



# HiPEAC *info*<sup>28</sup>

COMPILATION ARCHITECTURE

appears quarterly | October 2011

Network of Excellence on High Performance and Embedded Architecture and Compilation

- 2 Message from the HiPEAC Coordinator
- 3 Message from the Project Officer
- 11 HiPEAC Announce
  - Book on Reconfigurable Computing: From FPGAs to Hardware/Software Codesign
- 4 HiPEAC Activities
  - 4 - Mini-Sabbatical - Ozcan Ozturk
  - 5 - Barcelona Computing Week: PUMPS 2011
  - 5 - Mini-Sabbatical - Ramon Beivide
  - 6 - University of Edinburgh and ARM Ltd Form New Centre of Excellence
  - 7 - Joint Seminar, Tampere University of Technology and RWTH Aachen University
  - 8 - ACACES 2011 Report
- 2 Community News
  - 2 - Tobias Grosser Obtains a Grant from Google European Doctoral Fellowship Programme in Europe and Israel
  - 10 - Radically Scalable CRISP General Stream Processor Solves Pitfalls of Many-Core
  - 11 - European Research Center on Computer Architecture (EuReCCA): a HiPEAC EuroLab
  - 12 - MELT Plugin 0.9 for GCC 4.6 Released
  - 13 - Kactus2 Applies IEEE1685/IP-XACT to SW and MCAPI abstraction
- 8 In the Spotlight
  - 8 - FP7 DSPACE Project
  - 14 - FP7 parMERASA Project
  - 14 - FP 7 S(o)OS Project
- 3 HiPEAC Member
  - Prof. Ola Dahl, Linköping University, Sweden
- 15 HiPEAC Startups
- 4 HiPEAC Students
- 17 PhD News
- 24 Upcoming Events



Welcome to  
the Computing  
System Week in  
Barcelona, Spain,  
November 2-4  
2011

[www.HiPEAC.net](http://www.HiPEAC.net)

The 7th International Conference on High Performance and Embedded Architectures and Compilers (HiPEAC 2012), 23-25 January 2012, Paris, France

# Message from the HiPEAC Coordinator

Koen De Bosschere



Dear friends,

I hope you have all enjoyed a relaxing summer break. This summer, the international news was dominated by the financial crisis in the western world. Several EU countries face serious financial problems, and even the first economy of the world was struggling with its debt ceiling. Experts warn us that this situation might last longer than everybody hopes, and some believe that this is a long term consequence of the 9/11 attacks, or that it is the beginning of the collapse of capitalism, two decades after the collapse of communism. Recently, the Chinese premier Wen Jiabao said that China is willing to help the western world because the Chinese economy cannot grow in an isolated way. This shows that the world is changing rapidly. The 21th century will definitely be different from the 20th century, just like the 20th century was dramatically different from the 19th century.

In July, many of us enjoyed the annual ACACES summer school in Fuggi, Italy. Like previous years, this community flagship event received a high appreciation score by the participants. I want to thank all the instructors for their excellent contributions. Next year's summer school will again take place in Fuggi, on July 8-14, 2012. It will be announced in January 2012.

In October we organize our fall Computing Systems Week in Barcelona, again co-located with the traditional Barcelona Multi-core Workshop. The cluster events will take place on 3-4 November. I want to thank our local organizers in Barcelona to organize this outstanding networking event for our community.

In January, there is the HiPEAC Conference, which will take place in Paris. For that conference, we are

experimenting with a new publication model in which we have basically outsourced the paper evaluation and publication process to ACM TACO. We received 37% more submissions than the previous year. The TACO editors immediately conditionally accepted 7 papers, and 46 papers got an invitation to submit a revised version. The evaluation of the revised versions will be ready by mid November. All authors of the accepted papers will get an invitation to attend the conference, and to present their work. Their paper will be carefully copyedited, published in ACM TACO, and indexed in all major publication databases. The conference itself it turned into a major networking event for the computing systems community in Europe. There will be more than 20 workshops and tutorials taking place during the conference, and all

HiPEAC companies and FP7 computing systems projects will be invited to promote their research and activities at the conference.

This autumn, we will also start the preparation of HiPEAC3, the successor network which will formally start on January 1, 2012. The network will basically continue the activities of HiPEAC2, but it will put different accents: we will focus more on industrial needs, more on the integration of new member states in our community, we will more intensively promote Europe as a computing systems research area, and we will try to create more press coverage. In parallel we will also work hard to further increase the European computing systems research funding.

Take care,  
Koen De Bosschere



## Tobias Grosser Obtains a Grant from Google European Doctoral Fellowship Programme in Europe and Israel

Google recently announced the recipients of the Google European Doctoral Fellowship Programme in Europe and Israel for 2011. These awards have been presented to exemplary PhD students in computer science and related research areas.

Tobias Grosser, who graduated from the University of Passau (Germany), has been selected as one of the 16 recipients. The fellowship will support his PhD thesis at INRIA and Université

Pierre et Marie Curie, under the supervision of Albert Cohen. Tobias Grosser will work on integrated loop transformations and vectorization, with special emphasis on just-in-time compilation.

The Google Doctoral Fellowship Programme acknowledges the recipients' contributions to their areas of specialty, providing three years of funding for their education and research. The award covers stipend/

salary, health care, social benefits, tuition and fees, conference attendance and travel, personal computer, and a Google Research Mentor.

[http://research.google.com/university/relations/doctoral\\_fellowships\\_europe.html](http://research.google.com/university/relations/doctoral_fellowships_europe.html)



Tobias Grosser, a recipient of the Google European Doctoral Fellowship Programme



# Message from the Project Officer

As a result of a recent coordinated call for proposals between the Information and Communication Technologies programme of the European Commission and the Ministry of Education and Science of the Russian Federation, two joint research projects have started. These two projects bring together High Performance Computing researchers from the European Union and the Russian Federation.

## **HOPSA: Holistic Performance System Analysis**

The objective of the project is to create an integrated diagnostic infrastructure for combining application and system tuning. Starting from system-wide basic performance screening of individual jobs, the automated workflow will route the findings on potential bottlenecks either to application

developers or system administrators on how to identify their root cause using more powerful diagnostic tools. Developers will choose from a variety of mature performance-analysis tools developed by the consortium. The tools will be further integrated and enhanced with respect to scalability, depth of analysis and support of asynchronous tasking, a node-level paradigm, which plays an increasingly important role in hybrid programs on emerging hierarchical and heterogeneous systems.

Project web site: [www.hopsa-project.eu](http://www.hopsa-project.eu)

## **APOS: Application Performance Optimisation and Scalability**

The objective of APOS is to develop optimised versions of scientific and industrial codes which are scalable

and portable across heterogeneous and homogeneous architectures. Scalability to thousands of cores is a principal goal of the work. In parallel with the development of the codes, prototype tools, starting from the current state of the art, will be interactively developed, deployed and refined. Representative real-world applications from the areas of seismic modelling, oil- and gas-reservoir simulation, computational fluid dynamics, fusion energy, and molecular dynamics will be used to test and evaluate methodologies for developing massively parallel, highly scalable applications and for testing relevant tools.

Project web site: [www.apos-project.eu](http://www.apos-project.eu)

Panos Tsarchopoulos



## New Member

Ola Dahl is an Associate Professor in System Integration, at the Electrical Engineering Department in Linköping University, Sweden. Before joining Linköping University in 2011, he worked at ST-Ericsson in Lund, Sweden, as a system engineer. He was in charge of the development of virtual platforms for simulation of the baseband processing hardware in next generation 3GPP multimode wireless modems. He also participated in LTE system design and standardization. In his earlier career he has been a system engineer and software engineer, working with embedded systems, control systems, and simulation. He has also been an Assistant Professor in Computer Science and in Control.

Ola got his PhD in Automatic Control

## **Prof. Ola Dahl, Linköping University, Sweden**

from Lund University in 1992. The topic of the thesis was path constrained robot control, and being an interesting mix of control theory, real-time software development, and system engineering, it was significant for the main interests also to be present in the following career. The interest for real-time systems has persisted, and is today reflected in Ola's teaching work, where he uses a small real-time operating system specifically designed to let students get insight in, and the possibility to modify and explore, the internals of a real-time system.

His current research interests are in the areas of system modeling and simulation for embedded systems, with a special consideration for the interplay between hardware and software.

He expects to participate in the HiPEAC community in areas such as concurrent and parallel programming, system engineering, simulation and modeling.



Contact information:

Ola Dahl  
[ola.dahl@liu.se](mailto:ola.dahl@liu.se)  
<http://www.daisy.liu.se/~olad>



## Mini-Sabbatical - Ozcan Ozturk

NEC Labs has a global perspective with researchers from many countries all over the world. As part of this diversity, our initial collaboration discussions started in mid 2010 with Dr. Srimat Chakradhar. These initial discussions resulted with a HiPEAC-funded mini-sabbatical to visit the NEC Labs, Princeton from January to April 2011. The visit was very fruitful to consolidate our joint research activities. I was able to participate in several ongoing research projects during my stay in Princeton.

The Computing Systems Architecture group is one of the eleven research groups at NEC Labs, focusing on topics with strong innovations and technical breakthrough. It was a very rewarding experience to be present in an environment with many talented researchers. Moreover, research discussions with leading institutes such as

Stanford University, Georgia Institute of Technology, and Princeton University were possible.

During my stay, we specifically focused on parallel programming models that allow programmers to specify computation and communication patterns, in a manner that is independent of the target architecture. The broad goals of the runtimes and middleware are to achieve performance scalability, hardware virtualization, and data placement effectively. In this domain, we worked on high performance computing type of applications and specifically we focused on linear solvers. We are working on a paper to be submitted to a forthcoming conference or a journal.

We also collaborated on using custom accelerators. Many emerging applications ranging from automobiles to data centers require intelligent processing

of large amounts of data in real-time. We started to investigate novel, special-purpose systems such as custom many-core architectures as domain-specific accelerators. For data-intensive applications the goal is to achieve high-performance at low power consumption. We started to implement new techniques and we are in the process of collecting the corresponding experimental results. We are planning to extend our collaboration to effectively use next generation many-core processors.

Overall, it was a very beneficial stay that provided great opportunities to share ideas with researchers. I would like to thank HiPEAC for providing support during my stay at NEC Labs. I also would like to thank Dr. Chakradhar and the Computing Systems Architecture members for their hospitality.

Ozcan Ozturk  
Bilkent University



## Collaboration Grant Report - Sofia Padiaditaki

I am a Ph.D student at the University of Edinburgh, UK. My supervisor is Dr. Mahesh K. Marina and I am part of the Wireless & Mobile networking (WiMo) research group. My Ph.D thesis focuses on adaptive spectrum management mechanisms for capacity improvement in wireless networks. Thanks to the HiPEAC collaboration grant I received, I had the opportunity to collaborate with Dr. Maziar Nekovee by visiting the British Telecom research site at Ipswich, UK. While being there, my work focused on the area of dynamic spectrum management for wireless home networks using TV-white-spaces.

The future wireless home will consist of a number of wireless devices, spanning from laptops and PDAs to media servers (e.g High Definition TV) and consumer electronics (e.g. game consoles). Moreover, due to the continuously decreased prices in wireless solu-

tions, other household applications, such as gas and electricity, will in the future be controlled wirelessly, allowing easy configuration and maintenance. Most of these devices, however, use popular wireless technologies, such as WiFi (IEEE 802.11) and Zigbee (IEEE 802.15.4), which operate in already congested frequency bands. As the number of wireless devices grows and the network population density increases, home networks operating solely in these bands are expected to suffer severe interference, thus constraining the attainable bandwidth.

Currently most developing countries are in the process of replacing the analogue television signal by digital transmissions. This process, called the Digital Switch Over (DSO), was completed in the US in June 2009 and is expected to be completed in the UK in 2012. After DSO is complete, a number

of channels within a geographic area are excluded from TV transmissions (TV-white-spaces). In this work, we studied mechanisms, which utilise the cognitive radio technology to dynamically manage the unused TV-spectrum and increase the available capacity of the future home networks.

The HiPEAC Collaboration Grant gave me the opportunity to work for a leading telecommunication company in an important research area. Moreover, I had the chance to meet many researchers in this field, and establish a working relationship that I hope will last for many years to come.

Sofia Padiaditaki,  
University of Edinburgh, UK



## Barcelona Computing Week: PUMPS 2011



The Barcelona Supercomputing Center and Universitat Politècnica de Catalunya hosted the second edition of the Barcelona Computing Week: PUMPS 2011 (Programming and tUsing Massively Parallel Systems).

The summer school, organized by BSC, UPC, University of Illinois and the HiPEAC Network of Excellence, with the collaboration of NVIDIA and Microsoft Research, is a five-day course for international students on GPU and HPC programming taught by world-renowned experts in the field Wen-mei Hwu (University of Illinois) and David Kirk (NVIDIA).

The summer school co-directors Mateo Valero (BSC-UPC) and Wen-

mei Hwu (University of Illinois), and local organizer and group manager of the BSC's "Accelerators for HPC" team Nacho Navarro (BSC-UPC), worked hard to offer an extraordinary learning experience in the beautiful city of Barcelona to the 95 students of 29 nationalities selected out of more than 250 applications.

This year's summer school offered two different tracks for beginners and advanced students, with personalized hands-on labs adjusted to each track under the guidance of outstanding teaching assistants.

The lectures covered topics such as CUDA programming, CUDA threading and memory model, CUDA algorithm patterns, multi-GPU programming,

debugging and profiling techniques, GMAC, CUDA with MPI, OmpSs and algorithmic optimization techniques among others.

On Wednesday a poster session was held where a selection of students presented the attendees had the opportunity to showcase and explain their current research work. A jury composed by the summer school's teachers and BSC's senior researchers evaluated their work, and decided to concede the Best Poster Award to Theodoros Athanasiadis for his work "Parallel Computation of Spherical Parameterization For Mesh Analysis". The prize for the best poster was a Tesla 2070 card and a diploma, while the next top five posters received an honorable mention and a Geforce GTX 480 card.

The success of this second Computing Systems Week: PUMPS 2011 establishes this event as one of the most important GPU programming courses in Europe and consolidates the academic effort of the BSC CUDA Research Center.

We hope to be able to bring it on a yearly basis for future international researchers interested in the topic. ■

## Mini-Sabbatical - Ramon Beivide

From the first of March until the end of July 2011, I have been working as a visiting scientist at the IBM Research Zurich Lab. I was invited by Dr. Minkenber, manager of the System Fabrics Group, part of the Department of Systems. I have been collaborating with him and other researchers in his group working in the area of new technologies for supercomputers and data centers, with special emphasis on interconnection networks. Dr. Minkenber is also a member of HiPEAC.

Before commenting on some technical matters, I would like to devote a few words to the environment that surrounds the IBM Research Lab at Zurich. It is a wonderful campus located at Rüslikon, a beautiful village overlooking Lake Zurich with panoramic views to the Alps. At this idyllic place, some few hundreds of researchers from more than 40 different countries devote their efforts to both fundamental and applied research. Although students and scientists come from every-

where, a tight collaboration exists with ETH, one of the most qualified technical campuses worldwide. The scientific atmosphere at this IBM Lab seemed to me superb; there you feel as if you were in an outstanding university campus but, according to my personal experience, with more positive control and management functions over the different research duties.

The goals of the research activities carried out during this sabbatical period

have been focused on the evaluation of cost and performance metrics of the interconnection networks used in high performance computing and data centers. As is known, interconnection networks are a key architectural resource in the design of computing systems of any scale. Although on-chip networks are also of paramount interest, the present project has been focused on system area interconnection networks. The research activities developed during this sabbatical period have already been successfully pursued in collaboration by IBM Research Zurich, Universidad de Cantabria and BSC from UPC.

Currently, two kinds of system area interconnection networks are used in high-performance computing: distributed or direct (mainly low-degree tori) and centralized or indirect (mainly high-degree folded-Clos or fat-trees). Among the ten more powerful current supercomputers, four use low-degree switches and six high-degree ones. All

four systems using low-degree networks rely on tori and five of the six using high-degree networks employ InfiniBand technology over fat-tree topologies. In addition, there are several research and industrial projects under development using both low- and high-degree routers. In respect to low-degree, 3D tori are evolving to higher dimensions and some existing and forthcoming machines employ 5D torus. In respect to high-degree, new approaches such as pruning the fat-tree or employing hierarchical networks are emerging as alternatives to the traditional Clos networks. Low-degree networks have their own niche as many parallel applications use near neighbor communication. Thus, they will still be present in the supercomputing market but they are not the only alternative. Some recent proposals advocate for future Exascale supercomputers based on high-degree hierarchical networks. Hence, during this period we have concentrated our efforts on both indirect and direct

topologies but always using high-degree routers.

Although five months is not a significant period, some of the obtained results have been sent to conferences in the field in order to be considered for publication. In addition, our collaboration after this stay has experienced a new impulse and currently a student from the University of Cantabria is working at IBM Research Zurich pursuing her PhD on these topics. Moreover, the research collaboration among BSC, University of Cantabria and IBM Research Zurich on high-performance computers continues with heightened expectations. For all of that, our gratitude to the HiPEAC Network of Excellence has to be explicitly declared as it is powering successful research collaboration between European academia and industry.

Ramon Bevide  
University of Cantabria, Spain ■

## HiPEAC Activity

# University of Edinburgh and ARM Ltd Form New Centre of Excellence

The University of Edinburgh's world-leading School of Informatics announced on Jun 1st, 2011, that an agreement has been signed with ARM Ltd, to form a new centre of excellence within the School's Institute for Computing Systems Architecture.

ARM is a leader in microprocessor design and licenses fast, low-cost, power-efficient processors and related technologies to its customers. The University of Edinburgh, through its School of Informatics has world-class expertise in the field of compiler and architecture technology.

The new centre will use the University's

world-class computer science and artificial intelligence expertise to investigate new ways of improving how computer programs utilise their existing processors.

"We are very excited about working with the world's largest processor IP provider" says Prof. Michael O'Boyle, Director of the Institute for Computing Systems Architecture.

"Designing and programming data-centre scale, heterogeneous, multi-core systems is the key challenge for the next decade. Future systems will provide unprecedented levels of potential performance but there is no

clear way for application software to harness this. At Edinburgh we are investigating novel technique including probabilistic parallelism discovery and statistical optimisation of software to bridge this gap. Working with ARM, to investigate new ways of delivering energy-efficient large scale parallelism is a unique opportunity for us with truly massive potential."

ARM, based in Cambridge, UK, designs the technology that lies at the heart of advanced digital products, from wireless, networking and consumer entertainment solutions to imaging, automotive, security and storage devices.

“ARM is delighted to be working closely with such a world-renowned centre of excellence in energy-efficient programming. Multi-core solutions can already be found in many of today’s smart devices. As consumers demand an increasingly con-

nected life, the complexity of the multi-core semiconductor technology will only increase. Research to maximise processing capacity and energy efficiency is an important focus area.” Peter Hutton, General Manager, Media Division. ■



Peter Hutton (ARM) and Michael O'Boyle (Uedin) at the opening of the ARM Research Centre of Excellence

## HiPEAC Activity

## Joint Seminar, Tampere University of Technology and RWTH Aachen University



RWTH Aachen – TUT seminar attendees

Thanks to HiPEAC’s support for networking, its members have the opportunity to travel to different countries to learn about research activities carried out there and to explore new cultures. One of such opportunities is a joint seminar that involves two or more HiPEAC members usually from different countries. RWTH Aachen University has been organizing such seminars in cooperation with the other members on a regular basis.

This year in June we visited Prof. Takala’s and Prof. Nurmi’s group at Tampere University of Technology (TUT), who kindly offered to host the seminar.

The seminar opened with a welcome speech by Prof. Takala and Prof. Nurmi. They introduced us to TUT and the department of Computer Systems, and showed the challenges they are currently tackling. Their presentation was followed by several excellent technical talks. The first one was about joint

validation of application and platform models from UML Sequence Diagram specifications after which there was a talk on the state of the art software defined radio platforms. Among the other topics presented were novel scalable hardware architectures by dimensioning coarse-grain reconfigurable arrays, with FFT as a case study; the TCE tool for ASIP design; and the FUNBASE project for HW/SW co-design and interoperability including the Kaktus2 tool for IP components integration based on IP-XACT standard. The last talk from that session described abstract simulation techniques based on SystemC.

After lunch Prof. Leupers and Prof. Ascheid opened the session with an overview of RWTH Aachen’s ICE institute. Talks from the afternoon session were devoted to the Nucleus methodology for efficient and portable software defined radios, sphere decoder ASIC implementation for efficient MIMO demapping, MAPS programming envi-

ronment for heterogeneous multi-core chips, language-driven design of reconfigurable architectures and advanced techniques for fast system simulations with SystemC.

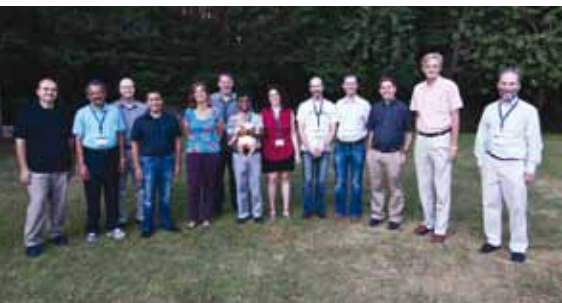
Both sessions were very lively with a lot of questions during and after each presentation. Many similar activities and interests were pointed out during the session breaks. In conclusion we were introduced to the excellent Finnish cuisine during the social event, where each of us got an opportunity to exchange more information about the research interests and culture, of course! The overall impression from the event was very positive for both sides and we believe that getting to know each other in person will certainly steer future cooperation between groups. A researcher exchange has already been agreed.

Anastasia Stulova  
RWTH Aachen University ■

# ACACES 2011 Report



Poster session



ACACES2011 teachers

This year, the HIPEAC summer school was held from the July 10 – July 16 in Fiuggi, Italy. I consider myself very fortunate to have received the opportunity to participate in it.

Fiuggi is about 60 kilometers east of Rome. The ACACES summer school 2011 commenced with the bus shuttle service, arranged by the organizing committee. The bus took us straight to our respective hotels in Fiuggi. The altitude of the town is about 700m above sea level. Fiuggi is very small and quite old, but I felt it was a perfect site to avoid digression from city life of Rome, and socialize, network and learn most from the lecture sessions. Fiuggi is famous for its water, which typically used to be bottled and exported to the rest of Europe in the past.

Registering for the courses was quite a challenging task as there were parallel sessions, and I found most of titles interesting. Finally, I selected two absolutely new topics, and two directly related to my own research and past experience, in order to focus both on the breadth and depth of the subject, while simultaneously not flooding my brain with too much new resources and information.

The accommodation of all the participants at the workshop were divided into three hotels (namely Silva Splendid, Ambasciatori and San Marco), and the facilities were quite satisfactory. Our handouts were already distributed at the respective hotel rooms. The first evening, we had our first keynote talk followed by dinner, where we all socialized and networked. From the second morning, lectures were held at the auditorium of the "Hotel Silva Splendid", and after every lecture, there was a typical coffee break or a meal break.

## In the Spotlight

### FP7 DSPACE Project: A New Digital Signal Processor for Space Applications



**Coordinator:** Annemaria Colonna, Sital Aerospace SRL  
**Website:** [www.dspace-project.eu](http://www.dspace-project.eu)  
**Partners:**  
 Sital Aerospace SRL  
 Intecs Informatica E Tecnologia Del Software S.P.A.  
 Consorzio Pisa Ricerche  
 Space Applications Services NV  
 RWTH Aachen University  
**Duration:** 2 years

The DSPACE Project, a new project supported by the EU FP7 Programme, has started with a kick-of meeting in Pisa, Italy in July 2011. The kick-of meeting served as a starting point for an initial get-together, discussions on organi-

zational details as well as first technical discussions. All project partners of the DSPACE project were present, including technical and managing staff from Sital Aerospace SRL, Intecs Informatica E Tecnologia Del Software S.P.A., Consorzio Pisa Ricerche Pisa, Space Applications Services NV and RWTH Aachen. This combination of SMEs, industrial partners, Universities and research organizations ensures that all the relevant stakeholders are present and different fields are taken into account for a successful completion of the project.

The new space missions, both scientific and commercial ones, require the capability to handle huge amounts of data and to process them on-board. These

always increasing calculation capabilities needs represent a critical issue for any spacecraft, emerging from complicated processing algorithms in the fields of earth observation and surveillance, SAR imaging, planetary observation and emerging manned space transportation system for human-interface displays. They all require a high, on-board calculation power in order to preprocess and process the acquired data before sending them to the Earth.

Available DSP based modules offer a typical computing power of 20 MIPS and noticeably 20 to 60 MFLOPS. The only European TSC21020 device currently suitable for these space applications is obsolete and US-made alter-



The halls and the seating facilities were quite extraordinary. As it was practically a crash course, the lectures were quite intensive and caffeine almost became indispensable. A couple of weeks after the summer school, all the respective course material was made available for download.

Along with the classes, and couple of keynote talks, there was a poster session. It was very impressive, and provided a better perspective of what different research groups at various universities and research facilities are currently working on. Also, we got to witness a lot of budding research

ideas, which are still at their nascent stages. Along with the research staff and students, there were people from the European Patent office, and discussions with them also provided very practical insights.

Every evening after the lectures, we typically enjoyed swimming in the sun, or in the spa. It was quite warm with a lot of sun, till late evening. A couple of evenings, I went biking to the old town on top of a hill. It was a great experience to see the old village and its narrow, cobbled streets.

The last evening, there was a party

organized in the open air, with a lavish dinner followed by live music and dancing till midnight. The next morning, the shuttle service took all the participants to either the Termini station or the Fumicino airport. That day, a spontaneous trip to the Vatican was organized, and I took this opportunity to visit the place and take pictures of the Sistine Chapel and the Vatican Museum.

In summary, ACACES 2011 was indeed a wonderful experience for me. This was my first summer school trip, and it's a pity that I did not participate in any of those while I was still a PhD student. I wish I can come back next year. I take this opportunity to thank the members of the organizing committee, the lecturers and all the fellow participants for making this event such a grand success.

Souradip Sarkar  
Intel Exascience Labs/ Ghent University,  
Belgium ■



DSPACE kick off meeting took place in Pisa, Italy

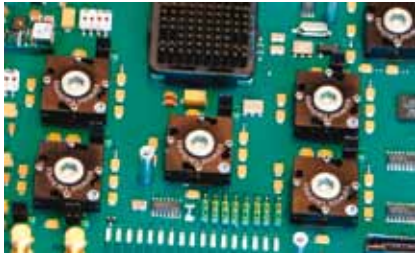
native products are subject to export restrictions (ITAR). Although it was considered sufficient a few years ago, future missions will require a much higher computational power. This requirement, together with the strong need to reduce the dependence on critical technologies from outside Europe,

makes it mandatory to develop the next generation of a European general-purpose high performance and radiation tolerant DSP. To make it easy to use and applicable to software developers of future space missions, it needs to come with an efficient and reliable Software application development environment.

Therefore, the DSPACE project aims to develop a new high-performance DSP core for space application up to 1GFLOPS that, meeting the scalability, multi-purpose and usability features. It is conceived to be used both as a stand-alone signal processor into embedded system and as a building block inside future scientific missions. The new DSP core will be designed taking into account both the requirements and limitations of European space technology. Moreover a complete software environment, building on top of an existing commercial or open-source environment, is expected to be developed together with benchmarks representative of existing and future space scenarios.

Maximilian Odendahl,  
RWTH Aachen University  
odendahl@ice.rwth-aachen.de ■

## Radically Scalable CRISP General Stream Processor Solves Pitfalls of Many-Core



CRISP PCB board with General Stream Processor chip

The CRISP consortium, consisting of Recore Systems (project leader), University of Twente, Atmel, Thales, Tampere University of Technology, and NXP Semiconductors, today presents the final demonstrators of the performant General Stream Processor chip. The term General Stream Processor truly applies to this chip, since it works equally well for simple streaming tasks such as position determination for Global Satellite Navigation (using 5 Recore-developed Xentium® DSP cores) as well as for complex tasks such as digital beamforming for radar processing (using over 120 cores simultaneously). Incidentally, the GSP chip is also a programmer's dream come true: the on-chip 'run-time resource manager' assigns tasks to resources during chip operation, and solves the multi-core programmer's nightmare of figuring out optimal use of resources on beforehand.

Programming many-core systems is known to be difficult. When adding more processing cores that can perform

DSP operations in parallel to make a faster system, cores that are waiting for each other may completely reduce the efficiency. The task of the multi-core programmer is to sensibly divide the main task in subtasks and optimally divide them over the available cores (resources). This is a complex problem to solve, even if the main task is fixed, and becomes almost impossible if the tasks change over time or if the number of resources varies over time. To make a programmer's life more pleasant, the consortium transferred the abstract optimization problem to an on-chip run-time resource manager. "Streaming applications usually involve real-time requirements. To guarantee these, application engineers traditionally need to make very conservative estimates on resource usage and do not tolerate changes to the platform or to neighboring applications" says Timon ter Braak, University of Twente. He continues "The right combination of reconfigurable hardware and run-time resource management can significantly improve the utilization of resources, in particular in the face of changes like for example a failing core". The resulting programmer-friendly, self-repairing CRISP chip is a big step towards facilitating the adoption of multi-core systems.

Another challenge of multi-cores is how many cores can actually be used at the same time without the system



On-chip runtime resource manager alleviates the programmer's nightmare of figuring out optimal use of resources

suffering from memory access bottlenecks, a prohibitive amount of cache coherency overhead, or other forms of 'on-chip traffic jams' when instructions and computation results travel back and forth from core to memory and core to core. The CRISP consortium implemented solutions based on techniques of distributed memory (memory tiles close to the DSP cores), distributed control, a scalable Network-on-Chip and the use of Xentium DSP cores - IP blocks optimized for use in low-power embedded streaming applications and with far less overhead than a 'standard' DSP core. "With the right architectural choices, reconfigurable many-cores can provide the predictability and flexibility needed to support run-time resource management." says Paul Heysters, Recore Systems. The CRISP consortium implemented the theoretical solutions for solving scalability issues using 3 PCB boards with General Stream Processor chips, resulting in a 138-core system. The run-time resource manager mapped a complex radar application on the system, to process 48 receive channels and compute 24 output beams on 117 cores simultaneously. Paul Heysters of project leader Recore Systems: "We could daisy-chain even more PCB boards to make a 1,000-core, and end up with an expensive gadget just to prove that we can do it. We're waiting for a sensible real-life application which will challenge us to go beyond the ground-breaking results we've already achieved in creating a universal multi-core General Stream Processor." ■

### About CRISP:

The CRISP (Cutting edge Reconfigurable ICs for Stream Processing) project researches optimal utilization, efficient programming and dependability of reconfigurable many-cores for streaming applications. Its goal is to develop a single, highly scalable, reconfigurable system concept that can be used for a wide range of streaming applications; from low-cost consumer applications to very demanding specialty applications.

Project website: [www.crisp-project.eu](http://www.crisp-project.eu)

# Book on Reconfigurable Computing: From FPGAs to Hardware/Software Codesign

Editors: **João M. P. Cardoso, and Michael Huebner**

Reconfigurable Computing has evolved to a mature level, with a wide range of applications, from high-performance to embedded computing domains. The research community has witnessed exciting advancements, promoted by technology and scientific findings. European industry and academia have been very dynamic on researching Reconfigurable Computing architectures, tools, and applications. In the last five years, the EU (European Union) has funded various projects through FP6 and FP7 frameworks. Projects such as MORPHEUS, hArtes, 4s, AETHER, ANDRES, REFLECT, CRISP, ERA, include or have included many of the academic groups working on reconfigurable computing and relevant companies interested in reconfigurable computing in Europe.

This book mainly consists of contributions from those projects. We believe this book is representative of the R&D about reconfigurable computing performed in Europe and partially funded by the EU. It can be a guideline to pre-

pare novel project proposals and can be used in courses about reconfigurable computing, in M.Sc. and Ph.D. programs, and as a library reference. It is a source of different perspectives, results, goals, etc. With this book, M.Sc. and Ph.D. students can learn about challenges addressed in recent EU projects. This book is organized in eleven chapters. Chapter 2, "The Relevance of Reconfigurable Computing", and Chapter 3, "HiPEAC – Upcoming Challenges in Reconfigurable Computing", present the relevance of reconfigurable computing and its main challenges, especially the most relevant ones as seen by members of the Reconfigurable Computing Cluster of HiPEAC NoE (High Performance and Embedded Architectures and Compilation Network of Excellence). The subsequent chapters present eight projects related to reconfigurable computing and partially funded by the EU: - six projects that have already finished (Chapters 4 to 9), "MORPHEUS – Exploitation of reconfiguration for



increased run-time flexibility and self-adaptive capabilities in future SoCs", "hArtes (Holistic Approach to Reconfigurable Real Time Embedded Systems)", "Smart Chips for Smart Surroundings - 4S", "Self-Adaptive Networked Entities: Autonomous computing elements for future pervasive applications and technologies – AETHER", "ANDRES – Analysis and Design of run-time Reconfigurable, heterogeneous Systems", "CRISP – Cutting edge Reconfigurable ICs for Stream Processing", and - two projects that started in the beginning of 2010 (Chapters 10 and 11), "ERA – Embedded Reconfigurable Architectures", and "REFLECT: Rendering FPGAs to Multi-Core Embedded Computing". We believe that these projects will help to shape the face of the reconfigurable computing field and we hope that the readers find this book an interesting resource to learn about them. ■

## European Research Center on Computer Architecture (EuReCCA): a HiPEAC EuroLab

One of the steps towards tighter integration in our community and sustainability of HiPEAC is the notion of EuroLabs: physically distributed but virtually integrated research laboratories, working together in specific areas.

EuReCCA, the European Research Center on Computer Architecture, is the first such HiPEAC EuroLab, being currently organized by Chalmers University of Technology (Sweden), Barcelona Supercomputing Center and UPC (Spain), and FORTH-ICS and Univ. of Crete (Greece), together with

other European Architecture academic and research Labs (INRIA-IRISA-ALF, U.Augsburg-SIK, U.Manchester-APT). EuReCCA is organized in close collaboration with many European companies. It is a pan-European virtual center of top-class research institutions in computer systems architecture, with complementary skills, cutting across computer systems layers.

Computer systems will increasingly face challenges concerning programmability, performance, energy efficiency, and reliability, driven by demands of

emerging ICT services and applications. EuReCCA, with its associated world-class research institutions, is committed to tackle these challenges, together with industry leaders in computing systems, by cutting-edge research, education, and innovation in computer systems architecture.

The three cornerstones that form the agenda for EuReCCA are:

**Research**, to tackle the most challenging long-term problems, at the same time as implementing working

prototypes. We aim at consolidating our research agendas, using our complementary skills, sharing knowledge, tools, and infrastructure, and being always inspired by the Architecture part of the HiPEAC Roadmap.

**Education**, to equip our future innovators with the right skill set for European industry leadership in computing systems. We aim for collaborative studies at multiple institutions, including course work, team projects, research, prototyping, industrial internships, and entrepreneurship, leading to Joint M.Sc. and Ph.D. degrees.

**Innovation Incubation**, to increase job opportunities for our graduates



through industry and entrepreneurship. We aim to create an ecosystem for accelerating innovation and commercial exploitation, by understanding industry's needs, concerns, and methods, by performing industrially-relevant research and education, and by encouraging the industrial exploitation of research results.

EuReCCA is needed because the complexity of computer design increases dramatically, and solutions to problems require holistic approaches. We aim to consolidate expertise in one virtual center, to exploit complementa-

rities and synergies of scale, to attract to Europe and keep here the best researchers, and thus nurture industrial development. EuReCCA's ambitions are to become a top-ranking world center in computer architecture, create the proper critical mass, provide the foundations for the next generation computing industry, raise public awareness of computing and architecture, and significantly improve the technology transfer climate and the number of successful startup companies in Europe.

Per Stenstrom,  
Mateo Valero,  
Manolis Katevenis  
<http://www.eurecca.org/>



## MELT Plugin 0.9 for GCC 4.6 Released

It is well known that GCC, the GNU Compiler Collection, is the leading free compiler (GPLv3+ licensed) for many languages (C, C++, Ada, Fortran, Go, Objective C, etc), processors (e.g. x86, ARM, Sparc, S390, PowerPC, etc) and systems (GNU/Linux, Solaris, Android, AIX, Windows, etc). Its current version is 4.6.1 (June 2011). It is extensible through plugins and provides many optimizations and features.

We are happy to announce the release 0.9 of the MELT plugin for GCC 4.6. MELT is a high-level domain specific language to ease development of GCC extensions. See <http://gcc-melt.org/> for details and download. MELT 0.9 is available as a free (GPLv3+ licensed) plugin for gcc-4.6. It has the following features:

- Extensions coded in MELT can add some new passes inside GCC which access or change internal GCC representations (notably in its middle-end).
- The MELT language handles both MELT values (dynamically typed), including closures, boxed GCC data, lists, tuples, objects, hash-maps, and

raw existing (statically typed) GCC stuff, in particular GCC Gimple, Generic/Tree, Control Flow Graph, internal representations.

- The MELT language enables high-level programming styles, notably functional or applicative programming (with anonymous lambda functions), object orientation (with single-inheritance and classes having dynamic method dictionaries SmallTalk or Ruby), reflection (run-time introspection), and powerful pattern-matching (notably on existing GCC internal representations).

- The MELT runtime fits well in GCC implementation; in particular, MELT garbage collector is an efficient, copying generational collector integrated into GCC.

- MELT is translated to C code (which is then compiled and dynamically loaded into gcc). The MELT translator is implemented in MELT (and its C translation is available).

- MELT code can include C code chunks, and the MELT language provides several devices to describe how

C code is generated (from the user's code in MELT).

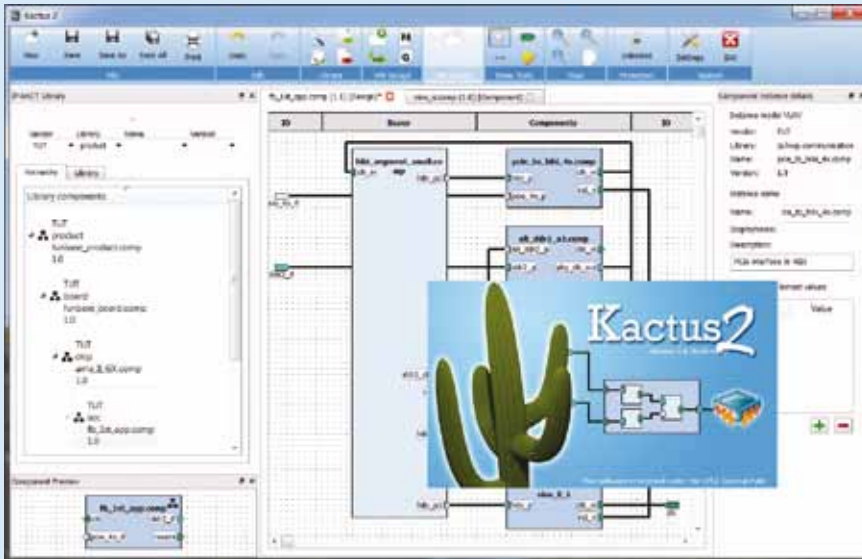
- MELT is a Lisp-like language, with syntax familiar to Lisp or Scheme.

MELT is developed mostly by Basile Starynkevitch at CEA LIST (Saclay, France) -funded through GGCC [ITEA] and OpenGPU [FUI] projects by DGCIS - with several external contributions (by Pierre Vittet, Jérémie Salvucci, Romain Geissler). The Talpo simple static analyzer, giving warnings on coding rules, has been developed in 4 months in MELT by Pierre Vittet, as part of a GSOC 2011 project (without any prior knowledge of GCC, MELT or compilation techniques). MELT is available as plugin for an unchanged gcc-4.6 compiler. See <http://gcc-melt.org/> for more (source code, mailing list, papers and slides, documentation, use cases). A tutorial on GCC extensions with MELT will be given at HiPEAC conference in Paris (January 2012).

Contact:  
Basile Starynkevitch  
[basile.starynkevitch@cea.fr](mailto:basile.starynkevitch@cea.fr)



## Kactus2 Applies IEEE1685/IP-XACT to SW and MCAPI Abstraction



Screenshots of Kactus2 development environment

Kactus2 is a new open source design environment for metadata-based design. It was developed at TUT with nine companies offering imaging and industrial embedded products. Common to all are very high speed IO and demanding real-time signal processing, with FPGAs for interfaces and CPU/DSPs for computation. The challenge is how to keep up products with rapid platform evolution and customer specific customizations also for small series. Immediate goals were “FPGA made easy to SW engineers” and systematic management of complete product with its HW, SW and configurations over the life cycle.

We selected a metadata-based design as the cornerstone, to solve the product management and practical component integration. Major benefit is reduction of errors in IP integration and exchange between partners. We use the IEEE1685/IP-XACT metadata standard, but apply its use in three ways.

First, we extend IP-XACT usage from IP-block and SoC levels to chip, board

and product levels. In practice users may draw PCB block diagram blueprints and store them as IP-XACT objects and designs instead of typical office tool drawings. This way all product hierarchies are kept in uniform, formal descriptions and associated design information can be found easily.

Our second extension is description of SW components and HW/SW mapping with VLNV references. That makes both HW and SW as independent metadata objects.

The third extension is abstraction of HW and SW for heterogeneous MP-SoC. We use Multicore Association communications API (MCAPI), not only for SW-SW interface but also for fixed HW blocks.

Kactus2 implements the extensions, but completely following the IP-XACT XML format. We cautiously introduce new vendor extensions to make tool development easier and to improve interoperability. A specific feature is absence of any Kactus2-specific

project files. All information is stored in IP-XACT XML metadata, e.g. the locations of the component symbols on screen. Setting up the design tool environment on different sites and exchanging IPs between partners is straightforward.

Kactus2 can be used for packetizing existing HW and SW components as well as integrating them in designs. Built-in generators are available for Modelsim, top VHDL generation and Altera Quartus project file generation. Standard SOAP/TGI interface is coming for external generators.

A handy feature is drafting: just draw empty components, interfaces and connections in any design and store new items as IP-XACT objects. For HW components, VHDL entity can be generated for real IP implementation. Fixed HW IP-blocks can be packetized with virtual MCAPI endpoints in IP-XACT library. User can define real endpoints for all processors in the product. The system design view shows flattened the components that may have endpoints, and the user can design channels between endpoints. Adoption of IP-XACT metadata-based design can be too difficult especially for SMEs even though it gives clear benefits. Our ambition with Kactus2 is to make the barrier lower by clear, easy to understand user interface and by focusing on essential design tasks. Being it open source we welcome academic and industrial contributions. Kactus2 and exemplar IP-library is released in Oct. 4th, 2011 and can be downloaded from <http://funbase.cs.tut.fi>

Prof. Timo D. Hämmäläinen  
Tampere University of Technology,  
Finland

## FP7 parMERASA Project : Multi-Core Execution of Parallelised Hard Real-Time Applications Supporting Analysability

Providing higher performance than state-of-the-art embedded processors can deliver today will increase safety, comfort, number and quality of services, while also lowering emissions as well as fuel demands for automotive, avionic and automation applications. Such a demand for increased computational performance is widespread among European key industries. Engineers who design hard real-time embedded systems in such embedded domains express a need for several times the performance available today while keeping safety as major criterion. A breakthrough in performance is expected by parallelising hard real-time applications and running them on an embedded multi-core processor, which enables combining the requirements for high-performance with time-predictable execution. The EC FP7 parMERASA project, starting Oct. 1, 2011, targets a timing analysable system of parallel hard real-time applications running on a scalable multi-core processor. Several new scientific and technical challenges will be tackled in the light of timing analysability: parallelisation techniques for industrial applications, timing analysable parallel design patterns, operating system virtualisation and efficient synchronisation mechanisms, guarantee of worst-case execution times (WCET) of parallelised applications, verifica-

tion and profiling tools, and scalable memory hierarchies together with I/O systems for multi-core processors. The objective of parMERASA is an at least an eightfold performance improvement of the WCET for parallelised legacy applications in avionics, automotive, and construction machinery domains in comparison to the original sequential versions. The execution platform, i.e. the parMERASA multi-core processor and system software, will provide temporal and spatial isolation between tasks and scale up to 64 cores. A software engineering approach will be developed to ease sequential to parallel program transformation by developing and supporting suitable parallel design patterns that are analysable. Verification and profiling tools will be developed, and we aim to provide recommendations to enhance both automotive and avionics standards. Most of the parMERASA partners met in the HiPEAC NoE and already collaborated on the EC project MERASA, 2007-2011, namely the partners University of Augsburg (project coordinator, parallelisation, system software, and processor architecture), Barcelona Supercomputing Center (processor architecture, simulator, and avionics software support), University Paul Sabatier of Toulouse (static WCET analysis), Rapita Systems

Ltd. (verification and profiling tools), and Honeywell International of Czech Republic as avionics application company. The additional parMERASA partners strengthen the application areas by DENSO Automotive Deutschland GmbH for automotive control units and by BAUER Maschinen GmbH for construction machinery; University of Dortmund will contribute to the processor architecture and system software development. parMERASA will impact new products for transportation systems and industrial applications. It will impact standards by introducing parallel execution and time predictability as key features. This will contribute to reinforce the EC position in the field of critical computing systems and yield an advantage for European industry in the highly competitive embedded systems markets.



Project website:  
<http://www.parmerasa.eu>  
 Project Coordinator:  
 Theo Ungerer,  
 University of Augsburg,  
[ungerer@informatik.uni-augsburg.de](mailto:ungerer@informatik.uni-augsburg.de) ■

## FP7 S(o)OS Project: A Status Update

The S(o)OS European Project (Service-oriented Operating Systems) aims to design a novel software stack that is suitable for future and emerging massively parallel, distributed and heterogeneous systems. The attention is strongly focused on the architecture of the Operating System and its kernel, and the interactions with the applications through suitable novel

programming paradigms. These will allow the majority of the programmers to take advantage of the computing power available on new generation hardware.

The project has just surpassed the first half of its overall duration, closing the first phase of the planned research activities. Among other documents

and deliverables produced by the Consortium, one of the key public documents recently produced is the Project Deliverable D5.3 "First Set of OS Architecture Models". This document reports the results of the preliminary investigations about general Operating System (OS) architecture models that will make it easier for developers to code applications on



**Partners:**

High-Performance Computing Center of Stuttgart (HLRS - DE), coordinator

Scuola Superiore Sant'Anna (SSSA - IT)

Ecole Polytechnique Fédérale de Lausanne (EPFL - CH)

Instituto de Telecomunicações (ITAv - PT)

University of Twente (NL)

**Website:**

<http://www.soos-project.eu/>.

**Duration:**

3 years

massively parallel and distributed systems as expected to be available in 10-15 years in the future. The discussion focuses first on a small set of

target application scenarios which are useful to highlight particularly critical requirements posed by the applications on the OS, as arising in the context of S(o)OS. These requirements are mainly related to scalability issues of nowadays Oses and run-time environments. Then, the OS architecture model is sketched out in terms of subcomponents, their interconnections and interdependencies and behaviour. The main focus of the document is the one to identify critical architectural elements and sub-components that, constituting major bottlenecks in nowadays Operating Systems, need to be reviewed and rethought for the purpose of being able to face with future massively parallel and distributed systems. This will allow for exposing the available computing power to a broad range of developers, ranging from average

developers to highly experienced ones.

The D5.3 public deliverable is available for download from the S(o)OS website at this URL:

[http://www.soos-project.eu/index.php?option=com\\_content&view=article&id=54&Itemid=60](http://www.soos-project.eu/index.php?option=com_content&view=article&id=54&Itemid=60)

where also the other public deliverables and technical reports can be downloaded. Should you be willing to follow more closely the project activities, and/or engage in interesting discussions on related topics, feel free to subscribe to the open LinkedIn group:

<http://www.linkedin.com/groups/S-o-OS-Serviceoriented-Operating-3591051?gid=3591051> ■

## HiPEAC Startups

## NovoCore Ltd



Two HiPEAC members from Imperial College London have formed a spin-out company, NovoCore Ltd, in collaboration with Imperial Innovations, one of the UK's leading technology transfer and spin-out incubation companies. NovoCore builds on the work of Dr Constantinides and Dr Bouganis in the Circuits and Systems Group at Imperial, a world-leading research group in FPGA based systems, acceleration, and embedded low power, high performance computing.

Our company offers licensing of FPGA IP cores for numerical computing, real-time, low latency and low jitter computing, embedded numerical

optimization, computer vision, and machine learning, as well as customized high-performance memory interfaces. We also undertake consultancy, and we are registered with the European Commission as an SME.

We would be very happy to hear from other HiPEAC members about possible research and development collaborations within our field of activity.

Website: <http://www.novocore.com>

**Dr George A. Constantinides** ([george@novocore.com](mailto:george@novocore.com)) is CEO of NovoCore Ltd, as well as Head of the Circuits and Systems research group and Director of the Centre for Digital Numerical Computation at Imperial College. Dr Constantinides is a Fellow of the BCS and a Senior Member of the



IEEE. He is an Associate Editor of the IEEE Transactions on Computers and the Journal of VLSI Signal Processing. He is CANDE Track Co-Chair of ISCAS in 2012, was Program Co-chair of the IEEE International Conference on Field-Programmable Technology in 2006 and Field Programmable Logic and Applications in 2003, and is a member of the steering committee of the International Symposium on Applied Reconfigurable Computing. He serves on the technical program committees of several conferences, including DAC, FPGA, FPT and FPL.

**Dr Christos Bouganis** ([christos@novocore.com](mailto:christos@novocore.com)) is CTO of NovoCore Ltd, as well as a member of academic staff and organizer of the MSc in Analogue and Digital Integrated Circuit Design at Imperial College. He leads



active research into architectures for signal processing, computer vision and machine learning algorithms. He is an Associate Editor of IET Computers and Digital Techniques, and Journal

of Systems Architecture. In 2007, he served as the Program Chair of the IET FPGA designers' forum, and in 2008 he was the General Chair of the Fourth International Workshop on

Applied Reconfigurable Computing. He currently serves on the Technical Programme Committees of several international conferences, including FPL, FPT, DATE, and SPPRA. ■

## HiPEAC Startups

Bmob Sagl is a start-up company founded in 2010 and located in Lugano (Switzerland). The company offers a city-wide parking monitoring system with the goal of helping authorities in improving parking management and of helping drivers finding free parking slots quickly.

Bmob is hosted by CPStartUp ([www.cpstartup.ch](http://www.cpstartup.ch)), the local start-up promotion center. Main partners of CP-StartUp are the University of Lugano (USI, [www.usi.ch](http://www.usi.ch)) and the University of Applied Science and Art of Southern Switzerland (SUPSI, [www.supsi.ch](http://www.supsi.ch)).



Parkomotivo: the parking monitoring wireless sensor network deployed in Lugano.

### The Bmob Parking Monitoring System

The Bmob parking monitoring system combines different vehicle detection methods with data mining and with web technologies. The system provides

information both to the drivers and to the city authorities. The drivers are informed about available parking slots through their smart phones or satellite navigators. By using these information, drivers can plan their trips better and they can find a parking spot quickly, thus reducing traffic and pollution in the city. City authorities are provided with statistics on usage of parking locations, detailed parking usage analysis, and suggestions on optimized transportation policies. These pieces of information can be used for planning more effectively the traffic flow and the use of parking locations.

Bmob achieved the aforementioned goals by installing a wireless vehicle detection sensor in each parking slot and through the use of video processing techniques. The information about parking status is then transmitted to a central server that performs data processing.

With the help of the Lugano municipality a parking monitoring system that covers 75 parking slots has been deployed in a district of Lugano in a project called Parkomotivo. Another 75 parking slots will be included in the system shortly.

## Bmob: We Know Where to Park

### The Position Detection System

In the framework of Parkomotivo, a collaboration has been established with the ALaRI Institute (Faculty of Informatics) of USI and with the Department of Innovative Technologies of SUPSI. During this collaboration, financed by CTI (the Swiss Commission for Technology and Innovation), a method for position detection of subscriber vehicles has been developed and integrated into the Parkomotivo installation. The aim of the position detection system is to allow parking subscribers to be identified and to allow the system to associate the position of the vehicle with a parking slot. The system is based on wireless sensor nodes placed inside the vehicles of subscribers. These nodes are switched on, by using an algorithm and a hardware also developed in the project, when a stop of the car is detected and they connect with nearby nodes. Information on these connections are used to compute a possible position of the vehicle. This position is then compared with the arrival times of vehicles in nearby parking slots and the vehicle is associated to a parking slot. The project was concluded successfully at the end of August 2011.

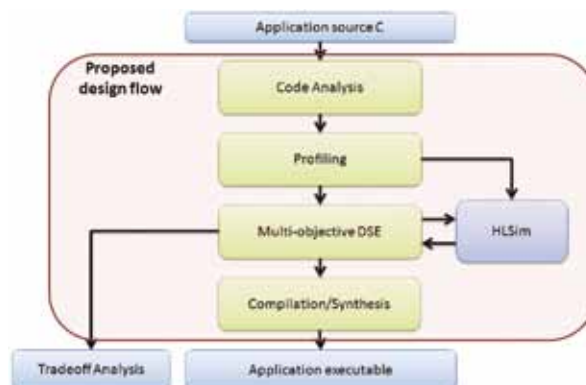
Alberto Ferrante,  
ALaRI institute, Switzerland ■

## Collaboration Grant Report - Giovanni Mariani



My name is Giovanni Mariani and in the current year I completed my PhD at ALaRI institute of the University of Lugano. During the PhD studies, I worked with Prof. Sami and Prof. Silvano in the context of the MULTICUBE project. After completing my PhD studies I obtained a collaboration grant from HiPEAC and I had the opportunity to move at TU Delft for three months to follow research in the field of multi-objective optimization of reconfigurable architectures. My work has been characterized by a strong collaboration with the local team lead by Prof. Koen Bertels.

In the target reconfigurable architecture, a General Purpose Processor (GPP) is complemented with reconfigurable logic. This heterogeneity is exploited to speedup the target applications by mapping computational intensive functions onto the reconfigurable hardware. When multiple applications are executed on the target reconfigurable platform, those will compete to access the GPP as well as the reconfigurable logic. At run-time, to efficiently allocate system resources to the active applications, detailed information on the performance/resource-cost trade-offs associated with each application are needed. In this context, we developed a novel automatic design flow that input is the high level description of an application



Proposed design flow

and which output is the mentioned trade-off information together with the application binaries to be executed on the GPP as well as the bitstreams necessary to configure the reconfigurable logic.

The proposed design flow has been implemented by integrating MULTICUBE Explorer with the hArtes tool-chain (respectively the outcomes of FP7 MULTICUBE project and FP6 hArtes project). The input of the design flow is the application description given in form of C code. The input application code is parsed and the intermediate representation is analyzed to identify the functions which can be executed on the reconfigurable hardware. The mapping of functions that present either direct or indirect recursion, as well as system calls invocations, is forced on the GPP. Those functions suitable for hardware execution are profiled to collect performance indices on the GPP

and on the reconfigurable hardware. At this stage, to avoid the long synthesis phase, hardware profiling is obtained using VHDL simulations, where the VHDL is generated using *dwarv*, a high-level-synthesis compiler.

Execution costs of the profiled functions are annotated to the application code and a High Level Simulator (HLSim) is automatically generated from the design flow. Then, Multicube Explorer executes a multi objective Design Space Exploration (DES) phase to map functions onto the available computing elements. The application executables are finally generated by compiling and synthesizing the different parts of the application. In particular, multiple versions of the same application are generated. Each version implements a different performance/resource-cost trade off. At run-time the trade off analysis generated during the DSE phase is used by the operating system (OS) to select, based on the system state and the application set to run on the platform, which application versions should be executed to maximize system performance while fitting in the available resources.

Giovanni Mariani,  
University of Lugano, Switzerland ■

## PhD News

By **Juan Manuel Cebrián**  
([jm.cebriangonzalez@gmail.com](mailto:jm.cebriangonzalez@gmail.com))  
Advisors: **Juan Luis Aragón,**  
**Stefanos Kaxiras**  
**University of Murcia, Spain**  
**September 2011**

## Efficient Power and Thermal Management Using Fine-grain Architectural Approaches in Multicores

In the last decade computer engineers have faced changes in the way microprocessors are designed. New microprocessors do not only need to be faster than the previous generation, but also be feasible in terms of energy

consumption and thermal dissipation. These microprocessors face constant thermal and power related problems during their everyday use. We have worked in the design, implementation and testing of

microarchitecture techniques for accurately adapting the processor performance to power constraints. We first designed “Power-Tokens”, to approximate the power being consumed by the processor in real time. For the single-core scenario we propose: Power-Token Throttling (PTT), that uses per-cycle power information to decide if a new instruction should enter the pipeline or not. Basic Block Level Management (BBML) stores power information about the next basic block to be executed and based on that information it selects the most

suitable power saving mechanism to reduce power consumption. Finally, we propose a Two-Level approach that uses Dynamic Voltage and Frequency Scaling (DVFS) as a coarse grain approach to lower the average power consumption towards the power budget and then use micro-architectural techniques to remove power spikes.

We ported all the previously proposed mechanisms to a multicore scenario and discovered that they are not suitable for parallel workloads due to synchronization points. Therefore, we

proposed the use of a Power-Token based load Balancer (PTB) that allows cores to run at full speed as long as there is power left to burn from idle/spinning cores.

Finally, we implemented a 3D die-stacked version of our processor design and checked the usability of the previously proposed mechanisms in this scenario. We proposed a new policy for the PTB that takes into account thermal and layout information. We also proposed some layout optimizations for 3D die-stacked vertical designs.

### Techniques to Cancel Execution Early to Improve Processor Efficiency

**By Mafijul Md Islam**

**(mafijul.islam@chalmers.se)**

**Advisor: Professor Per Stenstrom  
Chalmers University of Technology,  
Sweden**

**June 2011**

The evolution of computer systems to continuously improve execution efficiency has traditionally embraced various approaches across microprocessor generations. Unfortunately, contemporary processors still suffer from several inefficiencies although they offer tremendous computing capabilities. At the same time, the traditional approach of solely caring about performance is nowadays superseded with more critical and multi-dimensional constraints. This dissertation aims to address the prevailing inefficiencies and contributes with a number of techniques to improve performance as well as

energy efficiency of processors.

The first contribution is a novel scheme that detects and eliminates execution of trivial operations, such as multiplication by ‘0’ or ‘1’, early to improve energy efficiency. The second contribution is the finding that trivial computation and instruction reuse — two techniques to target program inefficiency — detect almost disjoint sets of instructions and may provide additive gains if combined. The next set of contributions increases execution efficiency of memory instructions by eliminating memory accesses early. To this end, the third contribution is a novel scheme that leverages frequent value locality and establishes that a significant fraction of memory instructions reads the value ‘0’ from memory. This dissertation then contributes with another micro-architectural technique that stores small values compactly to reduce

architectural inefficiency and eliminates unnecessary memory accesses to reduce program inefficiency. The penultimate scheme observes that a notable fraction of memory requests can be satisfied by the contents of physical register file and makes the associated memory accesses unnecessary. Finally, this dissertation presents a new unified scheme that employs a single structure to simultaneously target multiple forms of program inefficiency in memory instructions. Experimental results show that the proposed schemes improve performance and energy efficiency of processors. The proposed techniques are in general non-speculative and additional resource requirements are moderately low. Consequently, this dissertation contributes to resource-efficient and complexity-effective processor design.

### Integrated Code Generation

**By Mattias Eriksson**

**(mater@ida.liu.se)**

**Advisor: Prof. Christoph W. Kessler  
Linköping University, Sweden**

**June 2011**

Code generation in a compiler backend is commonly divided into sev-

eral phases that each take care of one particular aspect at a time, such as instruction selection, instruction scheduling, register allocation, spill code generation, and, in the case of clustered architectures, cluster assignment. However, these phases are usually interdependent; for instance, a

decision in the instruction selection phase affects how an operation can be scheduled.

In this work, we examine the effect of this separation of phases on the quality of the generated code. In order to study this, we have developed optimal methods for both integrated and

separated code generation scenarios, mostly based on integer linear programming.

For a generic class of instruction-level parallel target processors that includes VLIW and clustered VLIW processors, we first consider the case of acyclic code and then extend our method to modulo scheduling of loops. In particular, for a class of processor architectures that we call transfer-free, we derive an upper bound on the schedule length that allows to limit

the search space of our algorithm. Our experiments provide a quantitative assessment of the benefit of integration. For instance, when comparing optimal modulo scheduling with all above mentioned phases integrated to modulo scheduling where instruction selection and cluster assignment are done in a separate phase from the others, we find that, for a two-clustered VLIW architecture such as the TI C6x DSP series, the integrated method produces faster code than

the non-integrated method for 39% of the benchmark problem instances.

Another code generation problem that we studied is how to optimize the usage of the address generation unit in simple DSP processors. Here, the subtasks are: instruction scheduling, address register assignment, and stack layout; our results show that integration is beneficial when there are only a few (1 or 2) address registers available.

## Realistic Online Resource Management for Partially Reconfigurable Systems

By Yi Lu (luyi\_79@hotmail.com)

Advisors: Georgi Gaydadjiev,

Koen Bertels, Henk Sips

TU Delft, the Netherlands

July 2011

In this dissertation, we focus our research on the problems related to efficient configurable resource management for partially reconfigurable systems. FPGA devices are used to build such systems for various application domains with telecommunication and energy efficient high performance computing being two prominent examples. Dynamic management of FPGA hardware resources is an important problem and is the main motivation for this dissertation. Our research starts with investigating an abstract view of configurable resources which represents the essential properties of FPGAs in respect

to reconfiguration while leaving out many less important technology details. In this step, a realistic model with adequate complexity is exposed to the configurable resource management algorithms. Next, based on the abstract view, the hardware task's spatial requirements are studied and an efficient online task placement algorithm is proposed. Our placement algorithm dynamically redistributes the reconfigurable resources into blocks with various sizes and outperforms state of the art. In addition, a new model for measuring and analyzing the placement algorithms performance is built. In our next step we take into account also task's temporal requirements and consider holistic online task placement and scheduling. A novel algorithm with support for application specific scheduling heuristics is proposed. In

addition, a reuse and partial reuse mechanism is applied to alleviate the single configuration port limitation present in modern systems. After that, a communication model is introduced into the abstract view and the proposed online scheduling algorithm is extended to account the communication paths among data dependent hardware tasks and between tasks and external peripherals. In this step, the complete realistic configurable resource management problem is addressed. Furthermore, mechanisms for hardware reuse and interrupt handling are proposed. The hardware reuse mechanism gives the required hardware support for the reuse and partial reuse mechanisms. The hardware interrupt handling mechanism enables real time applications on reconfigurable systems.

## Efficient Runtime Management of Reconfigurable Hardware Resources

By Thomas Marconi

(marconi@comp.nus.edu.sg)

Advisors: Georgi Gaydadjiev, Koen

Bertels, Henk Sips

TU Delft, The Netherlands

June 2011

In this dissertation, we address three major challenges related to runtime management of hardware resources: efficient online hardware task schedul-

ing and placement, power consumption reduction and reconfiguration overhead minimization. Runtime reconfigurable systems built upon partially reconfigurable devices can provide reduction in overall hardware area, power efficiency, and economic cost in addition to the performance improvements due to better customization. However, the users of such systems have to be able to afford some

additional costs compared to hardwired application specific circuits. More precisely reconfigurable devices have higher power consumption, occupy larger silicon area and operate at lower speeds. Higher power consumption requires additional packaging cost, shortens chip lifetimes, requires expensive cooling systems, decreases system reliability and prohibits battery operation. The less efficient usage of silicon

real estate is usually compensated by the runtime hardware reconfiguration and functional units relocation. The available configuration data paths, however, have limited bandwidth that introduces overheads that may eclipse the dynamic reconfiguration benefits. Since hardware tasks are allocated and deallocated dynamically at runtime, the reconfigurable fabric can be fragmented. This can lead to the undesirable situation that tasks cannot be

allocated even if there would be sufficient free area available. As a result, the overall system performance is degraded. Therefore, efficient runtime management of hardware resources is very important. To manage hardware resources efficiently, we propose novel online hardware task scheduling and placement algorithms on partially reconfigurable devices with higher quality and faster execution compared to related proposals. To

cope with the high power consumption in field programmable devices, a novel logic element with lower power consumption compared to current approaches is proposed. To reduce runtime overhead, we augment the FPGA configuration circuit architecture and allow faster reconfiguration and relocation compared to current reconfigurable devices.

### Technology Aware Network-on-Chip Connectivity and Synchronization Design

**By: Daniele Ludovici**

**(daniele.ludovici@unife.it)**

**Advisors: Georgi N. Gaydadjiev,**

**TU Delft, The Netherlands**

**Davide Bertozzi, University of**

**Ferrara, Italy**

**June 2011**

NoCs have been considered as the new design paradigm for large MPSoC systems in the past ten years. In the beginning NoCs were radically different compared to the current state of the art mainly due to the unexpected unique challenges that system designers had to solve with the evolving CMOS technology. In fact, various hidden physical level issues may potentially degrade system performance, exceed the available power budgets or even endanger the overall design feasibility. The connectivity among different multi-core elements is such an issue that has

to be addressed during the design of the overall communication infrastructure.

Two different classes of implications related to the aggressive CMOS technology scaling, resulting in growing process variations, reduced power budgets per unit area and worsening signal integrity on chip, have to be considered. On one hand, a good topology is required to provide adequate sub-system connectivity while also satisfying the bandwidth and performance requirements.

On the other hand, increasing synchronization issues make the system design difficult and in some cases even impossible to be realized under a rigid synchronization model. For instance, the topology strongly depends on the physical effects as consequence of the wire delay reverse scaling while the synchronization issues are tightly related to process variation

effects. Therefore, in the current and the future CMOS technology nodes, ad-hoc counter measures must be adopted to cope with the above problems. In this thesis we propose a system-level analysis framework and design methodology both considering real layout effects. Our analysis is not only limited to classical layout effects such as the non-regularity of a rectangular tile; the real wire delays of inter-switch links; the number of pipeline stages required to provide the requested link performance; the maximum tolerated skew of a certain synchronization scheme; and more. We also consider the implications of the above physical phenomena while re-designing our architectural blocks. The ultimate result is a framework which is truly technology aware, ready to meet the challenges of the future CMOS technology landscape.

### Multipath Fault-Tolerant Routing Policies to Deal with Dynamic Link Failures in High Speed Interconnection Networks

**By Gonzalo Zarza**

**(gonzalo.zarza@caos.uab.es)**

**Advisor: Dr. Daniel Franco**

**Universitat Autònoma de Barcelona,**

**Spain**

**July 2011**

In this thesis, we present fault-tolerant routing policies based on concepts of

adaptability and deadlock freedom, capable of serving interconnection networks affected by a large number of dynamic link failures.

This work is framed in the context of high-performance computing (HPC) systems commonly used for running applications that require a large vol-

ume of data communication. These HPC systems are made of thousand of components, where the high-speed interconnection network plays a key role by allowing these systems to work as large coherent entities. Under these circumstances, it is imperative to keep the interconnection network up and running as long as possible since

networks failures have serious impacts on the overall computing system. This provides the motivation for this thesis.

The strongest point of this thesis is that it provides a simple but complete solution to the problem of dynamic fault tolerance in interconnection networks. The proposed solution does not require any information about network faults when the system is started or restarted. In this thesis,

we present the conception, design, implementation and evaluation of two contributions. The first of these contributions is the adaptive multipath routing method Fault-Tolerant Distributed Routing Balancing (FT-DRB). This method has been designed to exploit the communication path redundancy available in many network topologies, allowing interconnection networks to perform in the presence of a large number of faults. The second

contribution is the scalable deadlock avoidance technique Non-blocking Adaptive Cycles (NAC), specifically designed for interconnection networks suffering from a large number of failures. This technique has been designed and implemented with the aim of ensuring freedom from deadlocks in the proposed fault-tolerant routing method FT-DRB.

### Methodology for Efficient Execution of SPMD Applications on Multicore Clusters

**By Ronal Muresano**

**(rmuresano@caos.uab.es)**

**Advisor: Prof. Emilio Luque**

**University Aut3noma of Barcelona, Spain**

**Formal graduation date: July 2011**

The need to efficiently execute parallel applications in heterogeneous environments is a current challenge for parallel computing programmers. For this reason, the communication heterogeneities found in multicore clusters need to be addressed properly if programmer wish to obtain an improvement in efficiency and speedup. Under this focus, this work describes a methodology which is based on achieving the maximum speedup while the efficiency is maintained over a defined threshold. This goal is applied to SPMD (Single Program Multiple Data) applications which are designed with pure MPI

(Message Passing Interface). Also, this paradigm was selected because these applications present data synchronization and communications volumes which generate communication imbalance issues when we use hierarchical communication architecture such as Multicore cluster.

To achieve the goal, the methodology calculates the number of tiles and core that have to be assigned to each core. These tiles are divides in two groups internal and edge. This division enables us to apply an overlapping strategy between the internal computation and edge communication with the aim of hide the communication effects and improve the performance. Hence, the methodology was divided in four phases: a characterization (application and environment), a tile distribution model (determines the number of tiles and cores), map-

ping strategy (distribution of tiles over cores), and scheduling policy (defines the execution order).

Also, this works addresses how we can combine the efficiency and scalability in parallel applications. Using our method, we can observe how SPMD applications with some specific characteristics can present a linear speedup while the overlapping conditions are maintained. This methodology has been tested with different benchmarks and applications over different multicore cluster composed between 64 to 4096 cores and the results shows an improvement around 40% in efficiency in application tested.

This research was supported by the MICINN (Spain) under contract TIN2007-64974

### Programming, Debugging, Profiling and Optimizing Transactional Memory Programs

**By Ferad Zylkyarov**

**(ferad.zylkyarov@bsc.es)**

**Advisors: Mateo Valero, Osman**

**Unsal, Adrian Cristal**

**Universitat Polit3cnica de Catalunya, Spain**

**July 2011**

Transactional memory (TM) is a new optimistic synchronization technique which has the potential of making

shared memory parallel programming easier compared to locks without giving up from the performance. This thesis explores four aspects in the research of transactional memory. First, it studies how programming with TM compares to locks. During the course of work, it develops the first real transactional application – AtomicQuake. AtomicQuake is adapted from the parallel version

of the Quake game server by replacing all lock-based synchronization with atomic blocks. Findings suggest that programming with TM is indeed easier than locks. However the performance of current software TM systems falls behind the efficiently implemented lock-based versions of the same program. Also, the same findings report that the proposed language level extensions are not

sufficient for developing robust production level software and that the existing development tools such as compilers, debuggers, and profilers lack support for developing transactional application.

Second, this thesis introduces new set of debugging principles and abstractions. These new debugging principles and abstractions enable debugging synchronization errors which manifest at coarse atomic block level, wrong code inside atomic blocks, and also performance errors related to the implementation of the atomic block. The new debugging principles distinguish between debugging at the language level constructs such as atomic blocks and debugging the atomic blocks based on how they are implemented whether TM or lock inference. These ideas are demonstrated by implementing a debugger extension for WinDbg and the ahead-of-

time C# to X86 Bartok-STM compiler.

Third, this thesis investigates the type of performance bottlenecks in TM applications and introduces new profiling techniques to find and understand these bottlenecks. The new profiling techniques provide in-depth and comprehensive information about the wasted work caused by aborting transactions. The individual profiling abstractions can be grouped in three groups: (i) techniques to identify multiple conflicts from a single program run, (ii) techniques to describe the data structures involved in conflicts by using a symbolic path through the heap, rather than a machine address, and (iii) visualization techniques to summarize which transactions conflict most. The ideas were demonstrated by building a lightweight profiling framework for Bartok-STM and an offline tool which process and display the profiling data.

Forth, this thesis explores and introduces new TM specific optimizations which target the wasted work due to aborting transactions. Using the results obtained with the profiling tool it analyzes and optimizes several applications from the STAMP benchmark suite. The profiling techniques effectively revealed TM-specific bottlenecks such as false conflicts and contentions accesses to data structures. The discovered bottlenecks were subsequently eliminated with using the new optimization techniques. Among the optimization highlights are the transaction checkpoints which reduced the wasted work in Intruder with 40%, decomposing objects to eliminate false conflicts in Bayes, early release in Labyrinth which decreased wasted work from 98% to 1%, using less contentions data structures such as chained hashtable in Intruder and Genome which have higher degree of parallelism.

### Flexible Design and Dynamic Utilization of Adaptive Scalable Multi-Core Systems

**By Diana Goehringer**  
([diana.goehringer@iosb.fraunhofer.de](mailto:diana.goehringer@iosb.fraunhofer.de))  
**Advisors: Prof. Dr.-Ing. Juergen Becker, Prof. Dr.-Ing. Reiner Hartenstein**  
**Karlsruhe Institute of Technology, Germany**  
**July 2011**

Multiprocessor Systems-on-Chip (MPSoCs) are a promising solution to fulfill the performance requirements of embedded high performance computing applications, because of their parallel execution model. However, MPSoCs have a fixed hardware architecture where only the software can be adapted to a specific request. On the contrary, Field Programmable Gate Arrays (FPGAs) provide a very flexible and reprogrammable hardware architecture. This means, the hardware topology, the memory and the processor's microarchitecture of an FPGA

system can be adapted at design- and at runtime to the application requirements.

This dissertation introduces the holistic RAMPSoC (Runtime Adaptive MPSoC) approach, which combines the multiprocessor system performance, the reconfigurable system flexibility and the parallelism of FPGAs. RAMPSoC consists of three main components: the heterogeneous runtime adaptive MPSoC architecture, a user-guided design methodology and a runtime operating system.

RAMPSoC supports the runtime adaptation of the processors, the communication infrastructure and the accelerators resulting in a high energy efficiency. To support the runtime adaptation of the communication infrastructure and to analyze the communication patterns of the applications at runtime, the novel hybrid Star-Wheels Network-on-Chip was

developed. To hide the complexity of this novel hardware architecture, a user-guided design methodology was developed. This design methodology analyzes and partitions MATLAB code or C/C++ code, and it generates the appropriate hardware architecture including the configuration files for the FPGA. To manage the runtime adaptation of the software and the hardware and to virtualize the RAMPSoC architecture, a special abstraction layer called RAMPSoC Virtual Machine (RAMPSoCVM) has been developed, which uses a special purpose operating system called CAP-OS (Configuration Access Port - Operating System) to schedule and map the applications onto the RAMPSoC architecture. The benefits of the RAMPSoC approach have been demonstrated using various real-world applications from different domains, such as signal processing, image processing and bioinformatics.

## Optimizing Algorithms for Task Graph Mapping on Multiprocessor System on Chip

By **Heikki Orsila**

(heikki.orsila@iki.fi)

Advisor: **Professor Timo Hämäläinen**

**Tampere University of Technology,  
Finland**

**August 2011**

The Thesis analyzes and improves MPSoC design space exploration, specifically the task mapping using Simulated Annealing (SA) with fully automatic optimization. The work concentrates mostly on application execution time, but a trade-off between memory buffer and time as well as power and time optimization is also analyzed. Applications are represented as public Standard Task Graph sets and Kahn Process Networks (KPNs).

Main focus is on SA as the optimization algorithm for task mapping. A state of the art survey is presented, as well as analysis of the impact of SA parameters

on convergence. A new systematic method is developed for automatically selecting SA parameters. The method scales up with respect to increasing HW and SW complexity. It is called Simulated Annealing with Automatic Temperature (SA+AT).

SA+AT is compared with developed OSM (Optimal Subset Mapping), Group Migration, Random Mapping and a Genetic Algorithm. SA+AT gives the best results, while OSM is the most efficient algorithm.

Results include global optimum convergence properties and the convergence speed in terms of mapping iterations. This covers optimization of the runtime of mapping algorithms so that a trade-off can be made between solution quality and algorithm's execution time.

For a problem with 32 tasks and 2 processing elements SA+AT was able

to find the global optimum by testing 100k solutions on average, compared to 2T solutions by brute force. When trading off solution quality, solutions within 5 percent of the best solution was found by testing only 10k solutions on average.

As a conclusion of the work, SA+AT saves up to half the optimization time without significantly decreasing solution quality. Thesis makes detailed recommendations on how to use SA for distributing tasks on multiprocessor systems.

The thesis also presents a research tool called "DCS task mapper" including Simulated Annealing, Genetic Algorithms, Optimal Subset Mapping and other task distribution algorithms. The tool and benchmarks from the thesis are published with an Open Source license to facilitate further independent research and verification of results.

## Design and Implementation of Configurable Motion Estimation Architecture for Video Encoding

By **Jarno Vanne (jarno.vanne@tut.fi)**

Advisor: **Prof Timo D. Hämäläinen**

**Tampere University of Technology,  
Finland**

**October 2011**

Modern video compression involves complex encoding algorithms whose real-time execution necessitates efficient implementations. The emerging trend is that the same encoder is also compatible with plethora of video coding standards, so real-time encoding performance has to be met without compromising flexibility. In mobile devices, real-time constraints and flexibility expectations of the encoders are combined with limitations on cost, size, and power consumption.

The Thesis focuses on motion estimation (ME) that is the main encoding tool for removing temporal redundancy of video scenes. It typically accounts for 50 - 90% of the total video encoding complexity. Most of the current ME architectures are limited to a single video coding standard or ME algorithm. In turn, flexible architectures have large area, limited speed, unsustainable power budget, or restricted ME parameters.

The main contribution of the Thesis is a novel configurable ME architecture that overcomes all the crucial limitations faced by the contemporary approaches. The implemented hardware architecture is compatible with H.261/3, MPEG-

1/2, MPEG-4 Visual, H.264/AVC, and VC-1 standards. In addition, it can perform rate-constrained integer ME with various fast ME algorithms such as BBGDS, CDS, DS, HEXBS, and TSS.

The flexibility is not obtained at a cost of performance degradation, but the presented architecture is able to process 1080p resolution H.264/AVC ME at 30 fps in 50kgates and 364 mW power using 0.13um CMOS. The gate count is 40% less than in other proposals. Results are achieved through optimizing ME operations at algorithm and circuit level, designing a parallel memory system for ME, and designing generic control structures that are compatible with multiple ME algorithms.

## Parallel Video Decoding

By **Mauricio Alvarez-Mesa**

(mauricio.am0@gmail.com)

Advisors: **Alex Ramirez and Mateo Valero**

**Universitat Politècnica de Catalunya,  
Spain**

**September 2011**

The main objective of this thesis is to provide the performance required for

real-time operation of High Definition (HD) video decoding applications using programmable architectures. The proposed solution has been the simultaneous exploitation of multiple levels of parallelism. We made contributions in

two areas: enhancing the support for SIMD extensions and improving the parallel scalability on multi- and many-core architectures.

First, we performed a workload characterization of H.264/AVC for HD applications. We identified the main kernels and compared them with the kernels of previous video codecs. Due to the lack of a proper benchmark for HD video decoding we developed our own one, called HD-VideoBench. After that, we optimized the most relevant kernels of the H.264/AVC decoder using SIMD instructions. In order to improve SIMD efficiency

we implemented and evaluated the required hardware and software for supporting unaligned accesses in SIMD extensions.

Then, we developed an investigation on how to extract Thread-Level-Parallelism in H.264 decoding. We developed a new algorithm, called the dynamic 3D-wave, that is able to reveal thousands of independent tasks exploiting macroblock-level parallelism both intra- and inter-frame.

We implemented intra-frame macroblock-level parallelism on a parallel machine. But this implementation was not able to reach the maximum per-

formance due to the negative impact of thread synchronization and the effect of the entropy decoding kernel. In order to eliminate these bottlenecks we proposed a parallelization of the entropy decoding stage at the frame-level combined with a parallelization of the other kernels at the macroblock-level. A further performance increase was obtained by using different type of processors for each type of kernel. The overhead of thread synchronization was almost eliminated with a special hardware support for synchronization operations.

## Upcoming Events

**The 44th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO-44)**

3-7 December 2011, Porto Alegre, Brazil, <http://www.microarch.org/micro44>



**The 17th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2012)**

3-7 March 2012, London, UK, [http://research.microsoft.com/asplos\\_2012](http://research.microsoft.com/asplos_2012)



**The International Conference on Compiler Construction (CC 2012)**

24 March - 1 April 2012, Tallinn, Estonia, <http://conferences.inf.ed.ac.uk/cc2012>



**The International Symposium on Code Generation and Optimization (CGO 2012)**

31 March - 4 April 2012 San Jose, California, USA, <http://www.cgo.org/cgo2012/>



**The IEEE International Symposium on Performance Analysis of Systems and Software, (ISPASS 2012)**

1-3 April 2012, New Brunswick, NJ, <http://ispass.org/ispass2012/>



**The European Conference on Computer Systems (EuroSys 2012)**

10-13 April 2012, Bern, Switzerland, <http://eurosyst2012.unibe.ch>



**The 9th European Dependable Computing Conference (EDCC 2012)**

8-11 May 2012, Sibiu, Romania, <http://edcc.dependability.org/>

**The 17th International Conference on Reliable Software Technologies (Ada-Europe 2012)**

11-15 June 2012, Stockholm, Sweden, <http://www.ada-europe.org/conference2012>



**The Design Automation Conference (DAC 49),**

3-7 June 2012, San Francisco, CA, USA, <http://www.dac.com/dac+2012.aspx>



**The 18th IEEE International Symposium on High-Performance Computer Architecture (HPCA-18)**

25-26 February 2012, New Orleans, LA, <http://www.hpacconf.org/hpca18>



### Contributions

If you are a HiPEAC member and would like to contribute to future HiPEAC newsletters, please contact Rainer Leupers at [leupers@ice.rwth-aachen.de](mailto:leupers@ice.rwth-aachen.de)