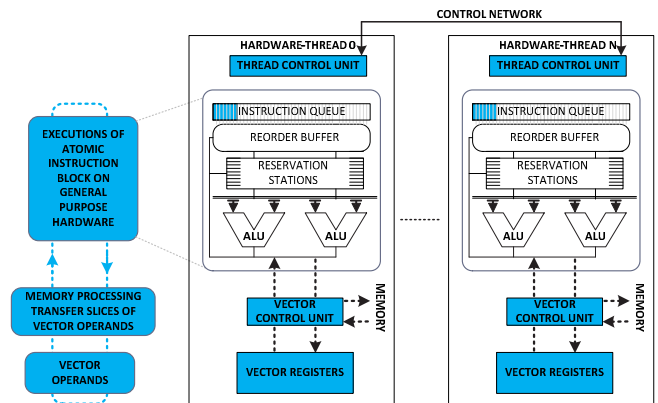


Figure 1. DVX on general purpose CMP utilizes a vector AIB to allocate the existing general purpose hardware of one or more cores that performs the compute operations over vector operands.

general purpose hardware. The hardware processes vector slices as same as scalar operands like in [1] and does not require any modifications. The number of executions is dynamically configured for each vector AIB, to resemble the execution of classic vector instructions that perform one operation over a configurable number of vector elements.

Rather than scheduling vector AIBs to vector cores like in [2], the AIBs leverage the existing issue logic, reservation stations (physical registers) and ALUs of the CMP substrate, as shown in Figure I. The reservation stations keep the temporal results between the instructions in a vector AIB. *Vector registers (VRs)* are an addition to the general purpose core, and they only hold the input/output operands of the AIB. This approach significantly reduces the size of a vector register file, the power-hungry structure that limits the applicability of vector design beyond supercomputer processors. *Vector control unit (VCU)* is a DVX dedicated unit added to the general purpose core to manage DVX. The VCU decouples the execution of vector memory instructions from the executions of vector compute instructions on the allocated resources. The decoupled memory execution is specialized to tolerate memory latency and provide sophisticated addressing modes required in diverse DLP workloads. Beside memory processing, the VCU reads/writes slices of large vector operands in the VRs and transfers them per execution of compute AIBs.

DVX allows for tuning of its resources to different application requirements by scaling the vector execution over multiple cores. Instead of the existing software approaches, DVX introduces lightweight threads controlled by hardware to dynamically spawn the execution of vector instructions over multiple cores. Such threads avoid startup and bookkeeping overheads in each software thread. DVX adds a simple dedicated *thread control unit (TCU)* to each core of the CMP and an additional on-chip-network between the cores to control the threads with vector instructions.

### III. CONCLUSIONS

In this paper, we have proposed dynamically-tuned vector execution that accelerates DLP workloads on the set of general purpose cores. The low cost DVX results evaluated in this work makes it a promising alternative to DLP accelerators in commercial processors.

### REFERENCES

- [1] Duric, et al. "EVX: Vector Execution on Low Power EDGE Cores." DATE'14
- [2] Krashinsky, R., et al. "The vector-thread architecture." ISCA'04

### MY PUBLICATIONS

- [1] "ReCompAc: Reconfigurable Compute Accelerator" RECONFIG'13
- [2] "EVX: Vector Execution on Low Power EDGE Cores." DATE'14
- [3] "DVX: Dynamic-Vector Execution on a General Purpose EDGE Chip Multiprocessor." SAMOS'14

# Methodology to select a I/O configuration (hardware resources and stack software) in cloud platform

Pilar Gomez-Sanchez, Dolores Rexachs  
Computer Architecture and Operating Systems Department (CAOS)  
Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain.  
{pilar.gomez, dolores.rexachs}@uab.cat

**Abstract-** Currently, there is an increasing interest about the cloud platform by the High Performance Computing (HPC) community, and the Parallel Input/Output (I/O) for High Performance Systems is not an exception. However, in cloud platform increase the number of parameters that user can select to I/O system. For this reason, we propose a methodology to help the user to select a configuration.

## I. INTRODUCTION

Nowadays, the interest about the cloud computing platform is increasing. The scientific community have interest about the cloud computing because some benefits of clouds are that users can acquire and release resources on-demand and they can configure and customize their own Virtual Cluster (VC) [1]. Parallel scientific applications that use parallel I/O can benefit of these platforms, because the user can create and configure the I/O system considering the application requirements which represents an advantage over the traditional HPC-IO systems. But, in cloud platform increase the number of parameters that user can select to I/O system.

### A. Methodology

For this, we propose a methodology to configure and select a configuration depending application requirements and user requirements. Our methodology has six steps:

- 1) Application Parallel I/O Characterization: The I/O characteristics are represented by an I/O model. We trace the parallel application with PAS2P-IO [2] and the traces are analyzed to obtain the I/O model. The I/O model allows us to meet the minimum application requirements.
- 2) Creation and Configuration of the Virtual Clusters: A VC is represented by the components shown in Table I. We can create a VC quickly with StarCluster [3]. We apply the following considerations as a starting point on the selection of the components for a VC that meets the user requirements.
- 3) Characterization of the Virtual Clusters: We use the IOzone [4] benchmark to obtain the average values for the transfer rate at local file system level. Normally, the user can check if he will obtain the waited performance.
- 4) Performance Evaluation on the Virtual Clusters for the application I/O model: IOR [5] benchmark evaluates the performance at global file system level. IOR is designed to measure parallel file system I/O performance at both the POSIX and MPI-IO level. The IOR performs writes and reads to/from files under several sets of conditions and reports the resulting throughput rates.

We analyze the access patterns of the I/O model at phases

Parameters	Description
Instance Type(*)	Number of cores, processor capacity, RAM memory size
Number of instances (*)	
Number of I/O nodes (-)	Data servers and metadata server.
Storage type(+)	Temporal and/or persistent.
Device type temporal (+)	HDD or SSD.
Device type persistent (+)	HDD or SSD.
Capacity of temporal storage(+)	As minimum the storage capacity required (expression 4).
Capacity of persistent storage(-)	
Network performance (+)	Low, Moderate, High, Unknown.
I/O Library (-)	MPI, NetCDF, pnetcdf, HDF5.
Local file system (+)	File system Linux ext3, ext4, xfs, etc.
Global file system (-)	Parallel, Distributed or Network file systems.
Stripe size (-)	Related by the parallel file system.

(\*) the parameters which can be selected by the user, (-) the parameters that the user must configure manually, (+) the parameters that the user cannot change because they are by default depending on instance type.

level and proposed an IOR configuration based on the application I/O model, where the relevant parameters are the numbers of processes (np), the number of segments (-s), block size (-b) and transfer size (-t). Table II shows input parameters for IOR based on the I/O model phase. The output of this process is the transfer rate expressed in MB/ sec, named  $BW_{CH}$ , and I/O time for application I/O model. The I/O model has been extracted executing the application once in the cluster.

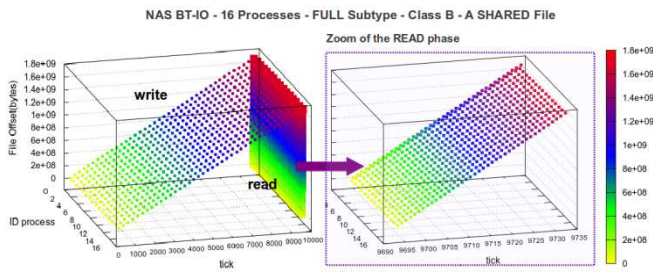
5) Cost Evaluation of the Virtual Clusters: Performance obtained using IOR for the application I/O model is used to calculate the cost. The total cost for a specific VC is composed of a variable cost and a fixed cost.

6) Comparison of the Performance-Cost Ratio for the Virtual Clusters: The performance and the cost for the VCs are presented to the user to simplify the decision making. To compare the performance-cost of the different VCs.

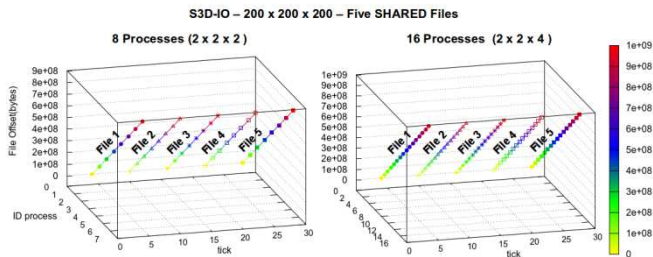
## II. EXPERIMENTAL RESULTS

In this section, we present the performance evaluation and the cost analysis for two scientific application such as I/O kernels NAS BTIO and S3D-IO that present different I/O access patterns. BT-IO and S3DIO have been traced using PAS2P-IO to extract their I/O models. The I/O model has been extracted executing the application once in the cluster.

1) Application Parallel I/O Characterization: The Figure 1 shows the BT I/O model and the Figure 2 shows the S3D-IO I/O model.



**Fig. 1.** The left picture shows the I/O model for the application and the right picture shows a zoom on the read operations. It can be observed that write and read are done in the same file offset. The application uses a shared file. Each MPI process performs a write operation every 122 communication events. This is done 40 times, and after, each process performs 40 read operations consecutively.



**Fig. 2.** The left picture shows the I/O model for 8 processes with a workload  $200 \times 200 \times 200$ . The application uses five shared files. All MPI processes write once on the File 1, after all processes write on the File2, and so on. This access pattern is representing the checkpointing process for the S3D application. The same behavior is observed in the right picture for 16 processes.

2) Creation and Configuration of the Virtual Clusters: We select some instances depending application and user requirements. See Table II and Table III.

Table II. Characteristics of the Amazon's instances selected

Instances	Processor	CPU	RAM (GB)	Storage(GB)	AWS Ireland (\$ Per Hour)	AWS Virginia (\$ Per Hour)
m1.small	Intel Xeon Family	1	1.7	1x160	0.047	0.044
m1.large	Intel Xeon Family	2	7.5	2x420 HDD	0.190	0.175
c3.xlarge	Intel Xeon E5-2680 v2 2.8 GHz	4	7.5	2x40 SSD	0.239	0.210

3) Characterization of the Virtual Clusters: We use this phase to meet the instance performance.

4) Performance Evaluation on the Virtual Clusters for the application I/O model: For example, Table IV show the input parameters to configure IOR from the I/O model phases of the BT-IO. From this process, we obtain the transfer rate ( $BW_{CH}$ ) and execution time for the BT-IO model. These values are used to calculate the variable cost.

5) Cost Evaluation of the Virtual Clusters: in this step we calculate the cost to use a determinate I/O configuration.

6) Comparison of the Performance-Cost Ratio for the Virtual Clusters: Finally, we compare what configuration is better depending of Performance-Cost ratio. We show the obtained result to BT in Figure 3

### III. CONCLUSIONS

In this paper, we have proposed a methodology to help the user to configure and select the configurations in a cloud platform depending of application requirements and the user requirements. Our methodology proposes to customize the IOR benchmark with the I/O model to evaluate quickly different configurations and reducing the evaluation cost because it is not necessary to execute the application every time. As future work, we will continue analyzing the influence

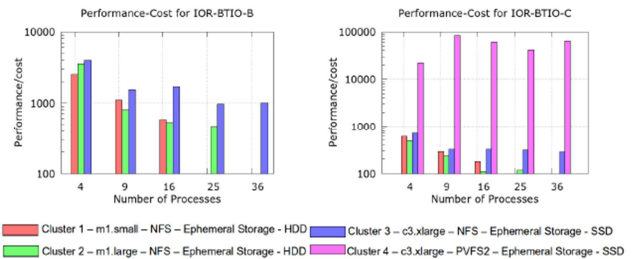
of the different components for the Virtual Cluster configuration.

Table III. Descriptive characteristics of the virtual clusters configured for the experiments (Step2)

I/O components	Virtual Cluster 1	Virtual Cluster 2	Virtual Cluster 3	Virtual Cluster 4
Instance Type	m1.small	m1.large	c3.xlarge	c3.xlarge
Number of Instances	17	17	11	11
Storage Type Temporal	Ephemeral	Ephemeral	Ephemeral	Ephemeral
Storage Type Persistent	EBS	EBS	EBS	EBS
Device Type Temporal	HDD	HDD	SSD	SSD
Device Type Persistent	HDD	HDD	HDD	HDD
Capacity of Temporal Storage	160GB	420GB	40GB	300GB
Capacity of Persistent Storage	8GB	8GB	8GB	16GB
Networking Performance	Low	Moderate	High	High
Number of data servers	1	1	1	1
Number of Metadata Server	1	1	1	1
File system Local	ext3	ext3	ext3	ext3
File system Global	NFS	NFS	NFS	PVFS2
Stripe Size				64KB
I/O library	mpich2_pnetcdf	mpich2_pnetcdf	mpich2_pnetcdf	mpich2_pnetcdf
EBS Fixed Cost EU(\$ per GB-month)	0.55	0.55	0.55	0.55
EBS Fixed Cost US-East(\$ per GB-month)	0.50	0.50	0.50	0.50

Table IV. IOR Input parameters from the I/O model phases of the NAS BT-IO subtype FULL-Collective Operations and  $s=rep=40$ . Outputs for the virtual Cluster 3.

np(ldPh)	b=rs(ldPh) (MB)	t=rs(ldPh) (MB)	$BW_{CH}$ (MB/s)	Time (sec)
<b>Class B</b>				
4	10.0	10.0	104	15.5
9	4.5	4.5	91	17.7
16	2.5	2.5	96	16.9
25	1.6	1.6	73	22.6
36	1.1	1.1	74	22.2
<b>Class C</b>				
4	40.6	40.6	87	74.5
9	18.0	18.0	83	77.8
16	10.1	10.1	84	77.2
25	6.5	6.5	82	79.0
36	4.5	4.5	78	82.9



**Fig. 3.** Performance-Cost ratio of the four Virtual Clusters using IOR configured for the BT-IO. The left picture corresponds to Class B and the right picture to Class C. Results are shown in logarithmic scale. Virtual Clusters with experiments without results are limited by storage capacity or the parallel degree required. Class B was not tested on Virtual Cluster 4 because the I/O workload is small for its I/O system.

### ACKNOWLEDGMENT

This research has been supported by the MINECO (MICINN) Spain under contract TIN2011-24384. The research position of the PhD student P. Gomez has been funded by a research collaboration agreement, with the "Fundación Escuelas Universitarias Gimbernat".

### REFERENCES

[1] M. Liu, J. Zhai, Y. Zhai, X. Ma, and W. Chen, "One Optimized I/O Configuration Per HPC Application: Leveraging the Configurability of Cloud," in Proceedings of the Second Asia-Pacific Workshop on Systems. ACM, 2011, pp. 15:1–15:5.

[2] S. Méndez, J. Panadero, A. Wong, D. Rexachs, and E. Luque, "A New approach for Analyzing I/O in Parallel Scientific Applications," in CACI12, 2012, pp. 337–346.

[3] StarCluster. (2014) An Open Source Cluster-Computing Toolkit for Amazon's Elastic Compute Cloud (EC2). <http://star.mit.edu/cluster/>

[4] W. D. Norcott. (2006) IOzone Filesystem Benchmark. <http://www.iozone.org/>

[5] W. Loewe, T. McLarty, and C. Morrone. (2012) IOR Benchmark. [https://github.com/chaos/ior/blob/master/doc/USER\\_GUIDE](https://github.com/chaos/ior/blob/master/doc/USER_GUIDE)



# Understanding Scientific Application's Performance

Oriol Tintó Prims

Universitat Autònoma de Barcelona (UAB), Institut Català de Ciències del Clima (IC3)  
oriol.tinto@ic3.cat

**Abstract-** *Improve Earth System model's scalability in High Performance Computing infrastructure is crucial to achieve efficiency in upcoming supercomputers. We study how to increase the scalability of the widely used Oceanic Model of the Nucleus for European Modelling of the Ocean (NEMO).*

## I. INTRODUCTION

A wide range of scientific fields are taking advantage of High Performance Computing and simulation is nowadays a key tool for many of these fields. This is especially true in Climate Science. The models in this field are using a massive amount of computation resources and not always with the proper efficiency.

In my research, I am focusing on the NEMO (Nucleus for European Modelling of the Ocean) model. The objective of this research is to improve its scalability and adapt the model to be executed using a bigger number of processors, doing a little step to exa-scale climate simulation.

## II. THE NEMO MODEL

NEMO is a state-of-the-art modeling framework for oceanographic research, operational oceanography seasonal forecast and climate studies. The NEMO research community is using a huge amount of computing resources and therefore a computational development of the application is mandatory in order to achieve a better usage of these resources.

### A. Parallelization

The NEMO model is parallelized using a Domain Decomposition Method that splits the global domain into smaller sub-domains and allows computing each of these sub-domains in a different processor core. The neighbour sub-domains require frequent communication in order to update the boundary conditions, and this is done using MPI.

### B. Configurations

NEMO is an Ocean Global Circulation Model and it is possible to run different configurations and resolutions. The most used configuration is the ORCA configuration, a global tripolar grid. There are several resolutions available for the ORCA configuration, and the trend is to keep increasing the resolution to improve the quality of the simulations. When the resolution of the model's grid increase, the resources needed to perform simulations also increase.

TABLE I  
ORCA RESOLUTIONS

Resolution	Domain Size			
	Latitude Points	Longitude Points	Vertical levels	Total Grid Points
ORCA2	182	149	31	1,09 M
ORCA1	362	292	46	4,86 M
ORCA025	1442	1022	75	110,53 M
ORCA012	4322	3059	75	991,57 M

In addition, if we increase the spatial resolution we must increase also the temporal resolution for numerical stability issues, and therefore the computational resources required will increase even more.

## III. PERFORMANCE ANALYSIS

In order to improve the scalability of the model the first step is perform a deep analysis of the application performance to know how it behaves and which are the bottlenecks constraining model's scalability. To do so I am been using the Barcelona Supercomputing Center Performance Tools . From the tests done using different configurations it was possible to find which is the scalability limit of each configuration and find out that the more important bottleneck constraining is the communication.

## IV. OPTIMIZATIONS

To solve the problems found we studied different solutions. Since the main bottleneck is the communication, the simplest optimization is to reduce the number of times that the model requires to communicate wherever is possible. It was possible to find two ways to reduce the communication overhead. By using this minor optimizations we could increase the speed-up of the ORCA1 configuration from 9,5 to 11,2 when using 16 nodes of 16 processor cores each one, improving the efficiency by a 10%. Future optimizations will focus the possibility to reduce communication overhead instead of reducing the number of communications by using asynchronous communication and MPI plus shared memory approaches.





**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*



**EXCELENCIA  
SEVERO  
OCHOA**