



HiPEACinfo²⁰

COMPILATION ARCHITECTURE

appears quarterly | October 2009

Network of Excellence on High Performance and Embedded Architecture and Compilation

2 Message from the HiPEAC coordinator

3 Message from the project officer

HiPEAC Activity:

- 4 - HyperTransport Tutorial at Stanford University
- 4 - Rainer Leupers on his mini-sabbatical at ACE by
- 5 - Joint Seminar: RWTH Aachen University visits FORTH

Community News:

- 2 - Gadgets could go greener with high-speed computer chip
- 5 - Newsletter spell checking transition
- 6 - Ozcan Ozturk received the IBM Faculty Award

Announcement:

- ALaRI institute invites to attend Doctoral School on Complexity Management in Embedded Systems

In the Spotlight:

- 9th International Forum on Embedded MPSoC and Multicore (MPSoC 2009)

New HiPEAC Member:

- RuChip, Russian startup in Moscow

8 HiPEAC Students and Trip Reports

12 PhD News

16 Upcoming Events



**HiPEAC
Computing
Systems Week
in Wrocław
October 26-28**

www.HiPEAC.net

**HiPEAC 2010 Conference
Pisa, Italy, January 25-27**

Message from the HiPEAC coordinator

Koen De Bosschere



Dear friends,

I hope all of you have enjoyed a relaxing holiday season this summer. At the personal level, vacations are important to work on personal relationships, to enjoy hobbies, and to re-energize. In short, to keep a balance in life. For a network of excellence, the situation is different: it never takes a day off, not even in the summer.

In June, HiPEAC2 underwent its first review. The reviewers concluded that the project successfully kicked off, that we correctly managed the transition between HiPEAC1 and HiPEAC2, and that all activities are showing a healthy level of activity. The steering committee and the staff is now working hard to implement the recommendations formulated by the reviewers, more in particular, increasing the industrial involvement in HiPEAC, stimulating additional research interactions between the different research clusters and task forces, and further stimulating mobility through collaboration grants, internships and sabbaticals.

In July, more than 200 of us enjoyed the yearly ACACES summer school in La Mola, Barcelona. The facilities were stunning, the local organization and the courses were excellent, and the participants were enthusiastic about the whole event. Our last minute

change to a new location did not have an impact on the appreciation for the summer school. We are already preparing the ACACES 2010 summer school, which will be officially announced in the January newsletter. In October we organize our fall computing systems week in Wroclaw. This is the very first time that HiPEAC organizes an industrial workshop and the co-located cluster meetings in a new member state. We hope that this event will help our colleagues in Poland to get familiar with our network, to get involved and to start collaborations. HiPEAC is strongly committed to build stronger links with colleagues in the new member states.

Finally, there is the HiPEAC Conference, currently being organized by our Italian colleagues in beautiful Pisa, Italy. The conference runs for three days in January 2010, and it is preceded by a very rich set of workshops and tutori-

als, of which several are directly linked to the HiPEAC research clusters. The whole event is expected to be a major networking event for our community. The European commission has recently started consultation meetings to prepare the next call in computing systems. The HiPEAC community is collaborating actively in this effort, through its roadmap process, and also through bilateral meetings with HiPEAC members. We hope that this joint effort will eventually lead to a better understanding of what is needed for the further development of the computing systems domain in Europe. Its conclusions should also inspire future calls that will fuel our research. I hope to meet you at one of our coming networking events,

Take care,

Koen De Bosschere

Community News



THE UNIVERSITY OF EDINBURGH
informatics

On June 01, 2009 the Processor Automated Synthesis by iTerative Analysis (PASTA) research group at the University of Edinburgh led by Prof. Nigel Topham, announced its first successful silicon implementation of a new and versatile microprocessor, using a generic 130-nm process.

The microprocessor, known as EnCore, delivers faster processing while using

Gadgets could go greener with high-speed computer chip

significantly less power and taking up less space than comparable devices. Power consumption between 18 and 24-mW under heavy load has been verified in the lab. This compares favourably with 48-mW for an ARM Cortex M3 at the same technology node. The die area of EnCore is 0.15 square millimeter compared with 0.86 square millimeter for the Cortex M3, and EnCore achieves 1.45-DMIPS/MHz compared with 1.25 for the Cortex M3.

The microprocessor is configurable by

means of user-defined instruction set extensions (ISEs) - meaning that it can be automatically customized for a particular application, and so is suitable for a variety of application domains. The adaptability of the processor means that performance is not compromised by energy efficiency.

Multiple EnCore processors may be used together, creating high-performance multi-core systems for more demanding applications.



Message from the project officer

Following the evaluation of the last Call for Proposals, the European Commission is currently negotiating the start of one Support Action and nine Research Projects (STREPs) in Computing Systems representing 25 million Euros of funding.

Support Action

The PLANETHPC Support Action will establish a free-to-join network to bring together the High-Performance Computing community in order to exchange knowledge, to identify shared and complementary goals, as well as research challenges. Members will enjoy an on-line forum, special interest groups, workshops etc. The project will contribute to the European research roadmap of HPC.

Research Projects

2PARMA focuses on the definition of a parallel programming model combining component-based and single-instruction multiple-thread approaches, instruction set virtualization based on portable byte code and design space exploration methodologies for manycore computing fabrics.

The aim of ADVANCE is a performance-directed software development approach for multi-core hardware. Application specifications defined in an

extended description language, will be processed by applying newly developed static analysis and compiler transformations.

ENCORE aims to make multicore processors with hundreds of cores programmable through a combination of programming models, runtime management systems and resource/hardware support technologies.

EuroCloud plans to create low-power processors integrating 3D DRAM for very dense low-power server systems targeted at mobile "cloud" services. EuroCloud targets 10 times improvement in cost- and energy-efficiency compared to current state-of-the-art servers.

HEAP aims to develop an innovative toolset that helps software developers to profile and parallelise existing sequential applications. This is done by exploiting top-level pipeline-style parallelism and a highly configurable cache architecture that can be tailored to an application by using profiling data.

IOLanes address inefficiencies associated with the I/O stack of multicore architectures. The proposal targets performance and scalability issues of the I/O stack on multi-core architectures, I/O

performance and dynamic resource management issues in virtualized environments, as well as on- and off-loading tradeoffs.

PEPPHER aims to develop a platform-agnostic, multicore-oriented software architecture that will provide a general-purpose compositional software framework exploiting adaptive and auto-tuning technologies to facilitate the programming of generic hardware platforms.

PRO3D aims to develop a system software flow that can operate transparently on parallel manycore platforms including 3D stacked architectures and develop formal methods for software design guaranteeing the compositionality and correct operation of both hardware and software.

The aim of REFLECT is to investigate a set of methodologies and approaches for the design of efficient FPGA-based heterogeneous multi-core computing systems with a key issue being the use of aspect-oriented programming to cover critical domain knowledge.

Panos Tsarchopoulos

The EnCore microprocessor's architecture is compatible with the ARC600 family of processors from ARC International, with whom the university has a long record of collaboration.

In order to support a variety of research areas addressed within the PASTA project, the group has also developed a number of software tools accompanying the EnCore processor. Among the tools are a fully automated ISE design and compilation flow, an energy-aware compiler based on the com-



EnCore microprocessor

mercial CoSy compiler development framework, a GCC compiler based on the ARC port targeting the special

features of the EnCore processor, high-speed simulators at various abstraction levels, and HW/SW testing and verification tools.



Nigel Topham

Project Lead: Nigel Topham
(npt@inf.ed.ac.uk),
The University of Edinburgh,
<http://groups.inf.ed.ac.uk/pasta/>

HyperTransport Tutorial at Stanford University

Jose Duato (UPV & HyperTransport Consortium), Robert Safranek (Intel) and Jasmin Ajanovic (Intel) presented a tutorial on System Interconnects on Sunday, August 23, 2009, in the framework of the Hot Chips 21 conference that was held at The Memorial

Auditorium, Stanford University. Jose Duato, on behalf of the HyperTransport Consortium and AMD, presented the main design goals and the different protocol layers for the HyperTransport technology, as well as the main innovations in generation 3 and current

trends towards more scalable multiprocessor platforms, including the recently announced High Node Count specifications. The three tutorial presentations were followed by a panel session where speakers answered questions from the audience.

Rainer Leupers on his mini-sabbatical at ACE bv

RWTH Aachen University and ACE have enjoyed a long-standing and successful research partnership. Among the highlights of previous collaborations there is, for instance, the Compiler Designer (a product of HiPEAC member company CoWare Inc. that builds on ACE's CoSy framework and the LISA processor description language) which enables the rapid generation of C compilers for embedded processors. So, it is no surprise that I took the opportunity of one of the first mini-sabbaticals funded by HiPEAC to reinforce the RWTH-ACE relations via a three months research stay in Amsterdam during April-July 2009. In practice, this meant somehow reducing my duties as a professor to one day in the Aachen office per week and significantly increasing the weekly mileage of my car.

ACE's flagship product is the well-known CoSy system (see also HiPEAC Info-19), a versatile retargetable com-

piler framework. Due to its flexibility and performance, CoSy has a unique position in the market, and ACE is also engaged in various services and side products around CoSy. However, no software tool is made for eternity. ACE has been thinking about how to shape the future of their products to keep pace with important trends in computer architecture, such as multi-core platforms. My role as an expert in Electronic System Level (ESL) design was to consult on different directions for this and provide fresh views from the EDA and hardware design perspectives. The major findings have been summarized in a white paper that was delivered to ACE. Among many other by-products of the sabbatical there was also a joint software demonstration organized at the Design Automation Conference in San Francisco in July.

In return, I was able to participate in the day-to-day operations of the company,

including attending the regular team meetings as well as joint customer visits, which provided valuable insights and allowed me to get to know the team much better than before, also on a personal basis. Indeed, the best way to improve the mutual understanding with your industry partners and their respective operation modes, behaviours, concerns, and constraints is to spend time with them! I strongly encourage the HiPEAC community to also seek opportunities for similar mini-sabbaticals at interesting places and teams. There are all too few companies operating in the European tools arena and researchers in the HiPEAC community can certainly help stimulate activity and innovation here.

Last but not least I should mention that ACE is certainly a very special company to be with. Being successfully active in the IT domain for 30+ years, they combine a relaxed working atmosphere with a constant flow of innovations, implemented via highly skilled technical staff and seasoned management. I experienced a great amount of openness and hospitality, ranging from early support in sabbatical logistics to the waterborne farewell dinner on the Amsterdam Grachten. In fact, every visitor at ACE is received with open arms. So even if you don't plan to spend three months there (which by the way also means to survive on Dutch food), you are encouraged to knock on their door to discuss any compiler issues.



R. Leupers among his ACE colleagues (J. van Vlijmen, M. Schoorel, R. Leupers, M. De Lange, M. Roodzant)

Newsletter spell checking transition



Igor Böhm

One of the important procedures in the whole newsletter production process is spell checking. For many years Leigh Murray from the University of Edinburgh was taking care of this procedure assuring the required quality of the HiPEAC newsletter. Starting from newsletter issue 20 Igor Böhm will take over the spell checking duties from Leigh. Igor is a second year PhD student and member of the Compiler and Architecture Design Group (CaRD) at the University of Edinburgh. Moreover he has been an active HiPEAC member during the past year. The HiPEAC community would like to thank Leigh for her reliable service and welcome Igor to the newsletter production team.

HiPEAC Activity

Joint Seminar: RWTH Aachen University visits FORTH

The Institute of Computer Science (ICS) of the Foundation for Research and Technology Hellas (FORTH) invited the Institute of Integrated Signal Processing Systems (ISS) of RWTH Aachen University to a joint seminar. The two HiPEAC partners met on June 1st, 2009 at FORTH near Heraklion.

The first presentations of the seminar were given by FORTH about new approaches for increasing performance and scalability of storage systems as well as about parallel programming of multi-cores.

Performance was also a main theme in the presentations from Aachen, which discussed two new tools targeted at MPSoC software development: an accurate retargetable source code profiler and a tool for early high-level MPSoC software development.

In the afternoon, the researchers from FORTH talked about their work on merging cache and scratchpad communications on multi-core architectures and on explicit synchronization. The second session by RWTH introduced Virtual Platforms using a case study and showed how to use those for early MPSoC design space exploration.

Although both groups are performing research on different topics within the HiPEAC field, it was very interest-



ing for the group from RWTH to see what might be the next generation of hardware that has to be supported by design tools. On the other hand, the tools presented can facilitate research of FORTH. Next to the exchange in research, the visitors from Aachen also enjoyed the cultural highlights of Heraklion and a nice Greek dinner at a traditional tavern together with ICS. Such joint seminars are a great opportunity to get to know different working approaches and cultures and can help to develop new ideas that finally

may evolve into joint project proposals. Thus, similar meetings of HiPEAC research groups are highly encouraged!



FORTH building

Ozcan Ozturk received the IBM Faculty Award



Ozcan Ozturk (third from the left) and other award recipients

Ozcan Ozturk, assistant professor of Computer Engineering, Bilkent University, was awarded the 2009 IBM Faculty Award. The IBM Faculty Awards program is an annual worldwide competitive cash awards program that fosters collaboration between researchers at leading universities and those in IBM research, development, and services organizations. It also promotes courseware and curriculum innovation to stimulate growth in disciplines and areas that are strategic to IBM. Awardees are nominated by IBM employees and winners have an outstanding reputation for contributions in their field.

Professor Ozturk received the prestigious award for his "Utilizing Heterogeneous Chip Multiprocessors through Efficient Parallelization" research.

Congratulations on this achievement!

Announcement



In 2008, the Swiss Government launched the nano-tera.ch initiative, centered on research, development and application of micro- and nano-information technologies to embedded systems, networks and software to support health, security and environmental monitoring.

The intrinsic value of the underlying research is to bridge traditional disciplines, ranging from electrical engineering and computer/communication sciences to micro- and nano-mechanical systems engineering, biomedical sciences, etc., with the objective of deepening the understanding of enabling technologies and applying scientific concepts to practice, as well as of mastering the novel challenges of engineering tera-scale complex systems.

The nano-tera initiative also foresees educational projects. COMES belongs to this class of projects; its most relevant offering is a doctoral school dedicated to the core theme "Dealing with Complexity in Embedded Tera-Systems: the facets of the problem". In fact, if technologies make it possible to create systems of tera-complexity, the challenge of designing, simulating and

ALaRI institute invites to attend Doctoral School on Complexity Management in Embedded Systems

managing such systems arises. Only if such challenges are overcome will the systems become actually viable. Not only are systems envisioned intrinsically very complex: beyond that, they are in most instances devised to interact with the physical world, facing challenges that go from the "modelling" aspects of the phenomena they should deal with to intrinsic non-determinism of such phenomena.

The Autumn School will last five days, starting on November 16, 2009; the location will be the University of Lugano. Subjects discussed will include design complexity of systems on chip, management of very complex – possibly distributed - systems, where real-time constraints have to be met, computational complexity (with particular reference to modeling), the problem of dealing with uncertainty and the concept of "probably approximate correct computation", control complexity, and various other specific design aspects of complex software systems. In depth discussions about various theoretical aspects and analysis of case studies are also on the agenda. In order to allow students' evaluation, in particular for granting ECTS credits to participating PhD students, projects will be assigned. After

the end of the Autumn School students will

receive a certificate recognizing their participation and demonstrating both, project evaluation and ECTS credits obtained.

The coordinator of the School is prof. Mariagiovanna Sami (Politecnico di Milano and University of Lugano). The faculty includes (but is not limited to) prof. C. Alippi (Politecnico di Milano), prof. R. Leupers (RWTH Aachen University), prof. M. Pezzè (University of Lugano), prof. M. Polycarpou (University of Cyprus), prof. Lothar Thiele (ETH Zurich).

While the School is designed for PhD students, other participants are welcome as well, as long as there will be openings.

For further information please visit www.alari.ch/COMES

Contact: Daniela Dimitrova (daniela.dimitrova@usi.ch)

Timeline: November 16-20, 2009

Location: University of Lugano, Switzerland



9th International Forum on Embedded MPSoC and Multicore (MPSoC 2009)

From the 2nd to the 7th of August in the beautiful city of Savannah, Georgia, the 9th International Forum on Embedded MPSoC and Multicore was held. The unique structure of the forum allowed for excellent presentations and intensive face-to-face discussions between world-class experts from industry and academia.

With an attendance of more than 50 world-class speakers, the ninth event of the MPSoC focused on research issues yet to be mastered. The 5-day forum gave an impressive overview of present and expected future challenges in the topics of applications, software and hardware. Examples of the broad range of topics are efficient hardware architectures for Software Defined Radios, 3D chip stacking and the ubiquitous quest for design space exploration of software and hardware.



MPSoC'09 forum hall in Savannah



Torsten Kempf giving speech at MPSoC forum

The assembly of researchers from both, industry and academia, at MPSoC'09 provides a great platform for guiding academia to the relevant design challenges the industry is facing today. In turn, executives and senior managers are encouraged to explore new ideas and to rethink their strategies.

Apart from the brilliant technical contributions at MPSoC'09, plenty of social events resulted in bringing people together in order to intensify net-

working. The wonderful dinner at the Savannah river made MPSoC'09 a conference to be remembered.

In a nutshell MPSoC'09 was a memorable and fruitful conference with its unique character of in-depth discussion and information exchange of researchers from all over the world. I hope to visit the next MPSoC and to meet you there.

Torsten Kempf,
RWTH Aachen University, Germany

New Members



In July 2009, the HiPEAC Network steering committee accepted Ru.Chip.Llc. as a new member of the HiPEAC community. RuChip is a bootstrapping fabless startup.

The core idea for this startup came from realizing current market needs in developing energy saving technologies for global Internet search.

As an effort to tackle the problem of energy saving, Ru.Chip came up with an idea of a special purpose microprocessor with ultra low power consumption and multicore architecture optimized

RuChip, Russian startup in Moscow

for search tasks. The company also plans to develop a specialized searching device consisting of a number of multicore processors placed on a mother board, compatible with a search system data center's infrastructure.

The development is based on the following core technical approaches:

1. Embedded realization of the Nutch/Hadoop platform, in which Map Reduce algorithms will operate. Map Reduce algorithms can be applied in many different applications including search systems.
2. Processor power management through switching off separate unused cores or lowering core operating frequency.
3. Embedded instruction customization algorithms.



A founder of RuChip Anton Gerasimov

Each search microprocessor is expected to consume at least 10 times less power compared to conventional PCs with equivalent performance, consequently the use of a specialized search device will allow to decrease the costs of servers in data centers by several orders of magnitude.

Business model

The new processor and the mother-

New Members

board are supposed to be produced in accordance with the fabless model implies no proprietary production line. Some core groups of future customers for the developed products have been identified: global and regional Internet search systems; corporate and state data processing centers, corporate search systems and other users.

About RuChip

Ru.Chip LLC is a startup that was estab-

lished in 2008 by a group of IT specialists and innovation managers. Initial financing was granted by the Russian Foundation for Assistance to Small Innovative Enterprises.

Ru.Chip is temporarily headquartered in Moscow. The core staff of the company has acquired their original experience from semiconductor and software industry working for world leading companies such as STMicroelectronics and Cadence

Design Systems. Ru.Chip team members have also acquired an extensive experience from participating in a lot of outsourcing software and hardware design projects in US.

Contact: Anton Gerasimov
(anton.gerasimov@ruchip.com),
123458, Russia, Moscow,
Tvardovskogo str. 8 building 1,
Office 608

HiPEAC Students

This summer I attended the ACACES 2009 summer school. After the tragic earthquake hitting L'Aquila hard earlier this year, the organizers had gone to great lengths to find an alternative venue: the conference hotel La Mola in Terrassa, near Barcelona, Spain. The following is a brief report of the school and the events surrounding it.

What had started out as a gorgeous Sunday, turned sour when the flight some colleagues and I had reserved a seat on to get from Brussels to Barcelona, was delayed. And delayed some more. After we finally boarded, we were entertained for a fairly uneventful flight. Once we regained solid ground under our feet, we decided to try the public transportation, but changed our minds when we saw the train schedule. A taxi brought us to La Mola quite swiftly - even though the driver was not sure of its location initially. Immediately, the rather dismal trip was made up for by the great looking facilities, the awesome room and the proximity of an outdoor swimming pool. Not to mention the fast and free WiFi.

The first evening is traditionally marked by a keynote talk, followed by dinner. Both were excellent, and left nothing else to desire. The keynote talk given

by Steve Furber described the ambitious SpiNNaker project. Its long-term goal is to build a huge system consisting of tens of thousands of nodes equipped with ARM processors, being able to simulate billions of neurons in real time. Cylons, anyone?

The dinner (and all subsequent meals the following days) was in the form of a rich buffet, including cold and hot dishes, and very tasty desserts. The only drawback of this venue seemed to be the rather pricey beer at the bar, especially compared to the prices at the previous ACACES venue in L'Aquila.

The classes started in earnest on Monday, with three parallel sessions, of which I attended the WCET analysis course by Peter Puschner. Because I was the appointed photographer of the event by the organizer Koen De Bosschere, Monday included racing from classroom to classroom and get decent shots of each teacher. Evidence of this effort can be observed at the Flickr ACACES group (<http://flickr.com/groups/acaces>). In the evening, Koen showed a number of pictures before the invited talk on Monday evening given by Alasdair Rawsthorne, who completed last years' invited talk by informing us how to get out of a startup (alive). Alasdair gave quite a nice talk, relating the story of Transitive

and its acquisition by IBM, while the audience was twittering on the second screen - set up at Alasdair's behest.

On Tuesday, the classes started to take shape as the introductory material was finished. I had my opinion of Paul McKenney (Real time in the Linux kernel) confirmed: he made a fabulous impression on Monday and repeated this on Tuesday; his was the class I enjoyed most. He was a good teacher, not afraid to pose challenging questions to his audience and to build on their answers. He also had the most awesome slide illustrations (drawn by his daughter, we learned), that really highlighted the key aspects Paul tried to get across. The other classes I took were on process virtualization by Kim Hazelwood, which was very good, and taught by one of the few women who are active in computer science (research), and the course on performance analysis taught by Lieven Eeckhout, my former PhD advisor.

On Wednesday, the afternoon was reserved for the poster session. Students could display their ongoing (or starting) work on an A0 sized poster and inform the inquisitive audience of their approach and (hopefully) get valuable feedback. I saw quite a lot of people discussing and the poster I pre-

sented with Kenneth Hoste drew forth a number of questions as well.

Thursday saw the classes go to a climax, when the teachers disseminated the highlights of their course. Finally on Friday, the courses were concluded with some more highlights and loads of interesting information. I think I learned quite a few things and refreshed some others. Sadly, one cannot attend all twelve courses: choices must be made. Hindsight is 20/20, but still I'd like to have seen some of the other teachers as well. The (formal part of the) day ended with Koen giving an overview of the school, inviting us for next year and thanking the teachers and speakers. And of course Nacho Navarro, who took upon himself a large part in helping to organize this year's summer school after the relocation was decided upon.

After all was said and done, Koen invited us for the group photo, the tasty barbecue and ... the pool party. At the pool, a live cover band was setting up to entertain us, a feat in which they



Teaching team of ACACES School

flawlessly succeeded. Until the power was cut and we were left standing while repairs were underway. After the successful restoration of electricity flow to the guitars, mikes and amplifiers, the (remaining) crowd danced some more and some of us took another dip. Finally, at around 2 AM, the party died. After some small talk with local Spanish students, I retired to my room and nodded off to the tunes of the wedding party being held at the poster venue.

After two hours of sleep, I rose on Saturday to share a taxi with several colleagues - our flight left too early for us to wait on the bus that would take most attendants to the airport. Luckily our plane was on time and it was with much joy that I rushed to my wife and two sons after landing at Brussels.

To conclude, I can heartily recommend attending the HiPEAC/ACACES summer school. It's a great way to learn new things from top notch experts



Kim Hazelwood discusses Super Pin

in their field and to get in touch with fellow students. This was the second ACACES summer school I attended and once again, I thoroughly enjoyed it. The summer school is also a great way to start collaborations and to get word of your work out to people with similar interests.

Andy Georges
(andy.georges@elis.ugent.be),
Ghent University, Belgium

Collaboration Grant Report - Daniel Cabrera



I am a PhD student at the Technical University of Catalonia (Spain) working with Dr. Daniel Jimenez-Gonzalez and Dr. Xavier Martorell. I work for the programming models group at the Barcelona Supercomputing Center (BSC). My research topic is about programming heterogeneous platforms

focusing on increasing the ease of use of reconfigurable devices.

Thanks to HiPEAC I had the chance to join the team of Dr. Georgi Gaydadjiev at the Technical University of Delft during three months. It was a great opportunity for me to improve my knowledge about reconfigurable devices and sharing my experiences with other students about compilers and heterogeneous architectures such as the Cell/B.E. processor.

In this collaboration we focused on the problem of using FPGA devices for C applications. We extended the OpenMP 3.0 task approach to run tasks on FPGAs. With our extensions a programmer can easily express the offloading of an already existing recon-

figurable binary code (bit stream) hiding all the complexities related with device configuration, bit stream loading, data placement and movement to the device memory.

Furthermore we implemented a prototype runtime system based on the SGI RASC system in order to test our extensions. Our runtime includes the following main features: (1) a bit stream cache and support for hybrid computation. So the runtime system avoids unnecessary configurations of the FPGA device. This is done by keeping the currently loaded bit streams in a cache. In the case a bit stream has to be loaded (cache miss), the runtime overlaps the load and FPGA configuration with execution of the algorithm on the host processor, since the former can be

significantly expensive. (2) Transparent change of memory association.

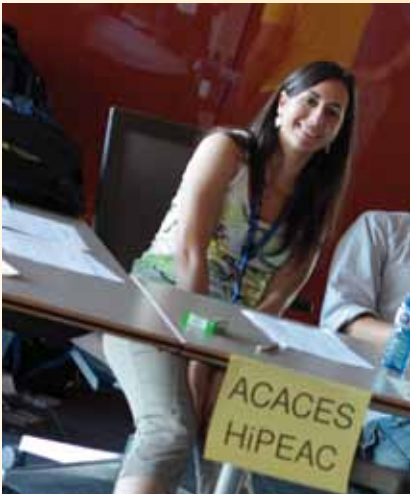
The runtime provides data packing and unpacking when transferring data between host and FPGA device, since data transfer can be a bottleneck, and (3) a multithreaded FPGA library interface. The runtime avoids any FPGA management operation if the operation blocks application execution. The

FPGA library interface is implemented using threads in order to avoid the application to be blocked during FPGA management operations.

Our proof-of-concept was successful and is explained in "OpenMP extensions for FPGA accelerators" in the 2009 proceedings of the International Conference on Embedded Computer

Systems: Architectures, Modeling and Simulation (IC-SAMOS), but still there are many things to do. We are currently working on supporting several FPGAs on SGI Altix Systems. Moreover our future working lines are the optimization of data movements as prefetching of data, new packing/unpacking techniques and partial runtime reconfiguration.

Collaboration Grant Report – Cecilia Gonzalez



I am Cecilia Gonzalez and currently I am working on my PhD at the Technical University of Catalonia (UPC) and the Barcelona Supercomputing Center (BSC). My advisors at UPC are Dr. Daniel Jimenez-Gonzalez and Dr. Carlos Alvarez, and my supervisor at BSC is Dr. Xavier Martorell.

My main topic of research is focused on hardware accelerators for the bioinformatics field. Concretely, I am interested in the automatic identification of instruction set extensions and their automatic implementation in a specialized processor.

Last year I received a grant from the HiPEAC Network to spend three months, from April to June, in the Computer Engineering department of the Technical University of Delft, under the supervision of Dr. Georgi Gaydadjiev.

There have been precedents of collaborations between TU Delft and the Computer Architecture Department at UPC, that share common interests in research of hardware accelerators and new architectures for scientific applications.

In this line of work, I went to Delft to further develop my master thesis. There, I have been trained on some topics that would be basic for my research. Those topics included the analysis of bottlenecks in bioinformatics applications, the identification of instruction set extensions, the use of FPGA-based architectures to evaluate application accelerators and the use of tools for automatic HDL code generation.

I worked on a toolchain for rapid generation of prototypes of accelerators that are identified with the analysis of the dynamic behaviour of the applications that we want to accelerate. In order to evaluate these prototypes, we have relied on MOLEN, a polymorphic processor developed at TU Delft. The toolchain is divided in two main parts: the detection of ISA extensions and the generation of hardware. In the detection of ISA extensions, we have used the Trimaran framework to profile the application and extract possible candidates for instructions extensions. Those candidates are subsequently pruned to suggest the final best set of new instructions. Besides, in the second part of the toolchain, we automati-

cally generate the code necessary to run the new instructions as accelerators in the MOLEN processor. Part of that generation is driven with the support of DWARV, a toolset also developed in TU Delft that translates C code to VHDL code that fits on the customized processor of MOLEN. The toolchain of prototyping has been tested with the bioinformatics application CLUSTALW. The tests show a speedup of up to 8.54x for a single accelerator, while the whole application can get a benefit of more than 2x of speedup.

The results of this collaboration have been presented at the 4th Workshop on Architectural Research Prototyping, held in conjunction with ISCA'09 in a paper entitled: Fast Evaluation Methodology for Automatic Custom Hardware Prototyping.

Although the stage had reached its end we have plans for future collaboration within the SARC project, to obtain accelerators of scientific applications. To continue the development of our toolchain of automatic generation of accelerators we need the expertise of VHDL coders and hardware designers, and this is the part of our project where we base our next collaboration with TU Delft. Likewise, people from our group contribute with the analysis of scientific applications that are useful for the engineers at Delft.

Collaboration Grant Report - Rafael Tornero

My name is Rafael Tornero and I am a PhD Student belonging to the Networks and Virtual Environments Group (GREV) of the University of Valencia, Spain. The GREV is part of the Advanced Communication and Computer Architecture (ACCA) Consortium. This Consortium, composed of a set of Groups each one belonging to a different Spanish university, is carrying out a project whose main goal is to significantly improve the performance and reliability of current server architectures for data centers and Internet servers, for a given cost and energy consumption budget. In this project I focus on improving the performance of on-chip interconnection architectures based on a Network-on-Chip (NoC) using communication-aware core mapping methods.



Last year, I did a collaboration with the University of Catania, Italy with Dr. Maurizio Palesi from the Department of Computer Engineering and Telecommunications. The collaboration was supported by the HiPEAC Network of Excellence under grant Cluster 1169.

The main goal of the collaboration consisted of integrating the core mapping technique for NoCs developed by the GREV group in the methodology for designing a routing algorithm for an application specific NoC (APSRA) developed by the group steered by Dr. Palesi.

The application mapping and routing strategy selection play an important role in NoC design, since they have a big impact on the communications exchanged by the application(s) running on the NoC Platform. Therefore,

the optimal NoC configuration cannot be found by addressing these problems independently.

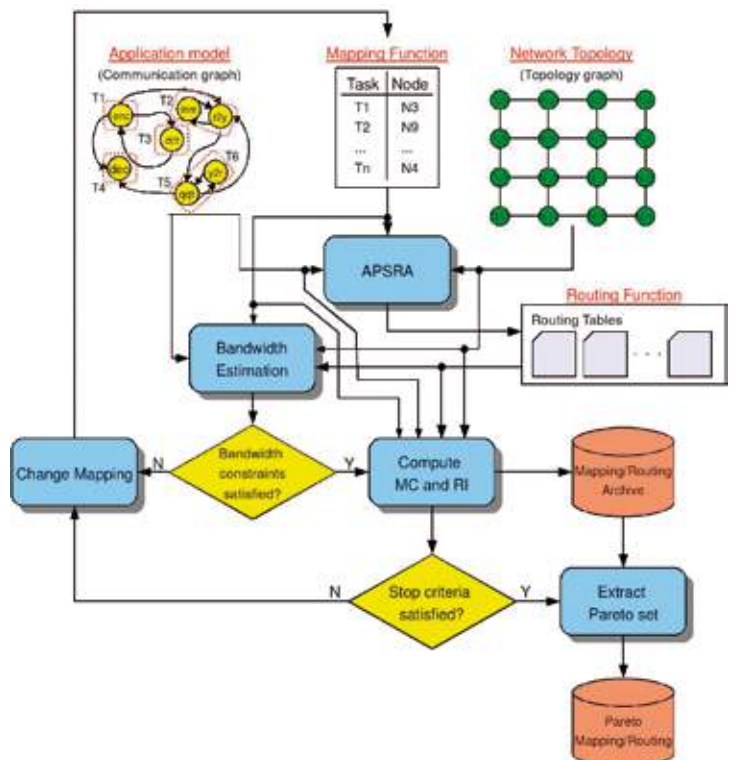
Before starting the collaboration, we knew that we wanted to jointly address these problems. When the collaboration started, we had a meeting in which we decided how to achieve our goals. We decided to implement a multi-objective design optimization strategy based on an evolutionary approach, in which the objectives to be optimized were the global communication delay and the fault tolerant properties of the NoC.

We spent the rest of the time implementing the design space exploration flow shown in the Figure. The inputs of the flow are the application model, a mapping from the cores of the application model to the NoC topology and the network topology of the NoC. The flow works as follows: in the first step, routes between any pair of communicating cores are obtained using APSRA. Once the bandwidth constraints of the application model are satisfied, the optimization indexes are computed. The Mapping Coefficient (MC) index is computed using the same approach as in the core mapping technique developed by the GREV Group. The Robustness Index (RI) is computed as an extension of the concept

of path diversity. Then, the current mapping along with the routing tables and the values of MC and RI are collected in an archive. In the next step, the mapping function is modified and another iteration is performed. When the stop criterion is satisfied, the result values are extracted from the archive.

A complete description of our work can be found in the reference paper: A Multi-objective Strategy for Concurrent Mapping and Routing in Networks on Chip. The authors of the paper are Rafael Tornero, Valentino Sterrantino, Maurizio Palesi and Juan M. Orduña. The paper has been presented at the 12th International Workshop on Nature Inspired Distributed Computing (NIDISC) held in conjunction with the 23th IEEE International Parallel & Distributed Processing Symposium (IPDPS) in May 2009, Rome, Italy. Currently, we have submitted an extended version of the paper to the International Journal of Foundations and Computer Science (IJFCS).

Design space flow



Collaboration Grant Report - Mounira Bachair



My name is Mounira Bachair. I am a PhD student at INRIA Saclay in Ile de France. At the end of 2006, I received a grant from INRIA Saclay to do my PhD studies about code size minimization in embedded systems within the ALCHEMY Group. I'm closely working with the assistant professor Sid-Ahmed-Ali Touati and the research director Albert Cohen.

High Performance Computing (HPC) techniques are increasingly used in embedded systems (multimedia applications, games, high-resolution printers, etc.). Decades of research in HPC have brought significant gains for classical HPC problems in scientific and numerical computing. For instance, software-pipelining methods become commonly used in many of the best optimizing compilers. However, current software pipelining techniques tend to be focused on producing the fastest possible code rather than dealing with the specific constraints of embedded applications, such as limited memory

for program code size. Establishing the fundamental relationship between code size and performance is still an open problem in computer science, even if several ad-hoc techniques have been developed in practice. Unfortunately, most of the existing ad-hoc techniques are simple heuristics that make tradeoffs between code size and performance whereas there does not exist any relevant study that proves that such tradeoff is mandatory.

A formal model of the relationship between speed and size of software-pipelined code would be of great value to the embedded software community. Many published software pipelining techniques under register, resource and sometimes code size constraints claim experimental improvements. But nobody can really check the real efficiency of such heuristics. However, if we are able to give a formal way to compute optimal solutions, it would be possible to compare all the existing techniques against the optimal solutions and hence we could objectively measure the efficiency of such methods.

Thanks to a HiPEAC PhD grant I did an interesting internship at Trinity College of Dublin under the direction of the assistant professor David Gregg.

I attempted to provide an optimizing compilation method that optimally reduces the code size of high performance loops (software pipelined loops) with the mathematical guarantee of non-losing performance. The technique is based on many fundamental results that prove some of our assertions. The different results from this internship can be found in our publication.

I think that staying in a foreign laboratory for three months is really exciting because one learns a lot from hearing other people's points of view and other ways of working. It is also an excellent way to determine if industry or academic is the best career option to pursue. Interns not only gain practical work experience in a field that they intend to pursue but also build experience in international platforms.

The HiPEAC network proposes great opportunities for PhD students. I hope that in the future more students will benefit from these experiences.

As conclusion, I would like to thank HiPEAC for this wonderful experience and Dr David Gregg and his research group at Trinity College of Dublin for their warm welcome and valuable advice.

PhD News

Fault-tolerant Cache Coherence Protocols for CMPs

by **Ricardo Fernández-Pascual**
(rfernandez@ditc.um.es)

Advisor:

José Manuel García Carrasco
and **Manuel E. Acacio Sánchez**
Universidad de Murcia, Spain
July 2009

We propose a new way to deal with transient faults in the interconnection network of future many-core CMPs that is different from the classic approach of building a fault-tolerant interconnection network. In particular, we provide fault tolerance mechanisms at the level

of the cache coherence protocol so that it guarantees the correct execution of programs even when the underlying interconnection network does not deliver all messages correctly. This way, we can take advantage of the different meaning of each message to achieve

fault tolerance with lower overhead than at the level of the interconnection network, which has to treat all messages alike with respect to reliability.

To demonstrate our approach, we design three fault-tolerant cache coherence protocols. First, we design FtTokenCMP, based on the token coherence framework. Secondly, we design FtDirCMP: a directory-based fault-tolerant cache coherence protocol with techniques inspired by the

previous work in FtTokenCMP. Finally, the same ideas are used to design FtHammerCMP: a broadcast-based and snoopy-like fault-tolerant cache coherence protocol based on the cache coherence protocol used by AMD in their Opteron processors.

We evaluate these protocols using full-system simulation. The results of this evaluation show that, in absence of faults, our techniques do not increase significantly the execution time of the

applications and their major cost is an increase in network traffic due to acknowledgment messages that ensure the reliable transference of ownership between coherence nodes, which are sent out of the critical path of cache misses. The results also show that a system using our protocols degrades gracefully when transient faults actually happen. Furthermore, we are able to support fault rates much higher than those expected in the real world with only small performance degradation.

Implementations of Baseband Functions for Digital Receivers

by **Perttu Salmela**
(perttu.salmela@tut.fi)

Advisor: **Prof. Jarmo Takala**

**Tampere University of Technology,
Finland**

August 2008

With ever-higher data rates, the complexity of baseband processing increases basically for two reasons. Firstly, the required processing rate is proportional to the bit rate and, secondly, with higher data rates, more demanding and sophisticated algorithms must be applied. For example, new wireless telecommunications systems like 3G long term evolution (LTE) can have a 100 Mbps data rate and multiple-input multiple-output (MIMO) transmission methods are applied. Thus, the prob-

lem domain of implementing baseband functions includes both addressing the high computational complexity and describing the implementations in a flexible way so that even complex algorithms can be used without extensive efforts.

In this thesis, implementations and implementation methods of baseband processing functions are proposed. Computational complexity and flexibility of implementation are approached with application-specific processors (ASP) and the transport triggered architecture (TTA) has been used as an architecture template. The computing demands can be met with high parallelism when parallelization of the target algorithm is possible, and the software

description of the computation possesses enough flexibility. Especially, the error correction decoding, matrix decomposition, and symbol detection tasks of the baseband processing chain are targeted in this thesis. Both processor implementations and implementations of assisting hardware units are presented.

As a result, the essential computational challenges and the design space of wireless receivers are clarified. The work in this thesis shows how the computation of the addressed baseband functions can be implemented efficiently when a programmable platform is targeted. The results show that the benefits of the programmability do not sacrifice implementation efficiency.

Implementing Fine/Medium Grained TLP Support in Multi-core Architectures

by **Nikola Puzovic**
(nikola.puzovic@gmail.com)

Advisor: **Prof. Roberto Giorgi**

University of Siena, Italy

September 2009

Future multi-core architectures should support a simple and scalable way to execute many threads that are generated by parallel programs. A good candidate to implement an efficient and scalable execution of threads is the Decoupled Threaded Architecture (DTA) that is designed to exploit fine/medium

grained Thread Level Parallelism (TLP) by using a hardware scheduling unit and relying on existing simple cores (there is no need for deep out-of-order pipelines, branch predictors or ROB).

The purpose of this thesis is to show that DTA can be flexibly adapted to different scenarios (from a standard Cell processor to more complex ScalAble aRchiteCtures, such as those envisioned in the SARC project) and efficiently include DMA-based prefetching mechanisms. Therefore, this thesis presents three case studies of a DTA implemen-

tation in multi-core architectures.

The first case study is an implementation of DTA support in the Cell processor. It shows that with small addition in hardware (around 2% increase in storage size), scalability of the system is near to ideal and execution time of several simple kernels is improved (execution time is shortened between 3%-58%), while avoiding the burden of specific programming models.

The second case study is an implementation of DTA support in the SARC architecture where it interacts with other

architectural components designed from scratch in order to address the problem of scalability.

The third case study presents a DMA-

based prefetching mechanism that complements the DTAs' preload mechanism in order to achieve non-blocking accesses to global data stored in main

memory. It is shown that this mechanism can greatly improve the execution time for several simple kernels (e.g. 13x in the case of matrix multiply).

Multithreaded Dataflow in Tiled Architectures

by **Zdravko Popovic**
(popovic@dii.unisi.it)

Advisor: Prof. Roberto Giorgi
University of Siena, Italy
September 2009

Multi and many core architectures are now widely used. Tiled architectures are easing design of the multi core architectures since they just replicate smaller tiles on a chip. The exploitation of multi-core systems for parallel processing led to the reviving of the dataflow paradigm, this time dataflow at the thread level.

In this thesis, the evolution of an architectural solution that employs both of these concepts, tiling and dataflow multithreading, will be presented. The architecture is called Decoupled Threaded Architecture (DTA). It clusters resources and uses a hardware scheduler to efficiently distribute threads among processing elements in order to achieve good scalability and performance.

Various tests will be presented in order to demonstrate that the architecture scales and performs well for applica-

tions with enough thread level parallelism. The applicability of DTA for real world problems is shown with the parallelization of a widely used video coding standard (H.264 de-blocking filter). At the end, a compilation tool chain that produces parallel code for this and other architectures that share dataflow multithreading concepts is studied.

The DTA architecture is proposed as a solution for some future multicore systems aimed to efficiently exploit thread level parallelism.

High-performance Visual Stimulation System for use in Neuroscience Experiments with the Blowfly

by **Mario A. Gazziro**
(mariogazziro@gmail.com)

Advisor: Jan Frans Willem Slaets
(University of Sao Paulo, Brazil)
and João Manuel Paiva Cardoso
(University of Porto, Portugal)
August 2009

This work describes the development of a system for generating visual stimuli to be used in experiments to

stimulate the vision of invertebrates, by reading signals from the H1 neuron of the fly. The developed system makes use of reconfigurable hardware technology (FPGA), generating images of 640x480 pixels with 256 levels of intensity at a rate of 200 frames per second for conventional tube monitors. These images are dynamically displaced horizontally in order to generate the desired stimuli.

We have developed a new architecture to integrate video memory with the scanning system, where a bigger image is sampled at two rates. The two images generated are showing separated frames in time, according with the stimulus to be presented. This generates a visual effect more sensitive to displacement that is very useful for experiments in neuroscience vision.

The TFLUX Platform: A Portable Platform for Data-driven Multithreading on Commodity Multiprocessor System

by **Kyriakos Stavrou**
(email.tsik@gmail.com)

Advisor: Paraskevas Evripidou
and Pedro Trancoso
University of Cyprus, Cyprus
June 2009

This work presents the TFlux (Thread Flux) Parallel Processing Platform, a complete system that offers an efficient dataflow-like thread-based model

of execution, namely Data-Driven Multithreading (DDM), to its users using commodity components (i.e. unmodified operating system, compiler and ISA hardware) making it applicable to off-the-shelf systems. TFlux provides a complete solution from the programming toolchain to the hardware implementation.

The abstraction layer TFlux exports to its users hides all the details of the

underlying machine allowing different hardware configurations to support its model of execution transparently to the programmer.

The user of TFlux can develop applications using a set of simple but powerful compiler directives. Then the TFlux-C-Preprocessor converts this code to an ANSI C program that includes runtime support for TFlux and all calls to the

system's scheduler. This code can be compiled with a commodity C compiler resulting in a binary that is executable by any commodity operating system and processor. The layered design of TFlux has been tested on different Unix-based multiprocessor systems. Moreover, this design enabled the porting of TFlux to different machines with minimum effort.

In this work, two TFlux implementations are presented: TFluxHard and TFluxSoft. For TFluxHard the Thread Scheduler is implemented as a hardware unit whereas for TFluxSoft, the Scheduler's functionality is provided at the software level. As such, TFluxHard is applicable

to systems that offer the ability to augment the machine with a small hardware module while TFluxSoft is directly applicable to any existing, off-the-shelf system.

For the applications of the evaluation suite, TFlux implementations show remarkable speedup and scalability. Although for most applications the performance of the two implementations is close, TFluxHard shows an advantage over TFluxSoft arising from offloading the Scheduler's functionality to the hardware module. In addition, the experimental results show that both implementations of TFlux are able to exploit more parallelism for

applications with complex dependency graphs, compared to traditional parallel programming model approaches.

Overall, TFlux is a platform characterized by four key components: (1) it can be programmed using a specially developed tool chain; (2) it virtualizes the details of the underlying machine which allows the applications to run on different TFlux implementations without any modification; (3) it is easily portable to systems that differ significantly compared to the original design and (4) it delivers high performance through its dataflow-like Thread scheduling scheme.

On the Road towards Robust and Ultra Low Energy CMOS Digital Circuits Using Sub/Near Threshold Power Supply

By Yu Pu (Y.Pu@tue.nl)

Advisor: Prof.dr. Jose Pineda de Gyvez and Prof.dr. Henk Corporaal
TU Eindhoven, The Netherlands
September 2009

This thesis presents our research work in design of robust near/sub-threshold CMOS digital circuits. While previous research uses ultra-low voltage operation only for low-throughput applications, we achieve medium through-

put using architectural-level parallelism. Several physical-level techniques are also proposed to mitigate yield loss due to process variations, such as balancing V_T of n/pMOS transistors, using V_T mismatch between parallel transistors to improve driving capability, selecting and modifying standard cells, etc. These ideas are demonstrated using SubJPEG, a state-of-the-art 65nm CMOS standard V_T JPEG co-processor. In the sub-threshold,

each DCT and Quantization engine dissipates only 0.75pJ per cycle with a 0.4V supply at 2.5MHz frequency, which leads to 8.3X energy reduction compared to using the 1.2V nominal supply. In the near-threshold, each engine dissipates only 1.0pJ per cycle with a 0.45V supply at 4.5MHz frequency, but the system throughput still meets the VGA standard requirement for 15 fps 640x480 pixel.

A Study of Spilling and Coalescing in Register Allocation as Two Separate Phases

by Florent Bouchez
(florent.bouchez@gmail.com)

Advisor:
Alain Darte and Fabrice Rastello
Université de Lyon, France
April 2009

The goal of register allocation is to assign the variables of a program to the registers or to spill them to memory whenever there are no registers left. The latter should be kept minimal since access to memory is much slower than to registers. In 1981 Chaitin et al. modeled register allocation as an interference graph coloring problem, which

they proved NP-complete. So, there is no exact way in this model to tell whether some spilling is necessary or not, and if it is, what to spill and where.

Recently (2004), three teams discovered that the interference graph of a program under Static Single Assignment (SSA) is chordal. Hence, coloring the graph becomes easy using a simple elimination scheme. Our hopes were that the spilling and coalescing might also get easier to solve, as we now have an exact coloring test.

Our first goal was to better understand from where the complexity of register allocation comes, and why SSA seems to simplify the problem. We came back to the original proof of Chaitin et al., finding that the difficulty comes from the presence of (critical) edges and the possibility to perform permutations of colors. We studied the spill problem under SSA and several versions of the coalescing problem. The general cases were proven NP-complete but we found one polynomial result: incremental coalescing for programs under SSA. We used it to design new heuristics to

better solve the coalescing problem, so that aggressive splitting can be used beforehand.

This coalescing performs well in an aggressive compiler. However, the high number of splits and the increased compilation time required is prohibitive for just-in-time (JIT) compilation. So,

we devised a heuristic, called "permutation motion," that is intended to be used with SSA-based splitting in place of our more aggressive coalescing in a JIT context.

All those results led us to promote a better register allocation scheme. While previous solutions gave miti-

gated results, our better coalescing allowed us to cleanly separate register allocation into two independent phases: First, spilling to reduce register, possibly by splitting a lot; Then color the variables and perform coalescing to remove most of the added copies.

Upcoming Events

22nd International Conference for High Performance Computing, Networking, Storage and Analysis (SC'2009)

November 14–20, 2009, Portland, USA, <http://staff.science.uva.nl/~delaat/sc09/>



16th IEEE International Conference on High Performance Computing (HiPC'2009)

December 16 – 19, 2009 Kochi(Cochin), India, <http://www.hipc.org/>



Asia and South Pacific Design Automation Conference 2010 (ASP-DAC 2010)

January 18-21, 2010, Taipei, Taiwan, <http://www.asp-dac.itri.org.tw/aspdac2010/index.html>



5th International Conference on High Performance and Embedded Architectures and Compilers (HiPEAC 2010)

January 25-27, 2010, Pisa, Italy, <http://www.hipeac.net/conference>



The Design, Automation and Test in Europe conference (DATE'10)

March 8-12, 2010, Dresden, Germany, <http://www.date-conference.com/>



International Conference on Compiler Construction (CC 2010)

March 20-28, 2010, Paphos, Cyprus, <http://www.cs.ucr.edu/~gupta/CC%202010.htm>



8th Annual IEEE/ACM International Symposium on Code Generation and Optimization (CGO 2010)

April 24-28, 2010, Toronto, Ontario, Canada, <http://www.cgo.org>



47th Design Automation Conference (DAC 2010)

June 14 - 18, 2010, Anaheim, CA, USA, <http://www.dac.com/>



37th Annual International Symposium on Computer Architecture, 2010 (ISCA 2010)

June 19-23, 2010, Saint-Malo, France, <http://isca2010.inria.fr/>



Contributions

If you are a HiPEAC member and would like to contribute to future HiPEAC newsletters, please contact Rainer Leupers at leupers@iss.rwth-aachen.de