

1st BSC Doctoral Symposium

2014

27th & 28th May, 2014

Book of abstracts



**Barcelona
Supercomputing
Center**



**EXCELENCIA
SEVERO
OCHOA**



1st BSC Doctoral Symposium

Book of Abstracts

Published at UPC Commons (<http://upcommons.upc.edu>) to
protect its contents and increase its visibility.





Acknowledgements

The BSC Education & Training team gratefully acknowledges all the experts and PhD candidates contributing to this Book of Abstracts and participating in the 1st BSC Doctoral Symposium.

BSC Education & Training team
education@bsc.es





CONTENTS

WELCOME	7
PROGRAMME	9
INVITED SPEAKERS	11
PRESENTERS	13







Welcome

WELCOME

It is my pleasure to welcome all the PhD students and advisors participating in the 1st BSC Doctoral Symposium.

The goal of the symposium is two-fold: first we aim at providing a framework to share research results of the thesis that are being developed by PhD students at BSC; and second, we offer training sessions on topics and skills that will be useful to them as future researchers and professionals.

The symposium has been conceived in the framework of the Severo Ochoa Program at BSC, following the project aims regarding the talent development and knowledge sharing. Taking that into account, the symposium provides an interactive forum for PhD students considering both the ones just beginning their research and the others who have progressed far enough to share some results.

I am very grateful to the BSC directors for supporting the creation of the symposium, to advisors, group leaders and department directors for doing a wonderful job encouraging the participation of the students in the first edition of the BSC doctoral Symposium. Moreover, I wish to specially thank the two invited speakers, Carles Comes and Fernando Cucchietti, for their willingness to share with us their deep experience on legal protection of computer programs and how to improve public presentations.

And last but not least, I would like to thank all PhD students for their presentations and effort and I wish them all the best for their on-going PhD work.

Dr. Maria Ribera Sancho
Responsible of BSC Academic Programs





Programme

PROGRAMME

First Day 27th May 2014		
9.30h	Welcome Address by Mateo Valero, BSC Director	
9.45h	Talk on 10 things you shouldn't do in a presentation	Invited Speaker: <i>Fernando Cucchietti, BSC</i>
10.45h	Poster Session and coffee	<i>Mercè Planas, LS</i> <i>Milos Panic, CS</i> <i>Nikola Rajovic, CS</i> <i>Lluis Gifré, CS</i> <i>Marc Guevara, ES</i>
11.30h	Cancer genomics at BSC	<i>Santiago González, LS</i>
12:00h	Atomic Dataflow Model	<i>Vladimir Gajinov, CS</i>
12:30h	Air Quality Impacts of Electric Vehicles in Barcelona	<i>Albert Soret, ES</i>
13:00h	CODOMS: Protecting Software with Code-centric Memory Domains	<i>Lluis Vilanova, CS</i>
13.30h	Lunch Break	
14.30h	PELE applications	<i>Israel Cabeza de Vaca, LS</i>
15:00h	Techniques for Improving the Performance of Software Transactional Memory	<i>Srdjan Stipić, CS</i>
15:30h	Coffee Break	
15:45h	Modelling volcanic ash dispersal and its impacts on civil aviation	<i>Chiara Scaini, CASE</i>
16:15h	Supporting Scatter/Gather Tasks in Manycore Architectures	<i>Tassadaq Hussain, CS</i>
16:45h	1 st day closing remarks	
17:00h	Adjourn	

Acronyms of the BSC departments
 CASE: Computer Applications in Science and Engineering
 CS: Computer Sciences
 ES: Earth Sciences
 LS: Life Sciences



Programme

Second Day 28th May 2014

9.15h	Talk on The legal protection of computer programs	Invited Speaker: <i>Carles Comes,</i> ZBM Patents & Trademarks
11:15h	Poster Session and coffee	<i>Karthikeyan Palavedu, CS</i> <i>Alvaro Villalba, CS</i> <i>Roger Hernandez, CS</i> <i>Victor Valverde, ES</i> <i>Vincenzo Obiso, ES</i>
12:00h	A new empirical approach to energy functions for protein-protein interactions	<i>Iain Moal, LS</i>
12:30h	Evaluation of the on-line NMMB/BSC-CTM model gas-phase results on the European domain for 2010 in the framework of the AQMEII-Phase2 project	<i>Alba Badia, ES</i>
13:00h	Per-task Energy Measuring and Accounting in the Multicore Era	<i>Qixiao Liu, CS</i>
13.30h	Lunch Break	
14.30h	Development and optimization of high-performance computational tools for protein-protein docking	<i>Brian Jimenez, LS</i>
15:00h	Source-level compiler analyses for OmpSs	<i>Sara Royuela, CS</i>
15:30h	Coffee Break	
15:45h	A CASE study	<i>Beatriz Eguzkitza, CASE</i>
16:15h	Hybrid and Resilient Monte Carlo Methods for Matrix based Linear Algebra Problems	<i>Janko Strassburg, CS</i>
16:45h	Closing remarks	
17:00h	Adjourn	

Acronyms of the BSC departments:

CASE: Computer Application in Science and Engineering

CS: Computer Sciences

ES: Earth Sciences

LS: Life Sciences



Contributors & abstracts

INVITED SPEAKERS

Fernando Cucchietti
Barcelona Supercomputing Center



Fernando Cucchietti specializes in scientific visualization and high performance computer simulations.

He works at the Barcelona Supercomputing Center since 2011, where he founded the scientific visualization team that received the 2012 award for the best visualization video of the year from the National Science Foundation and the Science journal. He obtained his PhD in Argentina, followed by postdoctoral positions in quantum computing in Los Alamos (USA) and the Institute of Photonic Sciences (Spain). Before entering physics, Fernando worked on TV advertising and 3D animation, which imprinted his career with a passion for the visual communication of quantitative data.

Talk “10 things you should not do in a presentation”

Good oral communication skills are increasingly more and more important for a successful scientist's career. However, they are almost never taught formally, and students have little time to learn them on their own. As a result, technical presentations are typically dull and ineffective. In this talk you will learn the 10 most important techniques that could save your next presentation.



Contributors & abstracts

Carles Comes ZBM Patents & Trademarks

Carles has experience in both IT and patents. Since the beginning of his career in patents, back in 1994, his work has been focused on the fields of electronics, telecom and computer-related inventions, where he keeps up-to-date with the latest technical and legal developments. His work includes drafting applications for Europe and the US and dealing with substantive examination, as well as assessing patent validity and risk of infringement.



He often acts as technical patent expert before Spanish courts and takes part in customs intervention cases dealing with the infringement of essential patents in audio technologies. He has been a partner of ZBM Patents since 2005.

Carles also lectures on patents, especially on subjects related to the patentability of computer-related inventions. He is partly qualified and preparing for the rest of the European Qualifying Examination.

Talk “The legal protection of computer programs”

The aim of the talk is to discuss some of the possible forms for protecting a computer program. Understanding the patents as the most efficient form of protection, the practice of the European Patent Office, the Spanish Patent Office, the U.S. Patent Office and the Japanese Patent Office will be analysed. The talk will also address how to write a patent with both electronic inventions as well as computer-implemented inventions.



Contributors & abstracts

PRESENTERS

Alba Badia
BSC-CNS, Earth Sciences
Atmospheric Modelling

Talk “Evaluation of the on-line NMMB/BSC-CTM model gas-phase results on the European domain for 2010 in the framework of the AQMEII-Phase2 project”



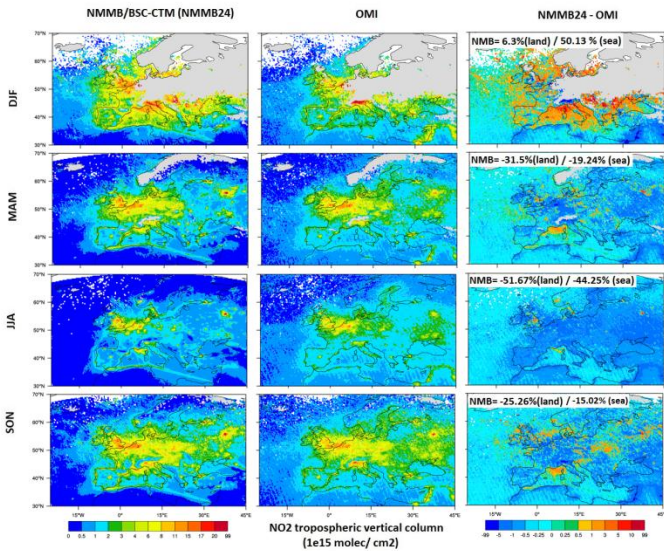
The Air Quality Model Evaluation International Initiative Phase2 aims to intercompare online coupled regional-scale models over North America and Europe. The NMMB/BSC Chemical Transport Model (NMMB/BSC-CTM) is a fully online integrated system for meso- to global-scale applications under development at the Barcelona Supercomputing Center. The NMMB/BSC-CTM is applied to Europe

for the year 2010 in the framework of the AQMEII-Phase2 intercomparison exercise. This contribution presents a spatial, temporal and vertical evaluation of the model results. This is the first time that the model has been evaluated on a regional scale over a whole annual cycle. The model is compared with available ground-based monitoring stations for relevant reactive gases, ozonesondes, and OMI and MOPITT satellite retrievals of NO₂ and CO. A comparative analysis of the present results and several European model evaluations is also presented here. The seasonal cycle for O₃, NO₂, SO₂ and CO is successfully reproduced by the model. The O₃ daily mean and daily maximum correlations for the analysed period are $r=0.68$ and $r=0.75$, respectively. The OMI tropospheric NO₂ column retrievals are



Contributors & abstracts

well reproduced, capturing the most polluted areas over Europe throughout the whole year (see figure). Modelled SO_2 and CO surface concentrations are generally underestimated, especially during the winter months. Two different vertical configurations of the model (24 and 48 vertical layers) are also analysed. Although model results are very similar, the simulation configured with 48 vertical layers provides better results regarding surface O_3 concentrations during summer. Compared to previous model evaluations, the NMMB/BSC-CTM's performance corresponds to state-of-the-art regional air quality models.



Comparison of modelled (NMMB24) NO_2 vertical tropospheric columns against satellite data (OMI) for (from top) winter (DJF), spring (MAM), summer (JJA) and autumn (SON) for the whole year 2010 in $1e^{15} \text{ molec}=\text{cm}^2$. NMMB24 data is displayed in the left panel, OMI data in the middle panel and NMMB24 minus OMI data in the right panel where the NMB (%) over land is added.



Contributors & abstracts

Israel Cabeza
BSC-CNS, Life Sciences
Electronic and Atomic Protein Modeling

Talk “PELE Applications”



Protein Energy Landscape Exploration (PELE) is a BSC in-house algorithm developed by Guallar’s Lab. In the Life Sciences Department. Originally, PELE was a Monte Carlo approach to study interactions between proteins and small molecules so called Ligands. PELE is an iterative procedure where each step is a combination of two main processes: Perturbation and relaxation. Perturbation consists of a random

translation and rotation applied to the ligand and a conformational change in the protein using Anisotropic Network Modelling (ANM). Relaxation has two parts, side chain prediction and global minimization. Side chain prediction techniques aim to find the best local minima by trying different rotamer configurations between side chain residues and ligand rotamers. Global minimization is used to relax the new configuration generated at the end of each step. The final structure produced in each step is accepted or rejected using a metropolis criteria.

Recently, a few changes in PELE have provided new capabilities. The addition of AMBER99SBBSC0 and a recent implicit solvent called OBC with the ACE non polar term has provided us the opportunity to explore DNA-ligand interactions. Nowadays, DNA-ligand interaction plays an important role in development of new drugs by pharmaceutical companies.

Another new PELE application is the study of protein intramolecular interactions. Experimental techniques, such as Atomic Force Microscopy (AFM), are able to measure the force generated during the forced mechanical unfolding of a protein. PELE has been modified to add a harmonic constraint to one



Contributors & abstracts

selected atom of the protein to a virtual point generated at the beginning in the same position. Moving the virtual point with a constant velocity (displacement per step) in a fixed direction generates a pulling force proportional to the atom displacement. PELE steps perturb and relax the protein generating a measurable resistance force proportional to the force measured in AFM experiments.

Beatriz Eguzkitza
BSC-CNS, CASE
Physical and Numerical Modelling

Talk “A CASE Study”

With my talk I would like to tell what means to do a doctoral thesis in CASE department. Through my personal experience and the results of my work I will try to show you the framework of our department, the profile of the students who carry out a PhD here and the great opportunities as well as the obstacles that one encounter when lives this experience.



The main result of my thesis is the HERMESH method which, a composition domain method. This term comes from the idea that HERMESH obtains a global solution of the problem from two independent meshes as a result of the union meshes. The global mesh maintains the same number of degrees of freedom as the sum of the meshes, which are coupled in the interfaces via new elements referred to as extension elements. The result of the global mesh is a non-conforming mesh in the interfaces between independent meshes due to these new connectivities formed with existing nodes and represented by the extension elements.

The first requirements were that the method be implemented in



Contributors & abstracts

an implicit strategy, be valid for any partial differential equation and not imply any additional effort or loss in efficiency in the parallel performance of the Alya code in which the method has been implemented. These properties constitute the main contribution in the coupling mesh for the computational mechanics framework

From these requirements, we have been able to develop an automatic and topology-independent tool to compose independent meshes. The method can couple overlapping meshes with minimal intervention on the user's part. The overlapping can be partial or complete in the sense of overset meshes. The meshes can be disjoint with or without a gap between them. And we have observed the flexibility of the method in the relative mesh size..

Lluís Gifré

BSC-CNS, Computer Sciences

Accelerators for High Performance Computing

Poster “Experimental Assessment of a High Performance Back-end PCE for Flexgrid Optical Network Re-optimization”

A specialized high performance Graphics Processing Unit (GPU)-based back-end Path Computation Element (PCE) to compute re-optimization in Flexgrid networks is presented. Experimental results show 6x speedups compared to single centralized PCE. This results in an increased probability of finding optimal solutions.





Contributors & abstracts

Santiago González
BSC-CNS, Life Sciences
Computational Genomics

Talk “Cancer genomics at BSC”



The first human genome was sequenced at the beginning of the 21st century in a common international effort that took more than 14 years. Nowadays the next generation sequencing machines (NGS) allow research centers to sequence hundreds of human genomes in a couple of months. The complexity of the genome and the total amount of data produced can not be understood

without a parallel evolution of computational programs to analyze them.

One of the leading fields in this genome analysis evolution is the study and characterization of the different types of cancers. International Cancer Genome Consortium (ICGC) is a common international effort that includes more than 70 different countries with the objective of uncover the genomic changes present in many forms of cancers and reduce their impact in the society.

As part of the ICGC and in collaboration with the sequencing centers, hospitals and research institutes, the Computational Genomics group at BSC is focused on the analysis of cancer data. One of our two main objectives is to remove the current limitations in the analysis of sequencing data enabling the analysis of complex diseases studies to the whole research community. The second is to contribute in the analysis of cancer DNA to move forward in our knowledge of the disease, to improve the diagnosis and to create new treatments.



Contributors & abstracts

Marc Guevara
BSC-CNS, Earth Sciences
Air Quality

Poster “Implementation of plume rise calculations and its impacts on emissions and air quality modelling over Spain”

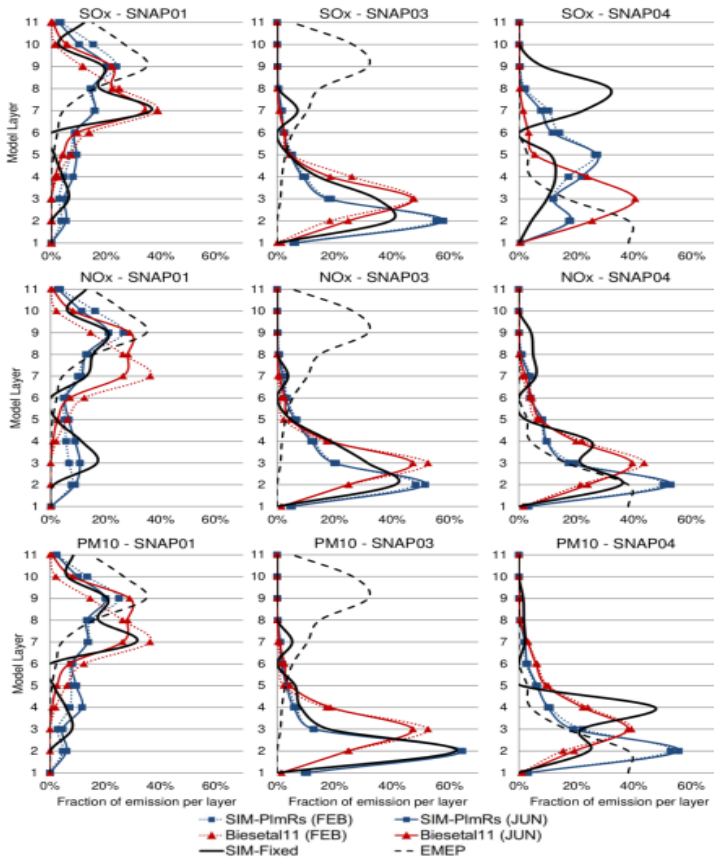


This work analyses the impact of implementing hourly plume rise calculations over Spain in terms of: i) vertical emission allocations, ii) modelled air quality concentrations. Two air quality simulations (4kmx4km, 1h) were performed for February and June 2009, using the CALIOPE-AQFS System (WRF-ARW/HERMESv2.0/CMAQ/BSC-DREAM8b) and differing only by the vertical allocation of point

source emissions: i) using fixed vertical profiles, ii) using an hourly bottom-up calculation of effective emission heights. Results shown that when using plume rise calculations, emissions are allocated to lower altitudes than when using the fixed vertical profiles, showing significant differences depending on source sector and air pollutant. In terms of air quality, it is shown that hourly plume rise calculations leads to improved simulation of industrial SO₂ concentrations, being the impacts on NO₂ and PM10 concentrations less significant. In order to maximize the performance of plume rise calculations, the use of stack parameters based on real-world data is mandatory.



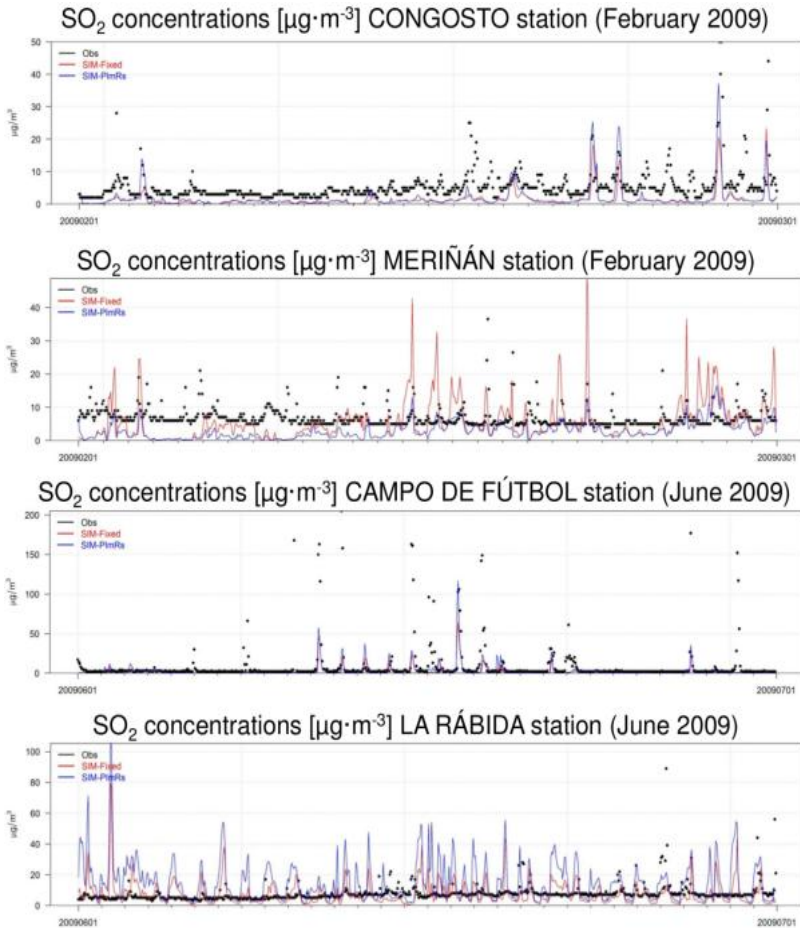
Contributors & abstracts



Comparison of SNAP-sectorial average SO_x, NO_x and PM₁₀ emission vertical profiles obtained with SIM-PlmRs, SIM-Fixed, EMEP and Bieser et al. (2011) (Biesetal11).



Contributors & abstracts



Observed and modelled time series of hourly concentrations for SO₂ [$\mu\text{g}\cdot\text{m}^{-3}$] at the stations of (a) Congosto and (b) Meriñán (February) and (c) Campo de Fútbol and (d) Rábida (June).



Contributors & abstracts

Roger Hernández

BSC-CNS, Computer Sciences

Autonomic Systems and e-Business Platforms

Poster “Experiences of using Cassandra for molecular dynamics simulations”

In response to the requirements of applications that work with large amounts of data (structure, scalability, etc.), various NoSQL databases have appeared to specifically deal with these challenges. This is a common practice in environments such as data analytics and OLTP, however these may not be the only data intensive applications that can benefit from these databases. In the life sciences domain, there are many applications that still use flat files as a medium to store data, and they see themselves very limited in terms of scalability and performance, in addition to code complexity. We present an analysis on the viability of using those databases for applications with large data demands that differ in some of the characteristics from what these systems were designed for. By using these databases, we can also observe that the design of the data model, queries and other configuration parameters can have considerable performance impacts, thus we present examples of different models and approaches to see how it affects performance. With the executions that are presented in this paper we are able to see performance gaps of a factor of up to almost 5 by using various models, queries and configuration parameters.



Contributors & abstracts

Tassadaq Hussain

BSC-CNS, Computer Sciences

Computer Architecture for Parallel Paradigms

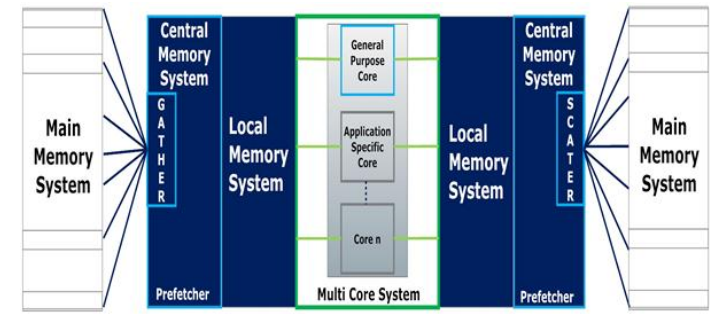
Talk “Supporting Scatter/Gather Tasks in Manycore architectures”

Heterogeneous multi-core architectures are increasingly popular due to their flexibility and high performance per watt capability. A kind of heterogeneous architecture, reconfigurable systems-on-chip, offers high performance per watt through the reconfigurable logic and flexibility via multiprocessor cores.

But in order to achieve the performance goals it is necessary to provide enough data to the accelerators.

In this work we describe a pattern-based memory controller (PMC) that aims at improving the performance of heterogeneous multi-core having general purpose and application specific cores.

These include scatter gather, irregular and strided 1D, 2D and 3D patterns. PMC can prefetch complete patterns into scratchpads that can then be accessed either by a microprocessor or by an accelerator. As a result, the microprocessors and accelerators can focus on computation and are relieved of having to perform address calculations.



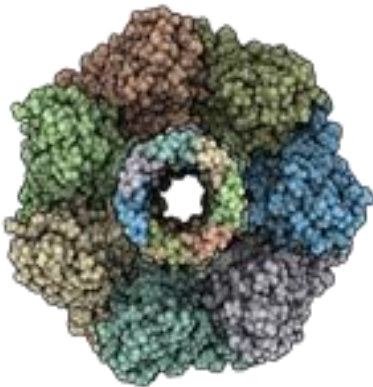


Contributors & abstracts

Brian Jiménez
BSC-CNS, Life Sciences
Protein Interactions and Docking

Talk “Development and optimization of high-performance computational tools for protein-protein docking”

Protein-protein interactions mediate most cellular functions, ranging from regulatory processes to immunologic reactions or signal transduction, and thus a detailed description of the association processes at molecular level is essential to comprehend the fundamental processes that sustain life. In such line, computational protein-protein docking tools aim to identify the native binding mode between two proteins, calculating the three-dimensional structure starting from their separate components. Predicting the structure of the complex is still a highly challenging task, but a valuable tool to complement experimental techniques which cannot provide structural information at a proteomics scale given their current technical limitations.



Here I present several optimizations of the PyDock rigid-body docking method developed within the Protein Interactions and Docking group as well as new methods of protein-protein docking using artificial intelligence algorithms.



Contributors & abstracts

Qixiao Liu

BSC-CNS, Computer Sciences

Operating System/Computer Architecture Interface

Talk “Per-task Energy Measuring and Accounting in the Multicore Era”

Energy is arguably the most expensive resource in a computing system. As multi-core processors continue to extend as the preferred hardware solution across different domains (e.g. datacenters, real-time systems or mobile devices), the need for accurately tracking (metering) energy usage in the system opens the door to relentless energy optimization techniques.

Quantifying the energy profile of one particular task running in a multicore

includes two fold: 1) per-task energy metering, consist of tracking the energy one task consumes when co-running with other tasks in a multicore system. 2) per-task energy accounting, consist of deriving the energy one task should consume with a given fraction of resources.

Both mechanisms have predominant applications in different domains open the new paths to enable energy-aware computing, power-secure system, energy/performance optimization and fair datacenter billings et al.





Contributors & abstracts

Iain Moal
BSC-CNS, Life Sciences
Protein Interactions and Docking

Talk “A new empirical approach to energy functions for protein-protein interactions”



Physical interactions between proteins are fundamental to almost all biological processes, including the attachment of viruses to their host, intra- and inter-cellular communication, and the assembly of macromolecular biomachinery. The ability to relate the atomic structure of an interaction to its binding strength, its free energy, is central to many studies, from determining the functional

consequences of pathological mutations and structural prediction, to the design of antibodies and peptide therapeutics. Here we present an approach to deriving intermolecular free energy potentials from experimental changes in binding energy upon mutation ($\Delta\Delta G$). First, the changes in intermolecular contacts between the mutant and wild-type structures are calculated. Then contact potentials are fitted to the $\Delta\Delta G$ data via least-squares regression, using the assumption that $\Delta\Delta G$ can be approximated by the sum change in contact energies. The derived potentials are validated by their ability to rank docking poses and predicting absolute binding free energies.



Contributors & abstracts

Vincenzo Obiso
BSC-CNS, Earth Sciences
Atmospheric Modelling

Poster “Modelling Direct radiative effect of mineral dust with the NMMB/BSC-CTM for dust outbreak events over The Mediterranean in Summer 2012”

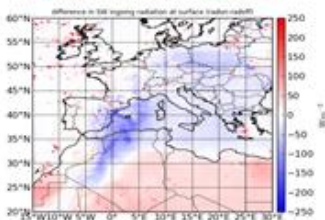
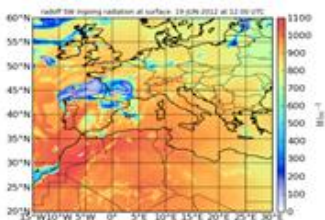
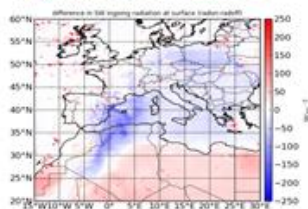
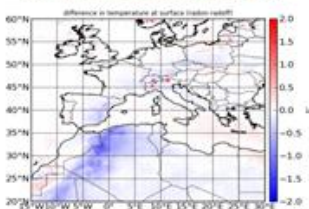
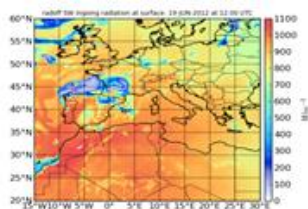
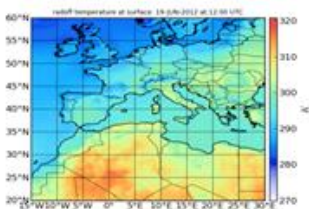
Aerosols interact with the atmospheric system scattering and absorbing solar radiation, with a significant impact on atmospheric energy and hydrologic processes. Radiative forcing associated with these perturbations affects climate and meteorology. In this contribution, we analyse model results of the Direct Radiative Effect (DRE) of mineral dust over the western-Mediterranean during summer 2012. For that, the NMMB/BSC Chemical Transport Model (NMMB/BSC-CTM) is applied on a regional domain at



0.1° horizontal resolution. The NMMB/BSC-CTM is a new on-line chemical weather prediction system coupling atmospheric and chemistry processes. In the radiation module of the model mineral dust is treated as a radiatively active substance interacting with both short- and long-wave radiation. The impact of the mineral dust outbreaks on meteorology is discussed by comparing model forecasts with atmospheric analysis and meteorological observations. The analysis focuses in the vertical structure of the atmosphere and the resulting surface meteorological conditions.



Contributors & abstracts



Maps of temperature at surface, ingoing SW-radiation at surface and outgoing SW-radiation at the TOA, with (RADON) and without (RADOFF) dynamic interaction between mineral dust and radiation, during a strong dust outbreak event on 19-JUN-2012.



Contributors & abstracts

Karthikeyan Palavedu
BSC-CNS, Computer Sciences
Heterogeneous Architectures

Poster “Energy Proportional Interconnects in HPC”



Energy efficiency has gained significant importance in HPC over the past decade and has become a key challenge in the design of Exa-scale systems. Advancements over the years in micro-architecture and process technology with focus on energy efficiency have been extremely successful at improving performance-per-watt while also sustaining tremendous performance scaling. In

this regard, with highly optimized processor and memory substrates, the power consumption of interconnect have become increasingly significant.

Interconnects contribute a significant portion of the system's energy consumption, of which, links represent about 65% of the interconnect energy. Links are of particular concern for energy efficiency as they are essentially ‘always-on’ regardless of usage. Recognizing the need, IEEE task force in Sept 2010 approved IEEE 802.3az Energy Efficient Ethernet (EEE) to exploit the low utilization patterns of Ethernet links to save energy. While the standard provides mechanisms to turn on or off links, when and how to do so are vendor decisions. Since EEE was primarily designed for Internet and data center workloads, in our research we investigate its potential for energy savings in HPC. In specific, to adopt such interconnects energy proportionality schemes in HPC, performance degradation due to the same must be controlled. In this poster, we present performance overhead aware energy savings schemes (PerfBound) targeted toward EEE for HPC systems. Our results show that with PerfBound, overhead can be bounded while still ensuring significant energy savings.



Contributors & abstracts

Milos Panic

BSC-CNS, Computer Sciences

Operating System / Computer Architecture Interface

Poster “A time-predictable many-core processor for critical real-time embedded systems”

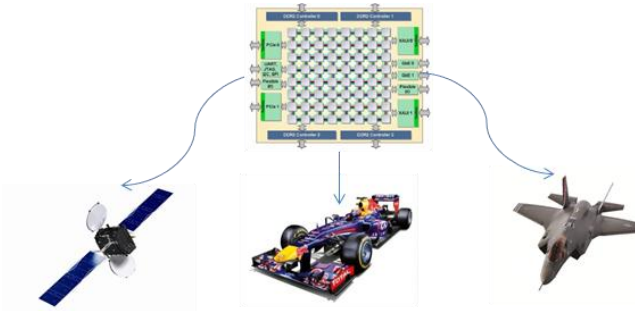
Critical Real-Time Embedded Systems (CRTES) in avionics and automotive domains use increasingly more sophisticated (complex) functionality, which requires higher and higher levels of computing power. Many-core processors are considered to cope with the performance and cost constraints imposed by future CRTES. On the one hand, many-cores allow scheduling mixed-criticality applications into the same processor, maximizing the hardware utilization while meeting size, weight and power constraints. On the other hand, many-cores improve application performance by exploiting task level parallelism.



Even though the use of many-core processors can satisfy the growing performance need of CRTES, it complicates timing analysis needed to prove that applications finish by their deadlines, as needed in CRTES (e.g. the airbag is instructed to open in less than 10ms). This is so because the timing behavior of the applications depends on the accesses to the shared hardware resources leading to inter-task interference. Inter-task interference occurs when some tasks share a hardware resource (cache, bus, memory, etc.) and try to access it at the same time requiring an arbitration mechanism to handle contention. The arbitration mechanism grants access to the shared resource to a task and stalls the others. This makes execution times and Worst Case Execution Time (WCET) estimates of tasks depend on the workload.



Contributors & abstracts



The goal of my work is devising many-core processors design guidelines to minimize the impact of inter-task interference on the timing behavior of applications as well mechanisms for concurrent execution of multiple parallel CRTES applications and novel WCET-aware thread to core mapping strategies that fully utilize the envisioned many-core.

Mercè Planas
BSC-CNS, Life Sciences
Computational Genomics

Poster “Metagenomics: Unveiling microbial communities across different environments”

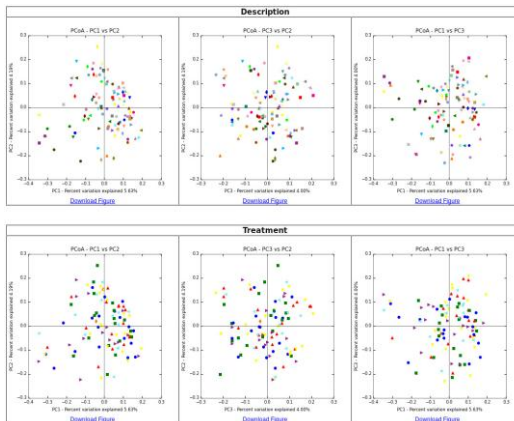
Metagenomics is a growing field in genomics that study the structures and functions of microbial communities where multiple microbial genomes are analyzed at the same time. The primary advantage is that metagenomics allows the analysis of genomes directly from environmental samples and does not require microbiological culture methods. Besides, it allows an environmental context study, a greater diversity and observes adaptation of community that is the process by which organism’s changes to become better suited to survive in their environment.



Contributors & abstracts

Metagenomics has innumerable applications and one of the most important is the study of diseases and human health. Our body contains a huge variety and number of microbes, and the study of these communities, called The Human Microbiome, enable us to deepen knowledge of complex diseases, such as cancer or obesity, as well as improving the prevention of diseases and the discover of new drugs.

Throughout metagenomics studies it has been shown that the loss or gain of specific genes activity have allowed the adaptation of microbiomes to particular environmental conditions. Nonetheless, is still not known whether and how the regulation of gene expression, the process by which an organism determines what genes will be express and when to becomes better suited to survive in their environment, also contributes to this environmental adaptation.



Principal Coordinate Analysis (PCoA) is commonly used to compare groups of samples based on phylogenetic or count-based distance metrics. We used it in order to observe if in case of treatment (Diabetes vs. Control) or in a particular month (Description) we can observe defined groups by their microbiome.



Contributors & abstracts

Our research interests are focused on the analysis of regulation of gene expression in metagenomic data so as to determine and explore how organisms are related to different environments. Furthermore, in collaboration with other groups, we are applying metagenomics to characterize microbial population correlated with different diseases, such as type 2 diabetes mellitus (T2DM), and tolerance in organ transplantation.

Nikola Rajovic
BSC-CNS, Computer Sciences
Heterogeneous Architectures

Poster “High Performance Computing with Mobile SoCs: Opportunities and Challenges”



In the domain of High Performance Computing (HPC) there is a constant need for higher computation performance. Over the time, there were different approaches in increasing the required level of performance due to new requirements, such as energy efficiency and economical market conditions.

This research study is a performance and feasibility study of Mobile SoCs

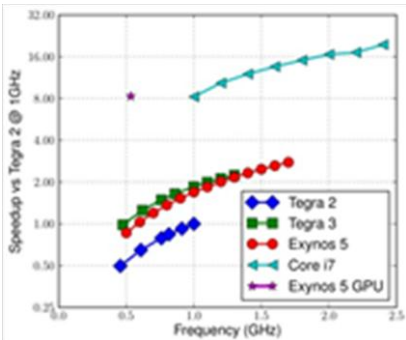
use in HPC environment – covering both homogeneous and heterogeneous computing scenarios with a mobile SoC being HPC node compute resource. To shed more light, we divide this study into four parts: mobile SoCs evaluation, HPC clusters deployment and evaluation, cluster modeling and simulation, and a future mobile SoC architecture proposal.

Our results show that each generation of mobile SoCs improves both performance and energy and reduces the gap between mobile SoCs and commodity processors used in HPC. In addition, we present mobile SoCs’ missing features for HPC. At

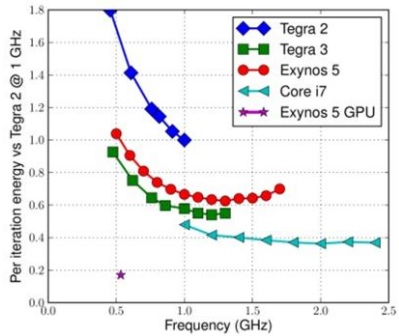


Contributors & abstracts

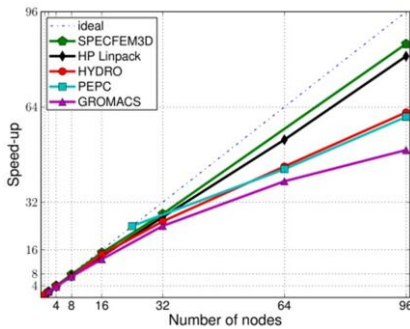
the end, we present scalability results of our first mobile SoC powered HPC clusters efforts, and cluster performance and power prediction methodology that we will use for design-space exploration of future mobile SoCs powered HPC clusters.



Performance comparison between three mobile multicore processors, commodity processor and a mobile GPU.



Energy comparison between three mobile multicore processors, commodity processor and a mobile GPU.



Application scalability on first Mobile SoC powered HPC cluster. (Tibidabo cluster)



Contributors & abstracts

Sara Royuela
BSC-CNS, Computer Sciences
Parallel Programming Models

Talk “Source-level compiler analyses for OmpSs”

Compiler analysis techniques have been developed for more than 40 years with the purpose of generating more efficient codes in three main aspects: execution time, memory usage and power consumption. Lately, compilers have been used to enhance programmability, offering automatic processes that lighten the use of programming models that otherwise can be in complicated or tedious.



During the analysis phase, the compiler gathers information that afterwards will be used to optimize the generated code. The more information the compiler has, the more aggressive it can optimize. Manpower is, however, an important factor so we have to identify the fewer analysis techniques that will allow more and most useful optimizations.

We have implemented a new service in the Mercurium source-to-source compiler with several analyses and optimizations. Two features are worth to highlight: a) the compiler shares a unique internal representation for C, C++ and FORTRAN and the analyses are built on top of it; b) Mercurium is a source-level compiler and therefore we have adapted low level analyses to deal with the abstraction of high-level programming languages (neither three addresses code nor SSA are ever generated).

We have extended some common control flow analyses (use-definition chains, liveness analysis, reaching definitions, simple strength reduction, constant folding and range analysis) to deal with parallel programming models such as OpenMP and OmpSs. With these analyses we are able to achieve good results in different fields:



Contributors & abstracts

I. Programmability:

1. Automatically detect the data-sharing attributes of variables involved in OpenMP/OmpSs constructs.
2. Automatically detect OpenMP/OmpSs tasks dependencies.

II. Correctness:

3. Detect errors or incoherencies in OpenMP/OmpSs constructs involving both scope of variables and dependencies between tasks.

III. Performance:

4. Assist user-directed vectorization based on directives.
5. In the future, implement a compile time mechanism to detect point-to-point dependencies among tasks.

Chiara Scaini BSC-CNS, CASE Environmental Simulations

Talk “Modelling volcanic ash dispersal and its impacts on civil aviation”



Our research aims at modeling volcanic ash dispersal in atmosphere and its impacts on civil aviation. This topic received a high attention after the 2010 Eyjafjallajökull eruption that provoked unprecedented impacts to all branches of society. Our objective is to support air traffic management during such events. We design different modeling strategies focused on both long-term (months to years) and short-term (hours to days) timescales. A

common long-term strategy is to combine the results of different simulations, obtained varying the model inputs, and produce



Contributors & abstracts

results representative of the expected conditions in a given study area. The methodology has been applied to many active volcanic areas and in particular to Iceland, given the high threat that its activity poses on European aviation. We thus produced the first hazard, vulnerability and impact assessment specific for volcanic ash dispersal in atmosphere. Short-term applications are usually associated to emergency management or air traffic, and at identifying the impacted elements (routes, airports and airspace sectors). In collaboration with aviation stakeholders, we produced an impact assessment tool that allows integrating and visualizing numerical modeling results and air traffic data in a Geographical Information System (GIS) framework. The tool has been applied to the Eyjafjallajökull eruption estimating its impacts on European air traffic. Further developments of these techniques are required in order to produce results at a higher temporal resolution (at quasi-real time). The application of HPC solutions could therefore support the definition of robust methodologies to support air traffic operations during volcanic eruptions.

Albert Soret
BSC-CNS, Earth Sciences
Air Quality

Talk “Air quality impacts of electric vehicles in Barcelona”

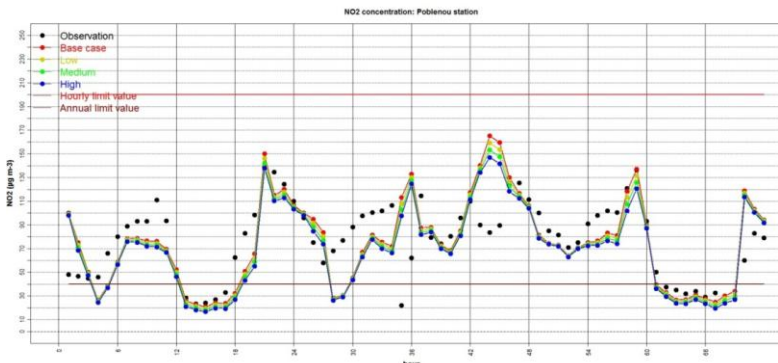
This work analyses the potential air quality improvements resulting from three fleet electrification scenarios (~13, 26 and 40%) by replacing conventional with electrified vehicles. This study has been performed in Barcelona, where road transport is the primary emission source. The WRF-ARW/HERMESv2/CMAQ model system has been applied at high spatial (1x1 km²) and temporal (1 h) resolution. The results show





Contributors & abstracts

that fleet electrification offers a potential for emission abatement, especially related to NO_x and CO . Regarding the more ambitious scenario (~40% fleet electrification), reductions of 11% of the total NO_x emissions are observed in Barcelona. These emissions reductions involve air quality improvements in NO_2 maximum hourly values up to 16%. Furthermore, an additional scenario has been defined considering electric generation emissions associated with EBVs and PHEVs charging from a combined-cycle power plant. These charging emissions would produce slight NO_2 increases in the downwind areas of $<3 \mu\text{g m}^{-3}$. Thus, fleet electrification would improve urban air quality even when considering emissions associated with charging electric vehicles. However, two further points should be considered. First, fleet electrification cannot be considered a unique solution, and other management strategies may be defined. This is especially important with respect to particulate matter emissions, which are not significantly reduced by fleet electrification ($<5\%$) due to the high weight of non-exhaust emissions. Second, a significant introduction of electric vehicles (26-40%) involving all vehicle categories is required to improve urban air quality.



NO_2 hourly air quality levels in the air quality station of Poblenou (Barcelona).



Contributors & abstracts

Srdjan Stipić

BSC-CNS, Computer Sciences

Computer Architecture for Parallel Paradigms

Talk “Techniques for Improving the Performance of Software Transactional Memory”

Transactional Memory (TM) provides software developers the opportunity to write concurrent programs more easily compared to any previous programming paradigms and promises to give a performance comparable to lock-based synchronizations.

Current Software TM (STM) implementations have performance overheads that can be reduced by introducing new abstractions in Transactional Memory programming model.

In this work we present four new techniques for improving the performance of Software TM: (i) Abstract Nested Transactions (ANT), (ii) TagTM, (iii) profile-guided transaction coalescing, and (iv) dynamic transaction coalescing.

ANT improves performance of transactional applications without breaking the semantics of the transactional paradigm, TagTM speeds up accesses to transactional meta-data, profile-guided transaction coalescing lowers transactional overheads at compile time, and dynamic transaction coalescing lowers transactional overheads at runtime.

Our analysis shows that Abstract Nested Transactions, TagTM, profile-guided transaction coalescing, and dynamic transaction coalescing improve the performance of the original programs that use Software Transactional Memory.



Contributors & abstracts

Janko Strassburg
BSC-CNS, Computer Sciences
Extreme Computing

Talk “Hybrid and Resilient Monte Carlo Methods for Matrix based Linear Algebra Problems”



Matrix inversion and solving systems of linear algebraic equations are problems found in a wide variety of scientific and engineering applications. Monte Carlo methods can be used to generate solutions quicker for suitable problems within this area.

Hybrid Markov Chain Monte Carlo methods for matrix inversion, sparse approximate inverse preconditioning and solving of systems of linear algebraic equations are considered.

Hybrid algorithms are built through the use of an iterative refinement scheme. Simulation of large-scale HPC environments has been used to analyse application behaviour and further optimisations. Considerations into resilience and fault tolerance for this class of algorithms are presented and an implementation detailed.



Contributors & abstracts

Víctor Valverde
BSC-CNS, Earth Sciences
Air Quality

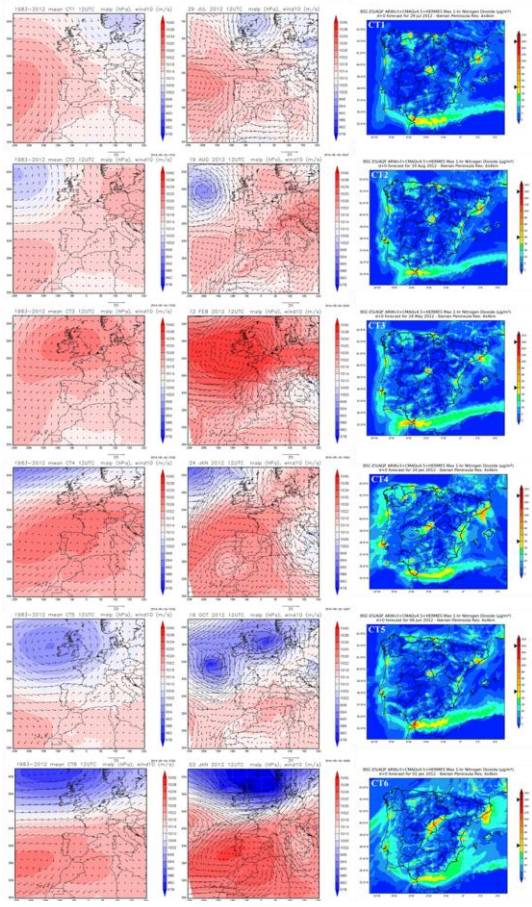
Poster “Characterization of atmospheric pollution dynamics in Spain by means of air quality modelling”

Air pollution is a matter of concern because it can adversely affect human health, Earth’s climate, ecosystems, visibility and materials. The concentration of pollutants in the air is directly associated to emissions and is strongly influenced by meteorology and topography. The aim of the PhD is to understand how typical meteorological patterns at synoptic scale explain air quality dynamics in Spain. An objective and automatic methodology has been developed in order to classify synoptic circulation on a climatic basis (1983-2012) using reliable reanalysis data from ECMWF. Sensitivity analyses to parameters affecting the resulting classification (statistical technique, number of patterns, meteorological variable used as proxy, vertical level, temporal and horizontal resolution, seasonality, and domain size) have been performed by means of the cost733 specific classification software in order to maximize its quality and to be applied on air quality characterization. A temporal stability analysis shows that 2012 synoptic classification is similar to the climatic one and can be therefore considered as a representative year. The relationship between synoptic patterns and surface concentration has been explored trough the analysis of 2012





Contributors & abstracts



Identified synoptic patterns for Spain for 1983-2012 (left) and representative days of 2012 (center). Mean sea-level pressure (hPa) and wind at 10 meters (m/s) at 12:00 UTC. NO₂ surface concentration forecast from CALIOPE at 12:00 UTC for representative days (right).



Contributors & abstracts

measurements at the Spanish air quality network. The CALIOPE high-resolution air quality forecasting system developed at BSC is currently used on representative episodes of the synoptic patterns in order to analyse urban and industrial pollution patterns for gases and aerosols.

The poster highlights the results obtained until now in relation to the synoptic classification and the analyses performed relating the identified synoptic patterns and air quality in Spain on representative episodes of 2012. The intended objectives of the on-going PhD are also presented.

Lluís Vilanova
BSC-CNS, Computer Sciences
Accelerators for HPC

Talk “CODOMs: Protecting Software with Code-centric Memory Domains”

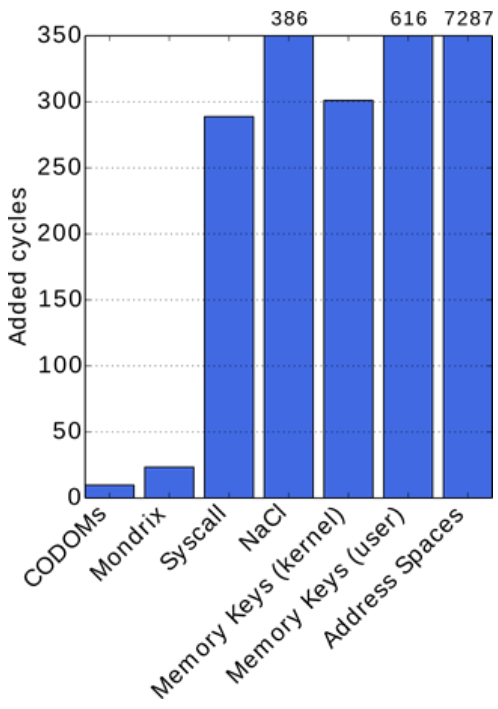
Today’s complex software systems are neither secure nor reliable. The rudimentary software protection primitives provided by current hardware forces systems to run many distrusting software components (e.g., procedures, libraries, plugins, modules) in the same protection domain, or otherwise suffer degraded performance from address space switches.

We present CODOMs (COde-centric memory DOMains), a novel architecture that can provide finer-grained isolation between software components with effectively zero run-time overhead, all at a fraction of the complexity of other approaches. An implementation of CODOMs in a cycle-accurate full-system x86 simulator demonstrates that with the right hardware support,



Contributors & abstracts

finer-grained protection and run-time performance can peacefully coexist.



Domain switch overhead comparison of CODOMs and other related systems. The overhead is depicted as additional cycles over a procedure call/return. Mondrix and Memory Keys are optimistically approximated by an instruction barrier. Our Mondrix approximation does not simulate the added costs of grants and revocations, which require costly operations akin to a TLB shutdown, which are not required in the CODOMs architecture.



Contributors & abstracts

Álvaro Villalba

BSC-CNS, Computer Sciences

Autonomic Systems and e-Business Platforms

Poster “The COMPOSE API for the Internet of Things”



The COMPOSE Project aims to provide an open Marketplace for the Internet of Things as well as the necessary platform to support it. A necessary component of COMPOSE is an API that allows things, COMPOSE users and the platform to communicate. The COMPOSE API allows for things to push data to the platform, the platform to initiate asynchronous actions on the things, and COMPOSE users to retrieve and process data from the thing.

In this poster we present the design and implementation of the COMPOSE API, as well as a description of the main key requirements that the API must satisfy. The API documentation and the source code for the platform are available at [<http://www.servioticy.com>].

Edited by Montserrat González F.
BSC Education & Training team
education@bsc.es
26th May, 2014
Barcelona (Spain)



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



EXCELENCIA
SEVERO
OCHOA