SEVENTH FRAMEWORK PROGRAMME

HiPEAC info 26

COMPILATION ARCHITECTURE

appears quarterly | April 2011

Network of Excellence on High Performance and Embedded Architecture and Compilation

Welcome to the HiPEAC Spring Computing Systems Week in Chamonix, France, 6-8 April 2011

www.HiPEAC.net

ACACES 2011, 10–16 July 2011, Fiuggi, Italy

Message from the HiPEAC Coordinator

Koen De Bosschere

Dear Friends,
The world news of the last quarter was dominated by a succession of events in the Middle East and North Africa. The events caught my attention for two reasons. First of all, we have a number of well-respected HiPEAC members who were born there and who still have family and friends in that region. It brings these terrible events really close to us. The second reason is that the Internet and other forms of information technology appear to have played a major role in the events. The global information sharing increasingly empowers individuals to influence the course of history. Whereas in Poland, the support of a strong union was needed to bring change over the course of almost 10 years, today an individual with a Facebook or Twitter account can start a popular movement that leads to change in just a couple of weeks. Clearly: information + communication = power, not only for governments, but also for individuals. It warns us that the world is changing continuously, and that the certainties of yesterday are the uncertainties of tomorrow, for all of us.

At the end of January, HiPEAC had its traditional semi-annual review. The conclusion was that the project is achieving good results. The reviewers appreciated the increased industrial focus in the network, the investments in organizing and preparing the European computing systems community for the upcoming research challenges, and the increase in the number of European contributions in international conferences. They want us to further stimulate research interactions and mobility in and beyond our community, to continue our efforts to attract industrial interest in our research, and to continue to reach out to the new member states. They also want us to seriously think about how to grow HiPEAC beyond its current size and impact.

I was happy to learn that the January call for projects has lead to about 60 project submissions, many of them involving HiPEAC members as partners. I hope that a large majority of these projects will make the selection threshold in order to allow them to compete for funding. I also hope that the projects that will eventually not be funded, will at least lead to future collaborations, as every above threshold project deserves to be realized, in one way or another. The HiPEAC network will be ready to help where it can.

In the meantime, we start preparing the 2013 call. The first step is to update and extend the HiPEAC vision document. I hereby invite the whole HiPEAC community to contribute to this effort. In parallel the European commission will organize a number of consultations on specialized topics like how to reduce the administrative overhead in FP7 projects, how to increase industrial uptake of FP7 project results, and how to build the future data centers. If you have a strong opinion on how to strengthen the European computing systems community – scientifically, organizationally, politically - feel free to contact me.

Take care
Koen De Bosschere ■

Per Stenström, a New Member of the Academia Europaea

Per Stenström, a professor of computer engineering at Chalmers University of Technology, has become a new member of the Academia Europaea. Stenström has been active in the HiPEAC Network of Excellence from the very beginning in which he has been particularly interested in making the HiPEAC conference an attractive annual networking event. As a long-term contributor to multiprocessor architecture he is also a key driving force in the multi-core architecture cluster. He says that this recognition is a great honor for him and encourages him to continue to work hard for science and its important role for a sustainable society.

The Academia Europaea was founded in 1988. It's an organisation of eminent, individual scholars from across the continent of Europe. Its members cover the full range of academic disciplines that comprises the humanities, social, physics and life sciences as well as mathematics, engineering and medicine. Currently there are over 2100 members.

Further information:
http://www.acadeuro.org/ ■

Per Stenström
Chalmers University of Technology, Sweden

Panos Tsarchopoulos
Panagiotis.Tsarchopoulos@ec.europa.eu

Message from the Project Officer

The European Commission has launched a consultation on major improvements to EU research and innovation funding to make participation easier, increase scientific and economic impact and improve value for money. The proposed "Common Strategic Framework", set out in a Green Paper, would cover the current Framework Programme for Research (FP7), the Competitiveness and Innovation Framework Programme (CIP) and the European Institute of Innovation and Technology (EIT). This will create a coherent set of instruments, along the whole "innovation chain" starting from basic research, culminating in bringing innovative products and services to market, and also supporting non-technological innovation, for example in design and marketing. The Commission's Green Paper also provides the basis for far-reaching simplification of procedures and rules. The changes aim to maximise the contribution of EU research and innovation funding to the Innovation Union and the Europe 2020 Strategy.

The proposed "Common Strategic Framework" combines three key aspects.

First, a clear focus on three mutually reinforcing objectives: giving the EU a world-leading science base; boosting competitiveness across the board; and tackling grand challenges such as climate change, resource efficiency, energy and food security, health and an ageing population.

Second, making EU funding more attractive and easier to access for participants, for example through a single entry point with common IT tools or a one-stop shop for providing advice and support to participants throughout the funding process. Furthermore, the Common Strategic Framework will allow a simpler and more streamlined set of funding instruments covering the full innovation chain, including basic research, applied research, collaboration between academia and industry and firm-level innovation. Flexibility will be promoted to encourage diversity and business involvement. Applicants should be able to apply for several different projects without repeatedly providing the same information.

Third, there will be much simpler and more consistent procedures for accounting for the use of the funds received. This may involve, for example, greater use of lump sum payments. Greater simplicity will make financial control of EU taxpayers' money easier and more effective.

Other ideas in the Green Paper include: further steps to pool Member States' national research funding; better links with cohesion funding; using EU funding to stimulate public procurement; more use of prizes; further strengthening the role of the European Research Council and of financial instruments such as the Risk-Sharing Finance Facility (RSFF) and the loan guarantee and venture capital investments; and drawing up a set of performance indicators to measure the success of EU research and innovation funding.

You can find more information and the online consultation on the Green Paper at http://ec.europa.eu/research/csfri/index_en.cfm. If you want to give your opinion, the consultation is open until 20 May 2011.

Panos Tsarchopoulos

HiPEAC News

Mateo Valero Awarded Honorary Doctoral Degree by the University of Zaragoza, Spain

The Spanish University of Zaragoza (www.unizar.es) conferred the honorary doctoral degree to Professor Mateo Valero (http://personals.ac.upc.edu/mateo/), director of the Barcelona Supercomputing Center (www.bsc.es) in Spain, on 11th February.

Mateo Valero was awarded as "an example of the necessary collaboration between science and society" and because "his ideas have enabled the development of processors and compilers for HPC in the most important companies of the sector".

Professor Valero, coordinator of HiPEAC 1, has been honoured with several awards. Among them, the Eckert-Mauchly Award, by the IEEE and the ACM, and "Hall of the Fame" selected him as one of the 25 most influents European IT researchers during the period 1983-2008. Mateo Valero, was also named Honorary Doctor by the University of Veracruzana, in Mexico; the University of Las Palmas de Gran Canaria, in Spain; the University of Belgrade; and by the Chalmers University of Technology in Sweden.

Mini-Sabbatical - Prof. José Miguel-Alonso

The Intelligent Systems Group (ISG) at the School of Computer Science of the University of the Basque Country UPV/EHU, to which I belong, has maintained for many years an excellent collaboration with the University of Manchester. Dr. M. Luján, member of the Advanced Processor Technologies (APT) Group at Manchester and alumni from my school, has intervened in several such joint research activities. The HiPEAC funded my mini-sabbatical program which allowed me to visit Manchester from October to December 2010. The visit was initiated with an invitation by Professor I. Watson (one of the leaders of the APT group and a HiPEAC member), and it was important to consolidate our joint activities.

During my visit in Manchester, I was able to get involved in several of the many research projects currently ongoing within the APT group. As one of my main lines of research is in the area of interconnection networks for massively parallel computers, I was glad to attend the meetings of the SpiNNaker project and to collaborate with Professor S. Furber and his colleagues in the joint preparation of two papers: one on the fault tolerance capabilities of this spiking neural network architecture, and another one on modelling the workloads that the SpiNNaker network will support, in order to assess its performance.

I enjoyed participating in the meetings of the UE-funded Teraflux project led by Prof. I. Watson at Manchester. This was an excellent opportunity to learn about innovative proposals for designing and programming future parallel computers. I also had the opportunity to contribute to some of the discussions regarding workload models and interconnection infrastructures. Loosely related with this activity, I spent some time with Dr. Luján studying the scalability of barrier implementations on multi-core systems. A research paper on the topic has been prepared, to be submitted to a forthcoming conference.

Part of our research at the ISG is related to the efficient implementation of machine learning techniques, and the utilization of these techniques in different aspects of computer systems. We share these interests with the Machine Learning and Optimisation group at Manchester. By invitation of Dr. G. Brown, I had the opportunity to talk about our research in a presentation entitled "Parallel Estimation of Distribution Algorithms".

In addition to providing a great working atmosphere, the members of the APT group made my visit a very enjoyable experience. I am very grateful to them and to the HiPEAC network to enable this fruitful collaboration that will continue undoubtedly with more joint work and publications.

Prof. José Miguel-Alonso
University of the Basque Country UPV/EHU, Spain

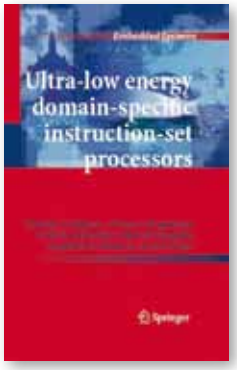
Book on Ultra-Low Energy Domain-Specific Instruction-Set Processors

By Francky Catthoor, Praveen Raghavan, Andy Lambrechts, Murali Jayapala, Angeliki Kritikakou, Javed Absar

Modern consumers carry many electronic devices, like a mobile phone, digital camera, GPS, PDA and an MP3 player. The functionality of each of these devices has gone through an important evolution over recent years, with a steep increase in both the number of features and in the quality of the services that they provide. However, providing the required compute power to support (an uncompromised combination of) all this functionality is non-trivial.

The design of processors to meet the demanding requirements for future mobile devices requires the optimization of the embedded system in general and of the embedded processors in particular, as they should strike the correct balance between flexibility, energy efficiency and performance. In general a designer tries to minimize energy consumption (as far as needed) for a given performance, with a sufficient flexibility. However, achieving this goal is already complex when looking at the processor in isolation; but, really, the processor is just a single component in a more complex system. In order to design such a complex system successfully, critical decisions during the design of each individual component should take into account effect on the other parts, with a clear goal to move to a global Pareto optimum in the complete multi-dimensional exploration space.

In the complex, global design of battery-operated embedded systems, the focus of the book "Ultra low energy domain-specific instruction-set processors" is on the energy-aware architecture exploration of domain-specific instruction-set processors and the co-optimization of the data-path architecture, foreground memory, and instruction memory organization with a link to the required mapping techniques or compiler steps at the



early stages of the design. By performing an extensive energy breakdown experiment for a complete embedded platform, both energy and performance bottlenecks have been identified, together with the important relations between the different components. Based on this knowledge, architecture extensions are proposed for all the bottlenecks. The book also puts together these architecture extensions to present a case study of a processor with all the low power features. We demonstrate these architectural extensions for a case study on a biomedical imaging application and processor instance that the proposed extensions can lead to very high-energy efficiency of up to 1000 MOPS/mW. ■

HiPEAC Activity

PEPPHER Workshop at HiPEAC'11

To foster exchange of ideas and experience in our community, the FP7 project PEPPHER (for PErformance Portability and Programmability of Heterogeneous many-core aRchitectures) was happy to use the opportunity to organize a workshop in conjunction with the 6th International Conference on High-Performance and Embedded Architectures and Compilers (HiPEAC'11), in Heraklion, Crete, Greece.

The workshop took place on January 22nd in the workshop weekend before the main conference. Workshop speakers were invited by the organizers based on recent and relevant contributions to heterogeneous many-core programmability and performance portability. The workshop consisted of a keynote, a PEPPHER overview, PEPPHER-specific presentations by consortium members, related projects presentations, and a panel discussion. The workshop was overall well-attended with about 30+ attendees, and many HiPEAC members seemed to find the way to attend some of the sessions.

Session 1, chaired by Sabri Pllana (University of Vienna), featured the keynote on "Automatic Performance Tuning and Machine Learning" by Markus Püschel (ETH Zürich), and Siegfried Benkner (University of Vienna) presented "PEPPHER Vision & Overview".

Bev Bachmayer (Intel) chaired Session 2 that included three talks from young scientists involved in PEPPHER. Martin Wimmer (University of Vienna) presented his research on "Work-stealing for Mixed-mode Parallelism by Deterministic Team-building". A talk on "Data Structures in Work-Stealing" was given by Daniel Cederman (Chalmers University). Cedric Augonnet (INRIA) presented "StarPU: A Unified Runtime System for Heterogeneous Multicore Architectures".

Session 3 was chaired by Koen De Bosschere (Ghent University). This session highlighted several relevant projects that address PEPPHER-related topics. Kunle Olukotun (Stanford University) presented his work on "Taming Heterogeneous Parallelism with Domain Specific Languages". A presentation on "Incremental Migration of C and Fortran Applications to GPGPU using HMPP" was given by François Bodin (CAPS entreprise). Ben Juurlink (TU Berlin) presented "The ENCORE Project - Enabling Technologies for a Programmable Many-core".

Jesper Larsson Träff (University of Vienna) moderated the workshop panel. Panelists François Bodin (CAPS entreprise), Ben Juurlink (TU Berlin), Christoph Kessler (Linköping University), and Kunle Olukotun (Stanford University) discussed "on the convergence/standardization / future of directive/annotation-based languages for programming heterogeneous multi/many-core systems".

Keynote speech of Markus Püschel on autotuning

Presentation slides and photos of participants are available online at www.peppher.eu/hipeac11.

The PEPPHER consortium is grateful for the generous support and invaluable advice for workshop organization to the HiPEAC NoE, in particular workshops chairs Eduard Ayguade (UPC/BSC) and Dimitris Nikolopoulos (FORTH-ICS/U.Crete).

The PEPPHER project (contract no. 248481) is part of the portfolio of DG INFSO G.3 Embedded Systems and Control Unit. Further information about the project can be found at www.peppher.eu.

Sabri Pllana
PEPPHER workshop co-organizer ■

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

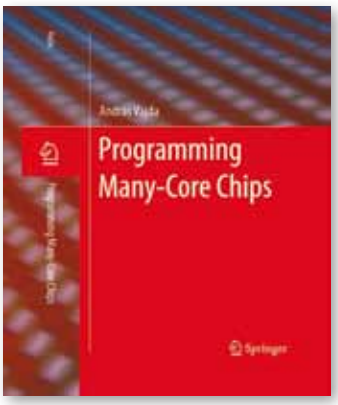
Joint Collaboration to Analyze Multicore Architectures for Space Applications

The European Space Agency (www.esa.int) and the Barcelona Supercomputing Center (www.bsc.es) have started a collaboration project to analyse multicore processors in the space domain. The "Multicore OS benchmark" Project (ESA/ESTEC contract #4000102623), which started in February 2011, focuses on the design of a set of benchmarks that allow to analyse the suitability of current multicore processors for space environments.

A group of five researchers from the CAOS group at BSC and experts of the ESA will pay special attention to the Next Generation General purpose Microprocessor (NGMP, http://microelectronics.esa.int/ngmp/ngmp.htm) currently in development by AEROFLEX/Gaisler (ESA/ESTEC contract #22279/09/NL/JK) for ESA. BSC will develop a set of benchmarks that can be used to test the suitability of this and other processors to the space domain, especially in their usage in satellites.

"This project will enable us to evaluate how current multicore processors can deal with the requirements of space applications in terms of performance and predictable performance. The BSC will apply its past experience in the real-time embedded domain to this project", says Francisco Cazorla, BSC Operating System Group Manager and leader of this collaboration. ■

Book on Programming Many-core Chips



By András Vajda, with contributions from Diarmuid Corcoran and Mats Brorsson

Based on current technology trends, in the near future programmers will have to program chips with hundreds or even thousands of processor cores, called many-core chips. Given the scale of parallelism inherent to these chips, software designers face new challenges in terms of operating systems, middleware and applications. This book will serve as a single-source reference to the state-of-the-art in the research and practical programming of many-core chips.

The book is structured in four main parts. In the first part we survey the state of the art in multi-core HW architecture and we evaluate different technologies related to instruction set architectures, cache and shared memory structure and scalability, on-chip interconnects, power design, but also novel ideas such as 3D stacking and innovative usage of frequency and voltage scaling.

The second part introduces currently available technologies for programming multi-core processors, covering all layers from operating systems to the application programming models. It lays the foundation for part three, through the detailed discussions of the concepts related to parallel programming in general. Besides introducing the terminology that we will use in the third part of the book, we also take a critical look at the bottlenecks and limitations with current solutions, thus setting the frame for the ideas presented in the third part.

The third part of the book targets novel approaches to programming true many-core chips, through the complete software stack. It describes the principles that shall underpin the architecture of operating systems for many-core chips: space-shared scheduling, support for heterogeneity, power awareness and the role of virtualization, illustrated

through some of the most promising research operating systems emerging from the OS research community. Then we explain and compare key concepts in the design of software for massively parallel systems, such as shared memory versus message passing approaches, data versus computation movement, as well as several emerging techniques. We explore the most promising programming models for many-core processors, focusing on scalability, such as the task-based model and the actor model. Finally, we survey and compare the currently available programming frameworks, such as OpenMP, Threading Building Blocks, the Erlang language as well as many other libraries and programming languages.
The book is available for pre-order from Amazon worldwide and is scheduled to be released in July.

HiPEAC Activity

Mini-Sabbatical - Benjamín Sahelices

From left to right Prof. Josep Torrellas, Benjamín Sahelices and Xuehai Qian

I received my MS degree in Computer Science in 1991 and a PhD in Computer Science in 1998, both from the University of Valladolid in Spain. I'm Associate Professor in the Computer Science Department, also at the University of Valladolid in Spain. Since year 2000 I have had a non-stop research collaboration with Professor J.M. Llabería from Polytechnic University of Cataluña and with Professor Víctor Viñals from the University of Zaragoza in different research projects in the Computer Architecture Group of Zaragoza (gaZ). Our main research interests include multicore architectures, cache memory hierarchies and synchronization operations.

Thanks to a mini-sabbatical from HiPEAC I spend four months working in the I ACOMA group led by Professor J. Torrellas in the University of Illinois at Urbana Champaign (http://iacoma.cs.uiuc.edu), as a Postdoctoral Research Associate. The I-ACOMA group research activity focuses on new processor, memory, and system technologies and organizations to build novel multiprocessor computer architectures. Their goal is to propose new designs for high performance multiprocessor computers focusing on programmability, performance/cost ratio and energy efficiency.
The main effort in the I-ACOMA group is the Bulk Multicore architecture. This architecture focuses on easing the task of programming high performance multiprocessors. This is done with an architecture that relieves the programmer from managing low-level task, and with an architecture that helps minimize the chance of parallel programming errors. The Bulk Multicore architecture is based on two key ideas. First, the hardware executes all software as a series of atomic blocks of thousand of dynamic instructions called Chunks. Chunks put no restriction on the programming language. They can either be invisible to the software or actually be generated by the compiler. Second, the use of Hardware Address Signatures as a low-overhead mechanism to ensure atomic and isolated execution of chunks and that also maintains hardware cache coherence.

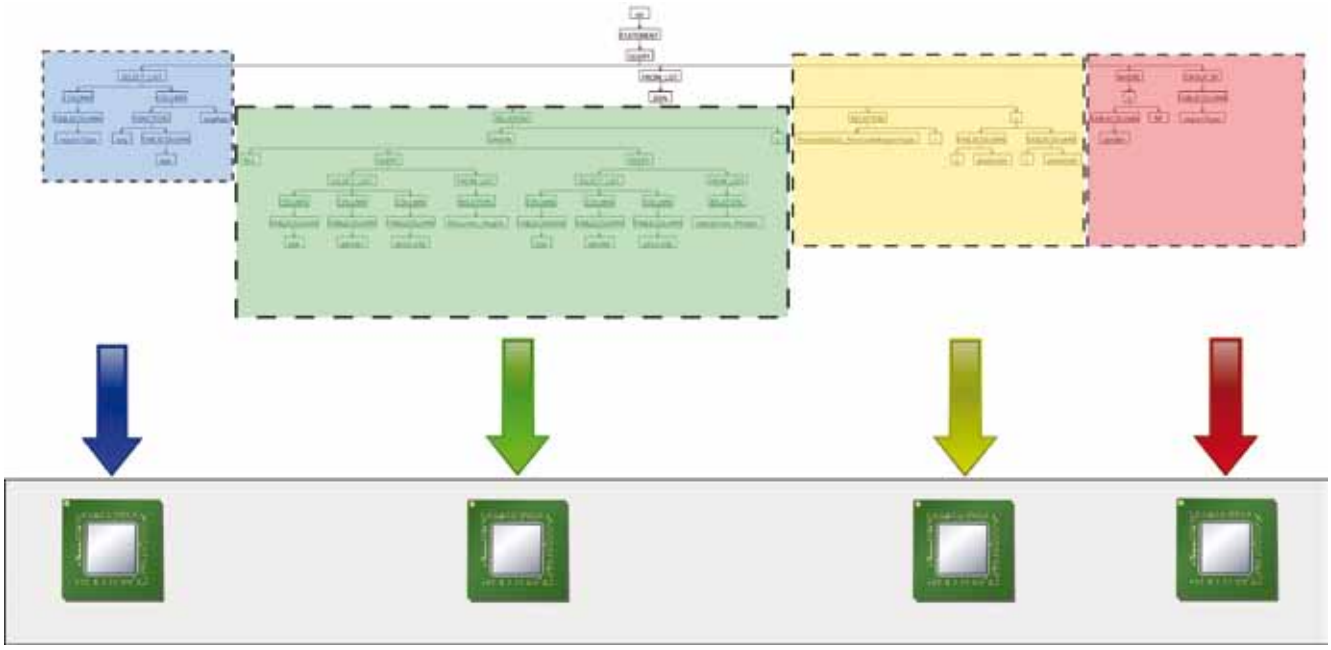
The Bulk Multicore architecture eliminates one of the traditional tenets of processor architecture, namely the need to commit instructions in order, providing the architectural state of the processor after every single instruction. This apparent need contributes to the complexity of current designs because memory-system accesses take many cycles, and multiple loads and stores from both the same and different processors overlap their executions. In the Bulk Multicore, the execution mode of a processor is to commit Chunks of instructions at a time. This mode actually improves performance because the processor hardware is allowed to reorder aggressively and overlap the memory accesses of a program within chunks without the risk of breaking their expected behavior in a multiprocessor environment. Moreover, if the compiler generates the chunks, it can apply aggressive code optimization inside chunks.

While in the I-ACOMA group, I have been working with graduate student Xuehai Qian and with Professor Torrellas (see photo) in chunk execution designs for Bulk Multicore architecture. Both from the personal and professional standpoints this collaboration has been very fruitful. I hope that some publications came soon and, the most importantly, that the basis of future collaborations in high performance multiprocessor research projects are set.

Benjamín Sahelices
University of Valladolid, Spain

Stefano Crespi-Reghizzi Receives a Google Research Award

Parallel Parsing on Multicore and Handheld Devices



Suggested mapping of a syntax tree on several processors.

The team: Matteo Pradella (associate professor), Alessandro Barenghi (post-doc), and master students Valerio Ponte and Ermes Viviani

The objective is to develop practical data-parallel parsers that exploit parallel processors on multicore and handled devices effectively , as well as operate within a prescribed power envelope. Similar to Bison, the parser will be automatically generated from a grammar specification. Measurements by benchmarks will compare: parallel parsing algorithm versus the sequential parser on a single CPU or on a multicore. If successful, the parser could become a standard open source component of browsers on forthcoming battery-operated devices.

Parsing is ubiquitous in browsing, compilation, anti-virus applications, NLP, genome sequencing, etc. Classical algorithms (e.g., in Crespi-Reghizzi, "Formal languages and compilation", Springer, 2009) are not easy to split into balanced, loosely synchronized threads. Though less important for compilation, this is critical for browsing, for semistructured data searching, and for NLP. Chip manufacturers forecast that on future battery-operated and handheld devices, single CPUs will be simpler and slower and speed must come from parallel execution; for more motivation see also UC Berkeley Par Lab Tech. Rept. "Browsing Web 3.0 on 3.0 Watts: why web browsers will be parallel". Unfortunately, applying data-parallel decomposition (in particular map-reduce) to a text is difficult because a, say, LR(k) parser starts in a state that depends on the parsed prefix, thus imposing serialization. Parallel parsing algorithms have been theoretically investigated in the past, but we have found little of practical value, since the costs incurred by synchronization are neglected in most studies. More to the point, recent practical efforts, such as (B. Shah et al, "A data-parallel algorithm for XML DOM parsing", Proc. 6th Int. Symp. on DB and XML Techn., 2009), identify certain tokens as handles that force the parser into a known state. This approach is rather ad hoc, and recent results are not too encouraging.

We want to split the text into balanced chunks, to parse them by independent threads, and to combine partial parses into the final tree. For this to be possible, we need formal properties that are not available for grammars, not even of the LR(k) class: closures under concatenation, star, prefixing and suffixing, and maybe Boolean operations. On the other hand, these properties are present in certain new studied grammars, the best known case being the Visibly Push Down languages (Alur and Madhusudan), which mimic XML parenthesis structure but are too naive for real application, say, in NLP. Inspired by them we discovered (Crespi-Reghizzi and Mandrioli, "Operator Precedence and the Visibly Pushdown Property", LATA 2010) that the needed properties hold for Floyd's Operator Precedence Grammars, extensively used in early compilers and still popular for fast parsing of expressions: they are our candidate for data-parallel algorithms. Two types of target platforms are considered: an x86-based multicore and a smaller architecture based on ARM or simpler cores.

Stefano Crespi-Reghizzi
Politecnico di Milano, Italy

ARTEMIS RECOMP Project: Reduced Certification Costs Using Trusted Multi-core Platforms

Coordinator:
Mäkitalo Jarkko (Jarkko.Makitalo@kone.com), Kone (FI)
Website: www.recomp.eu
Duration: 3 years
Start date: 01/04/2010
Total Costs: 25.8 M€

RECOMP seminar in Munich (Oct 2010)

The RECOMP (Reduced certification cost for trusted multi-core platforms) research project aims to establish methods, tools and platforms for enabling cost-efficient re-certification of safety-critical systems and mixed-criticality systems. Applications addressed are automotive, aerospace, industrial control systems, lifts and transportation systems.

RECOMP recognizes the fact that the increasing processing power of embedded systems is mainly provided by increasing the number of processing cores. The increased numbers of cores is commonly regarded as a design challenge in the safety-critical area, as there are no established approaches to achieve certification. At the same time there is an increased need for flexibility in the products in the safety-critical market. This need for flexibility puts new requirements on the customization and the upgradability of both the non-safety and safety-critical critical part. The difficulty with this is the large cost in both effort and money of the re-certification of the modified software.

RECOMP, started in April 2010, will provide reference designs and platform architectures together with the required design methods and tools for achieving cost-effective certification and re-certification of mixed-criticality, component based, multi-core systems. The aim of RECOMP is to define a European standard reference technology for mixed-criticality multi-core systems supported by the European tool vendors participating in RECOMP.

The RECOMP project will bring clear benefits in terms of cross-domain implementations of mixed-criticality systems in all domains addressed by project participants: automotive systems, aerospace systems, industrial control systems, lifts and transportation systems.

The participating and collaborating partners aim to achieve at least the following advantages:

1. Lower total system cost by being able to use multi-core the overall system cost is reduced as the full application can be run on a single circuit board module that saves mass, volume and power.
2. Lower cost of the certification of first version of the product; by using a certified computing platform the cost of certification will be reduced.
3. Shorter time-to-market as the certified platform will reduce time required for the certification of the application.
4. Lower re-certification costs since non-critical parts of the application can be updated without touching the safety critical parts of the application.
5. Supplier-based development process as the possibility to develop pre-certified components will enable component suppliers to take on a bigger share of the product development share, distributing costs over several customers, and thus lower overall costs of the development process.

The approach will take into account that depending on the industry and application, different procedures can lead to a cost-efficient product, because this always requires a return on investment analysis: e.g. for low-volume products, the cost of certification can possibly be reduced by deliberately over-dimensioning the hardware, which means an increase in unit cost to reduce development cost, while for high-volume products, the unit cost is the driving factor, and a much higher certification cost can be accepted. Thus each system, architecture, and application will have its own "sweet spot" between certification cost and unit cost. RECOMP will put particular emphasis on this topic when defining processes and validating the certification.

Collaboration Grant Report - Rubén Titos

My name is Rubén Titos and I am a PhD student at the University of Murcia, Spain. I started my PhD in 2007 at the Computer Architecture Group under the supervision of Manuel Acacio and José M. García. My research focuses on the design of efficient and scalable hardware transactional memory systems (HTMs) and their implementation on chip-multiprocessor architectures.

Last fall, HiPEAC gave me the opportunity to visit Professor Per Stenström at the Chalmers University of Technology. During my stay in Göteborg, I worked closely with members of his research group, some of whom work in the very same topic that my research focuses upon, i.e. HTM systems. It was a very enriching and fruitful experience, as not only did the constant discussions improve my understanding of some key aspects of HTM design, but they also proved to be breeding grounds for several interesting ideas.

In this collaboration we investigated the benefits and implementation costs of hybrid-policy Hardware Transactional Memory (HTM) systems. In prior research, HTMs have either fixed policies of conflict resolution and data versioning for the entire system, or allowed a degree of flexibility at the level of transactions. We found that making such decision at the time of design results in a system susceptible to pathologies, lower average performance over diverse workload characteristics or high design complexity. We explored a new dimension along which flexibility in policy can be introduced. Recognizing the fact that contention is more a property of data rather than that of an atomic code block, we developed an HTM system that allows selection of versioning and conflict resolution policies at the granularity of cache lines. We discovered that this neat match in granularity with that of the cache coherence protocol results in a design that is very simple and yet able to track closely or exceed the performance of the best performing policy for a given workload. It also brings together the benefits of parallel commits (inherent in traditional eager HTMs) and good optimistic concurrency without deadlock avoidance mechanisms (inherent in traditional lazy HTMs), with little increase in complexity. This work has recently been accepted for publication at the ICS'2011 conference, in a paper titled: "ZEBRA: A Data-Centric, Hybrid-Policy Hardware Transactional Memory Design".

We are now taking this collaboration further, working on a different HTM design that attempts to combine the

Rubén Titos

advantages of both eager and lazy policies of data version management, with minimal changes to the behaviour of standard write buffers. A paper based on early results of our work has been accepted for publication in a workshop held in conjunction with IPDPS'2011, titled "The Impact of Non-Coherent Buffering in Lazy Hardware Transactional Memory Systems".

Thanks to this HiPEAC grant, a strong bond has been created between the research groups at the University of Murcia and Chalmers. I would like to thank HiPEAC for providing such a unique opportunity to work with such brilliant researchers from other European institutions, and to establish long-term collaborations. I strongly encourage other HiPEAC students to take advantage of this program.

Rubén Titos ■

High-level Synthesis Symposium in Ghent

FlexWare

On 1st December 2010, Ghent University hosted the symposium "A new era for high-level synthesis" in the beautiful University Congress Centre "Het Pand." More than 90 people were registered but due to heavy snowfall only a little over 60 attendees actually made it to Ghent. Unfortunately, some of our invited speakers were also stranded at European airports. Luckily, their presentations could be live streamed to the audience through the WebEx system, with the much-appreciated help of the HiPEAC team in Ghent. This way, the participants were still able to enjoy the five invited presentations. After a brief introduction by Dirk Stroobandt (Ghent University) on the FlexWare project (that was also presented on posters during the breaks), Paolo Ienne (EPFL, Switzerland) convinced us of "the irresistible charm of automation." After his live streamed

Internship Report - Richard Membarth at ARM

Generating GPU Code from a High-level Representation for Image Processing Kernels

I am a PhD student in Computer Science at the University of Erlangen-Nuremberg, Germany. My advisor is Professor J. Teich. The focus of my work is on parallelization frameworks for standard shared memory multi-core processors as well as for many-core accelerators like graphics cards for medical imaging. After investigating existing frameworks in order to see what is possible already today, we are moving now along to bridge the gap of these frameworks: to create a high-level framework that will generate optimized low-level code for the target system.

To provide the basis for such an abstract framework, I received last year a HiPEAC Industrial Internship Grant that gave me the opportunity to work with Anton Lokhmotov at ARM's Media Processing Division. During my visit in Cambridge, we created a framework for representing image processing kernels based on decoupled access/execute metadata, which allow the programmer to specify both execution constraints and memory access patterns of a kernel.

The framework performs source-to-source translation of kernels expressed in high-level framework-specific C++ classes into low-level CUDA C or OpenCL code with effective device-dependent optimizations such as global memory padding for avoiding partition conflicts. The source-to-source compiler we developed is based on Clang, an open-source frontend for C-family languages. The Clang front end parses the source files and generates an AST representation of the source code. Our back end uses this AST representation to generate host and device code in CUDA C or OpenCL.

Our first preliminary experiments show that code generated from our abstract description is as fast as hand-written and optimized CUDA C code for vertical mean filtering. While the performance for images that lead to misaligned memory layouts decreases almost by 50%, our compiler pads the memory layout so that almost no performance penalty can be observed. Supporting different backends, our source-to-source compiler produces CUDA C and OpenCL code. We showed that our framework can generate source code for convolution kernels that run on graphics cards as fast as the corresponding GPU implementations in OpenCV, a widely used computer vision library.

Richard Membarth

In the following months, we will continue our work to provide support for automatic border handling and device-dependent optimizations based on decoupled access/execute metadata. This allows programmers to concentrate on developing algorithms and applications, rather than on mapping them to the target hardware. We are looking into releasing the framework under an open source license, so that also others can benefit from our research.

I would like to thank HiPEAC for generously supporting my trip to Britain. It was a great opportunity to establish a collaboration that is continuing after my visit. I would also like to thank Anton and other colleagues and interns at ARM for a very pleasant stay.

Richard Membarth ■

presentation, two speakers that actually made it to Ghent took the floor: Jason Villarreal (Jacquard Computing, Inc., USA) presented some "design challenges in the development of the Riverside Optimizing Compiler for Configurable Computing" and Paul Stravers (VectorFabrics, The Netherlands) tried to provoke the audience with the title "Need parallel programs? Think sequential!" After the break and poster session, it started snowing heavily in Ghent, as if to underline why the two remaining presentations also had to be streamed: Nigel Topham (University of Edinburgh, UK) helped us in overcoming "the inevitable complexity of customization" and Steve Teig (Tabula, USA) gave an entertaining look on his new tech-child Tabula in "Spacetime: a programmable fabric beyond the FPGA." More info and the pdf-files of these presentations can be found on http://flexware.elis.ugent.be/hls-symposium.

Dirk Stroobandt
Ghent University, Belgium ■

HiPEAC'11 Conference Report

During the HiPEAC'11 opening

The sixth edition of the HiPEAC conference took place this January in the beautiful island of Crete, Greece. Over 190 people from the five continents enjoyed the remarkable hospitality and kindness of our Cretan hosts, who made sure everything ran smoothly and without flaws for what turned to be yet again another successful HiPEAC conference. The conference general chairs were Manolis Katevenis from FORTH-ICS and Margaret Martonosi from Princeton University, while this year's program chairs were Christos Kozyrakis from Stanford University and Google, and Olivier Temam from INRIA.

The three-day conference was preceded by a range of workshops in the topics of computer architecture and operating systems co-design, design for reliability, interconnection network architecture, programmability issues for heterogeneous multicores, programmability and portability, methods and tools for rapid simulation and performance evaluation, and reconfigurable computing. Together with tutorials on OpenCL programming and the ILDJIT compilation framework, these events gathered 140 attendees and served as a perfect prelude for the main event.

The invited keynote speakers are two influential researchers in their area. Antonio Gonzalez, director of the Intel Barcelona Research Center and UPC Professor talked on Monday about Moore's Law implications on energy reduction. On Tuesday, Saman Amarasinghe, Professor in the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, presented his work on PetaBricks, a language and compiler based on autotuning.

Conference attendees practicing traditional Cretan dance.

This year's conference again emphasized the convergence of challenges faced by the embedded and high performance computing fields. The 2011 call for papers encouraged the submission of cross-cutting research and innovative ideas that attempt to solve these challenges that have emerged in the last few years. 85 submissions were reviewed by 119 reviewers for a total of 423 reviews, guaranteeing that almost all papers received five reviews, at least three by program committee members and at least one by an external expert. The final 20 selected papers were decided at a Paris meeting by the 27 program committee members, who were able to achieve consensus on several of them.

This year, the Conference featured a best paper award. The program committee co-chairs selected 5 candidates from the final papers; all PC members read these 5 papers, and voted on them –the winner was: "Fast Modeling of Shared Caches in Multicore Systems" by David Eklov, David Black-Schaffer, and Erik Hagersten (Uppsala University). David received the award during the banquet at the beautiful Zacharioudakis Winery.

And of course, no HiPEAC conference would be such if it were not for the social events organized. This year the attendees had the pleasure of going on a guided tour through the impressive Knossos Archaeological Site, followed by a visit to Zacharioudakis Winery placed on a remarkable spot at the top of the hill Orthi Petra, where they learned the process of producing wine by the words of the winery owner. The night concluded with a banquet at the winery, where the attendees were served delicious dishes of traditional Cretan cuisine and entertained by a group of traditional Cretan musics and dancers.

Víctor Garcia
BSC, Spain ■

Collaboration Grant Report - Daniel Sánchez

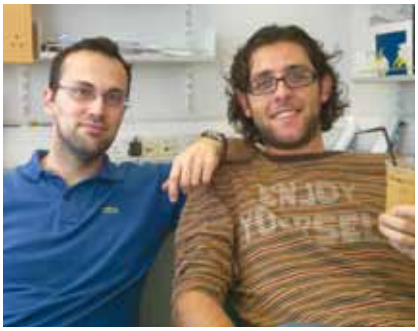
Hi! My name is Daniel Sánchez. I'm a PhD. student in the Computer Engineering Department at the University of Murcia, south-east Spain. Currently, I'm finishing my PhD focused on the study of different techniques to deal with hardware errors.

Among others I was awarded with a HiPEAC 2010 collaboration grant. This offered me the opportunity to take a long trip over the Mediterranean, headed to the small island of Cyprus. There, I had the great pleasure to work with Assistant Professor Yiannakis Sazeides, head of the Computer Architecture Research Group at the Department of Computer Science at the University of Cyprus. I also had the opportunity to meet a lot of great people including Isidoros Sideris, Nikolas Ladas, Bushra Ahsan, Marios Kleanthous, and Lorena Ntrou. During the three months I spent in Cyprus, we studied the impact of errors on memory devices.

The ratio of permanent faults in memory cells increases exponentially within every new scale generation. The reasons for these devices to fail are multiple. Among them, it is worth mentioning fabrication defects and physical wear-out, including problems such as electromigration and oxide breakdown. But the problem we face exacerbates even further due to the use of different techniques to limit power consumption. For example, DVFS (Dynamic Voltage and Frequency Scaling) attempts to reduce the Vcc of structures such as caches. However, voltage scaling in the presence of process variations can cause a large number of circuits to fail.

One mechanism, which is already present in caches and memories, is Error Correction Codes (ECC), which results inexpensive to detect and correct faults in caches. However, ECC affects the performance since, potentially, every access to a faulty block incurs the ECC repair overhead. Furthermore, ECC soft-error capabilities are reduced when some bits protected by the ECC code are already faulty. Thus, ECC may not be the best option to repair permanent or wear-out faults in the cache.

Another alternative is the disabling of faulty portions. Different techniques have been so far proposed with different granularities, from bytes to words or even blocks. Nonetheless, there is a lack of knowledge regarding what the real impact of permanent errors over the performance of the cache is. Previous studies rely on the use of randomly generated fault maps to measure their impact. Because of that, the results they provide are limited to the number of faulty maps used (which is far from the total number of total maps). Hence, it remains unclear whether these previous studies are representative or not.



Isidoros Sideris and Daniel Sánchez

To solve this issue, we have developed an analytical model for the exact calculation of the expected number of cache misses in the presence of random permanent faults. For that purpose, we only require a memory trace, the cache configuration and the probability of a cell to fail. Therefore we avoid the use of fault maps.

It is our belief that this model can help designers to easily evaluate the impact of faults in caches. In addition it provides with a valuable insight for developing new mitigation techniques, which we will carry out as a follow-up to the collaboration between the University of Cyprus and the University of Murcia.

Daniel Sánchez ■

Honeywell International s.r.o.

Honeywell's technology development programmes drive innovation and create new solutions to meet the unique and evolving needs of government agencies and aviation customers. Our dynamic and vibrant aerospace technology enterprise is enabled through dedicated investments in local resources to establish research, technology and manufacturing operations that work in close collaboration with government agencies and aviation customers to identify next-generation requirements and accelerate the production of cutting-edge solutions.

The technology teams at our regional Centres of Excellence design, develop and implement technological advancements that enhance the delivered value of our products, systems and solutions for key aerospace applications:

- Crew Interface and Platform Systems – human centred design studies to enhance safety, understand pilot workload, and create

and assess new user interfaces (HMI) and avionics visualisations.
- Advanced Systems and Prototyping – SMARTlab visualisation, simulation and modeling with architectural studies for avionics, mechanical and electrical systems, and vehicle health management development.
- Communication, Navigation and Surveillance Systems – pioneering research and development work to support Air Traffic Management (ATM) modernisation with research on cooperative and non-cooperative surveillance, Global Navigation Satellite Systems (GNSS)/Inertial Navigation System (INS) communication, 4D trajectory management and self-separation concepts with technical and operational validation, and platform system development.
- Engines and Air Management – development of advanced turbine technologies and application work on our Catalytic Oxidation (CATOX) technologies that neutralise biological and chemical agents.

Our local engineering and research teams additionally provide ongoing innovation for the following aerospace technologies:
- Advanced Sensors and Microsystems – development and application of tactical-grade MEMS gyro- and accelerometer-based inertial sensor technologies.
- Power and Control Components – industry-leading research and development of More Electric Aircraft (MEA) systems and components.
- Aeronautical Communication Systems – concept and development of new information exchange technologies between aircraft and ground.

From ATM modernisation to energy efficiency and environmental mandates, we are active in creating partnerships and joining industry programmes to discover and apply the region's best technologies to energise ongoing aerospace initiatives, while helping aviation customers ensure compliance with current and emerging government standards and legislative requirements.

Honeywell Technology Development provides:
- Customer Focus – our experienced organisation works closely with customers to identify and develop the appropriate technology solutions.
- Operational Performance – Six Sigma plus and other proprietary processes ensure delivery of peak performance and optimal value.
- Advanced Technologies – broad technical capabilities deliver solutions that improve performance, increase reliability, reduce maintenance spend, and lower total cost of ownership.
- Local Resources – regional technology teams and Centres of Excellence present the most advanced technologies and products to the marketplace.
- Global Network – worldwide resources provide customer care, technical support and mechanical services.

Zlatko Petrov
zlatko.petrov@honeywell.com
Honeywell International s.r.o.
Brno Czech Republic

Internship Report - Alexander van Amesfoort at Codeplay, Edinburgh

Large C++ applications that need to run much faster, such as modern computer games in development, must be parallelized and further optimized. Codeplay's Offload technology allows programmers to quickly offload C++ tasks to other cores with private memory (spaces) without intrusive modifications. Additional optimization steps can focus on improving memory bandwidth and locality. C++ applications typically access data through STL data structures. General-purpose hardware/software caching is only really effective if properly developed for. The programmer's knowledge of access patterns and concurrency type can be used to improve data handling, especially on Cell/B.E., GPUs, and mobile platforms. One, or multiple "accessor objects" are wrapped around (parts of) an STL object, and the programmer indicates access pattern type, storage size and policies. The accessor then caches, (pre) fetches, and/or write-combines as indicated possibly using platform specific features. For applications that need more intrusive refactoring, accessors force programmers to start on the inevitable changes providing a clear, efficient access pattern goal.

As a PhD student working on programming multi-cores at Delft University of Technology, I have worked at Codeplay, Edinburgh from mid-October until mid-January.
A large part of this project was devoted to supporting STL classes with Codeplay's Offload, starting with STL vector and map, since this is needed before creating accessors for them. Directly offloading any routine is not always possible. Offload automatically duplicates call graphs, deduces if a pointer refers to a local or remote (SPE or GPU) memory space and generates cross memory space transfers, but rejects dubious cross memory space constructs. It proved undesirable to cleanly extend problematic STL code in separate files, especially for any STL library installed at a user, so Codeplay has to supply an STL with Offload. Electronic Arts STL (EASTL) vector and map have been adapted, but RDESTL is also suitable, as both libraries are

well-written, fast, and designed with computer games in mind, clearly an advantage for Codeplay's business. A few improvements to Offload will allow most STL functions to work without modifications.
The Offload compilation toolchain is complex software that needs an extensive test suite to keep track of what STL functionality works. An extensive test suite has been developed for all functionality of STL vector and map containers, applied to all Offload contexts and for various data types. It generates status reports, and timing information to quickly verify computational complexity promises. These tests have been added to Codeplay's automated testing system. Support for other containers can follow quickly: some modifications to map have already been applied to set, multimap and multiset.
Further, we agreed on accessor class semantics, developed an initial data sorting example, and a variant with asynchronous offloading. This work will be extended in the next intern project at Codeplay, currently underway.
We also developed a small synthetic bandwidth benchmark for Cell/B.E. and concluded that the QuantLib computational finance library is suitable as a large C++ demonstrator. Some progress has been made to get its benchmark working with Offload.

While the work on STL for Offload took too much time to also work-out and demonstrate the gains of various accessor classes, Codeplay staff and I agreed to continue working towards publishing a paper on this topic. We have discussed with the next intern how to combine our (projected) results. In the meantime, I have written a technical report on my experiences. Codeplay is also willing to provide Offload updates to me and to support student projects working with Offload.

I would like to thank Codeplay staff for supporting me throughout this internship and HiPEAC for their academia-industry grant.

Alexander van Amesfoort
a.s.vanamesfoort@tudelft.nl

An important first step for CHA'N'GE

We have to go back to July to find the moment when Hank and I decided to start this adventure of establishing a new workshop called CHA'N'GE (Computing in Heterogeneous, Autonomous 'N' Goal-oriented Environments). Our goal was to open a dialog between researchers in systems and those exploring self-adaptive and autonomic techniques. Our hope is that establishing self-aware techniques for hardware, compilers, operating systems, and system software will help cope with the skyrocketing complexity of modern computing platforms. These ideas have been much on our minds for the last two years, although we didn't understand how much work we were taking on in establishing this workshop. However, after holding the first edition of CHA'N'GE on March 6th (collocated with ASPLOS) we can say all that work was surely worthwhile. Special thanks have to be given to HiPEAC, IEEE, the ASPLOS general chair Rajiv Gupta, and ASPLOS workshop chair Evelyn Duesterwald for their valuable support.

CHA'N'GE speakers

The participation for CHA'N'GE's first year (find out more information at www.change-conference.org/) was quite high with five papers included in the proceedings. The workshop started on a great note with the keynote speech "Liquid Metal, Change you can believe in", given by Rodric Rabbah, from IBM Research. Rodric gave an insightful talk presenting IBM's Liquid Metal project (https://researcher.ibm.com/researcher/view_project.php?id=122). Rodric convincingly demonstrated that enabling introspective and adaptive computing systems requires a comprehensive and interdisciplinary approach including research into programming models, compilation methodologies, runtime systems, and heterogeneous architectures combining the advantages of multicores, GPUs, FPGAs, and other processing accelerators. The remaining talks presented the accepted papers and spanned a wide range of topics related to CHA'N'GE, including operating systems, reconfigurable and adaptable FPGA-based systems, evolutionary computation, mobile systems, dynamic compilation, and multicore architectures.

At the end of this first edition we would like to thank again our technical co-sponsors: HiPEAC and IEEE, just to mention few names, Koen De Bosschere, Hazel Kosky, Allison Mohn, and Sushil K. Prasad, but the list can be definitely longer. Additional thanks go to Kevin Skadron and Sudhanva Gurumurthi editor-in-chief and associate editor-in-chief of Computer Architecture Letters, which is considering the best two papers for CAL submission. Finally, thanks go out to all the TPC members, the reviewers, our supporters (ALTERA, the Progetto Roberto Rocca and iDRESD) and all the people who attended CHANGE at Newport Beach. We hope to see you and many new faces at CHANGE 2012!

Marco Santambrogio
Hank Hoffmann ■

PhD News

Non-Speculative Enhancements for the Scheduling Logic

By Rubén Gran Tejero (rgrantejero@gmail.com)
Advisors Names: Enric Morancho, Àngel Olivé, José María Llabería
UPC, Barcelona, Spain
November 2010

One of the tasks that a hardware designer must face up is the balance of conflicting design goals. This thesis focuses on one of these conflicts of interest between desirable objectives in High-Performance Out-of-Order processors.
In particular, we refer to the difficulty that exists between operating at a high frequency and the ability to expose and exploit instruction level parallelism of the executed code. In Out-of-Order processors, the issue queue and the scheduler are related to exposure and exploitation of instruction level parallelism. One of the design parameters of these two elements is the number of entries (instructions) of issue queue. This parameter is directly related to the amount of parallelism that the processor will be able to expose. On the other hand, another design parameter is the number of instructions that can be issued to execution in parallel, the issue-width. This parameter determines how much instruction level parallelism the processor can exploit. A high-performance processor requires that the values of these parameters to be as high as possible. However, the cost (area and delay) of the issue queue and the scheduler also grows with these parameters, thereby increasing the value of these parameters may compromise the processor cycle time.
In this thesis, we employ two different techniques that can alleviate the conflict between frequency and cost (area and delay) of the scheduling stage: pipelining the scheduling stage and slicing the selection logic. However, these techniques involve some performance losses when applied in a straightforward manner. In this thesis, we have proposed several enhancements to overcome the upcoming problems.

Parallel Algorithms and Architectures for LDPC Decoding

By Gabriel Falcao (gff@deec.uc.pt)
Advisors: Prof. Leonel Sousa and Prof. Vitor Silva
University of Coimbra, Portugal
December 2010

Because Low-Density Parity-Check (LDPC) codes allow working very close to the Shannon limit and achieve excellent BER performance, they have been adopted by modern communication and storage standards. Due to the irregular and computationally intensive nature of LDPC decoders, VLSI architectures able to provide high throughput performance have been proposed over the last 10 years. However, the development of these systems presents important challenges.

This thesis shows that these hardware dedicated systems can be implemented more efficiently with a reduced number of processors supported by a minor reconfiguration of the memory blocks of the system. By using a small number of processors, the routing complexity of the design has been significantly simplified, allowing savings in terms of area and development costs, namely for the particular case of demanding DVB-S2 communications.

While dedicated hardware still imposes some restrictions, the fact that multi-cores have gone mainstream encouraged the development of flexible solutions towards the computation of LDPC decoding. This thesis proposes new approaches for these computationally intensive algorithms, by performing parallel LDPC decoding based on ubiquitous multi-core homogeneous or heterogeneous systems, namely on the Cell/B.E, on GPUs and on general-purpose homogeneous multicores. This thesis extensively addresses the challenges faced in the investigation and development of new and efficient parallel algorithms, with focus mainly given on scalability, efficiency regarding memory accesses and multithread execution, and throughput and BER performance.

Experimental results for programmable approaches show throughputs above 100Mbps and BER curves that compare well with VLSI dedicated hardware for a wider range of possible data-precision representations.

System Level Design Space Exploration for MPSoC: Methods, Algorithms and New Infrastructure

by Jia Li Zai Jian
(cjia@iuma.ulpgc.es)
Advisors: Prof. Antonio Nunez and Dr. Tomás Bautista
Universidad de Las Palmas de Gran Canaria, Spain
January 2011

In the embedded systems domain, system-level design space exploration (DSE) is a key element in the design of complex heterogeneous MP-SoCs. Previous work has proven to be efficient to explore various alternatives for mapping a specific application onto a target MP-SoC architecture. But significant effort is still required to (re-)write scripts that control the evaluation mechanism (analytical model or simulator) during the search through the design space, devoting repetitive processes to build customized scripts and/or architecture models for every different kind of DSE experiment. Automating such a process becomes essential in terms of reusability and flexibility for larger design space explorations in the design of a complex heterogeneous multiprocessor architecture.

In this thesis, we present a new and generic system level MP-SoC DSE infrastructure, called NASA (Non Ad-hoc Search Algorithm). This highly modular framework uses well-defined interfaces to easily integrate different system-level simulation tools as well as different combinations of search strategies in a simple plug-and-play fashion.
CASSE, SESAME, and PISA are example tools that can be plugged to NASA. As a result, the potentials for reuse of the framework are significantly increased since each DSE experiment can be performed without the need of preparing experiment-customized scripts, but it only requires a simple change of the user's input constraint values. Second, we have implemented and integrated a new approach in NASA to gradually and automatically generate simulatable system models that are used for obtaining system metrics to evaluate design decisions. Thus, the entire DSE process (composed of searching, system models generation and design point evaluation) is performed in an automatic and systematic fashion, thereby improving design productivity and decreasing the designer's efforts. Third, NASA deploys a novel dimension-oriented DSE approach in which the design space is explicitly separated into dimensions, which could represent design decisions that are orthogonal to each other such as mapping, architectural components, and platform. The designer can choose to simultaneously explore all dimensions, or to fix one or more of these dimensions (e.g., a fixed platform) and to focus the exploration within one or two dimensions (e.g., mapping exploration only). To this end, designers are allowed to configure the appropriate number of, possibly different, search algorithms to simultaneously coexplore the various design space dimensions. Several DSE experiments are presented in which we, e.g., compare NASA configurations using a
single search algorithm for all design space dimensions to configurations using a separate search algorithm per dimension. These experiments indicate that the latter multi-dimensional coexploration can find better design points and evaluates a higher diversity of design alternatives as compared to the more traditional approach of using a single search algorithm for all dimensions.

Techniques to Reduce Inefficiencies in Hardware Transactional Memory Systems

By M. M. Waliullah
(waliulla@chalmers.se)
Advisor: Prof. Per Stenstrom
Chalmers University of Technology, Sweden
January 2011

The recent trend of multicore CPUs pushes for major changes in software development. Traditional single-threaded applications can no longer get a sustainable performance boost from this new generation of CPUs that consist of multiple processors. Applications must be programmed in a parallel fashion to take advantage of their performance potential. Traditional lock-based parallel programming models are considered to be too difficult and error prone for average programmers. Furthermore, lock-based synchronization is blocking because execution of critical sections are serialized. This serialization would not be needed in case there are no data races.

Transactional memory has been proposed to simplify parallel programming and increase concurrency by using non-blocking synchronization. In transactional memory systems, multiple transactions (e.g., critical section invocations) from different threads can be executed speculatively in parallel. Data integrity, hence the program correctness, is maintained by isolating the speculative execution and committing the end result atomically. Data sharing conflicts between two

transactions restrict only one of them to commit successfully. This thesis deals with inefficiencies of transactional memory that is implemented in hardware (HTM).

In an HTM system that detects conflicts lazily, transactions from one thread can repeatedly squash a transaction from another thread which can lead to a starvation problem for the latter. A novel solution that uses squash counts for individual transactions is proposed to avoid starvation. At a data conflict, HTM systems squash the speculative executions and re-execute transactions from the beginning without considering the fact that the entire execution is not unsafe. The thesis proposes a scheme that smartly takes intermediate checkpoints so that the safe part of the execution is not squashed. To isolate the effects of the speculative execution, a private buffer is used to store the speculative data. The problem of speculative buffer overflow is also addressed and a scheme is proposed that decouples the read set from the speculative buffer to reduce overflows.

Application characteristics play a significant role in performance of HTM policy decisions. To adapt conflict resolution policy to the application behavior a flexible HTM infrastructure is proposed. Finally, conflicts are quantified in different classes to better understand the root causes of HTM inefficiencies and techniques are introduced to reduce conflicts from each of these classes.

Design Methodologies for Improving Embedded Systems with Hardware Accelerators

By Christian Pilato
(pilato@elet.polimi.it)
Advisor: Prof. Fabrizio Ferrandi
Politecnico di Milano, Italy
February 2011

Heterogeneous multiprocessor architectures are the de-facto standard for embedded system design. They are usually composed of several general purpose, digital signal and hardware accelerators, interconnected through various communication mechanisms.

When developing an application on such systems, the designer has to determine the scheduling and the mapping of the tasks and the communications, depending on a set of constraints and dependences, in order to optimize some design metrics, e.g., the program execution time. Moreover, custom hardware accelerators are known to outperform the corresponding software solutions by different orders of magnitude, but the resources for their implementation are usually limited in the final architecture. This research aims at developing a methodology to efficiently generate and optimize multiple hardware implementations, and combine them with existing software ones in order to improve the overall performance of the application.

The proposed methodology starts from a partitioned application (i.e., a task graph) and a model of the heterogeneous target architecture, along with information about the software implementations of each task on each processing element. Then, a multi-objective design space exploration framework for high-level synthesis has been developed for generating multiple hardware implementations for each task. An additional logical optimization can be also performed on these resulting descriptions to further reduce the requirements of resources for their implementation based on the target technology. Finally, the initial description of the partitioned application is analyzed with a novel algorithm, based on Ant Colony Optimization, for mapping and scheduling on the given target platform. In particular, it optimizes the execution time of the entire application by identifying the proper combination of hardware and software implementations and the proper order of execution of the different tasks and communications, respecting all the constraints imposed by the target architecture (e.g., the limited area for hardware accelerators).

The results show that the logical optimization is able to reduce both area occupation and timing after the place and route by more than 10% in average. The design space exploration performed at architectural and system level is then able to obtain solutions 16% better in average compared with traditional methods.

A Methodology for Performance Estimation of Heterogeneous Multiprocessors Embedded Systems

By Marco Lattuada
(lattuada@elet.polimi.it)
Advisor: Prof. Fabrizio Ferrandi
Politecnico di Milano, Italy
February 2011

In the last years, multiprocessor heterogeneous systems have become the de-facto standard in the embedded systems. Despite of the complexity of these architectures, we need to guarantee that the analyzed applications meet the performance constraints. Direct measure of the performance onto the target architecture is not always possible, so we need to esti-

mate the execution time of the overall application.
This thesis proposes a new methodology flow for fast evaluation of design solutions which combines profiling, static analysis of GNU GCC compiler internal representations and Hierarchical Task Graphs. The proposed methodology flow is mainly composed of three parts. The first part exploits the static analysis and automatic annotation of source code to collect characteristics of applications and architecture components. The second part exploits the results of the previous phase and the linear regression technique to automatically build the performance models of the single processing elements. Next, these models are combined with host profiling, mapping information and synchronization costs to produce the overall performance estimation of the application.
The proposed flow has been validated by applying it to embedded system benchmarks targeting some architectures based on ARM processor, LEON3 processor and MAGIC Digital Signal Processor.
The results show that the proposed flow is able to estimate the performance of the single tasks (e.g., 7.38% estimation error on LEON3) and of the whole task graph with good accuracy (4.1%).

A Design Space Exploration Methodology Supporting Run-Time Resource Management for Multi-core Architectures

By Giovanni Mariani (marianig@alari.ch)
Advisors: Prof. Mariagiovanna Sami and Prof. Cristina Silvano
ALaRI, University of Lugano, Switzerland
March 2011

Multi-core systems are currently designed by using platform-based synthesis techniques. In this approach, a wide range of platform parameters are tuned either at design-time or at run-time to provide the best trade-off in terms of the selected figures of merit (e.g. system performance and platform power consumption).
The number of possible configuration alternatives is huge and the evaluation of a single alternative requires accurate performance evaluation which is time consuming. Today a gap exists between the number of system configurations that are needed to be considered as possible solutions and those that could be investigated in an acceptable time. This dissertation proposes a solution to this problem providing a methodology for both design-time and run-time optimization of multi-core platforms. The methodology is twofold. First, a design-time Design Space Exploration (DSE) flow leverages computationally inexpensive analytical models of platform behavior by approximating the objective functions target of the optimization. These models (e.g. linear regression, artificial neural networks, etc…) are trained to fit the simulation results and can then be used for speeding up the DSE. Second, advanced Run-time Resource Management (RRM) techniques leverage information gathered during the design-time DSE to quickly derive an optimal configuration of the run-time tunable parameters once user requirements and system state are given.
The proposed DSE methodology has been implemented and validated on a set of industrial case studies exploiting multi-core computing platforms.

On the Compilation of a Parallel Language Targeting the Self-Adaptive Virtual Processor

By Thomas A.M. Bernard (tho.bernard@gmail.com)
Advisor: Prof. Chris Jesshope
University of Amsterdam, The Netherlands
March 2011

This thesis investigates the changes and the challenges facing the integration of concurrency idioms and assumptions from a concurrent execution model - Self-Adaptive Virtual Processor (SVP) - into an existing sequential-based imperative-language compiler. The compilation schemes of the concurrent language implementation reuse and modify the C-language standards. This thesis focuses on the differences of SVP compilation against conventional sequential compilation. Furthermore, this thesis investigates the impact of such an integration and reveals the feasibility of reusing existing antecedent compiler technology that has been proven to work. Nonetheless, such integration is a significant challenge and the side-effects are relevant to understand the quest of multicore programming tools. At first glance, this work targets a specific audience of technical computer scientists involved in compilation. This work also represents a reflection on the limits of existing engineering methods when used to tackle the challenges of dealing with concurrency. The essence of the technical contribution of this work is then used as a theory of engineering on the limitations of current computing systems and other concurrency-based systems. In the context of facing the multicore programming menace, this work becomes relevant to a wider audience dealing with issues of concurrency-based compilation, language design, and multicore programming issues.

Upcoming Events

The 32nd ACM SIGPLAN conference on Programming Language Design and Implementation (PLDI 2011)
4-8 June 2011, San Jose, CA, USA, http://pldi11.cs.utah.edu/

The 48th Design Automation Conference (DAC 2011)
5-10 June 2011, San Diego, CA, http://www2.dac.com/dac+2011.aspx

The 9th IEEE Symposium on Application Specific Processors (SASP 2011)
5-6 June 2011, San Diego Convention Center, San Diego, CA, http://www.sasp-conference.org

The 6th IEEE Symposium on Industrial Embedded Systems (SIES 2011)
15-17 June, 2011, Västeras, Sweden, http://www.mrtc.mdh.se/sies2011/

The 7th International Conference on Wireless and Mobile Communications (ICWMC 2011),
19-24 June 2011, Luxembourg, http://www.iaria.org/conferences2011/ICWMC11.html

International Conference on Localization and GNSS (ICL-GNSS 2011),
29-30 June 2011,Tampere, Finland, http://www.icl-gnss.org/

The 2nd International Summer School on Information Security and Protection (ISSISP 2011),
4 - 8 July 2011, Ghent, Belgium, http://issisp.elis.ugent.be/

International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS XI) 18-21 July 2011, Samos, Greece, http://samos.et.tudelft.nl/

The 17th International European Conference on Parallel and Distributed Computing (Euro-Par 2011),
29 August – 2 September 2011, Bordeaux, France, http://europar2011.bordeaux.inria.fr

The International Conference on Parallel Computing (ParCo 2011),
30 August - 2 September 2011, Ghent, Belgium, http://www.parco.org/

The Symposium on Parallel Computing with FPGAs (ParaFPGA 2011),
30 August - 2 September 2011, Ghent, Belgium, http://parafpga2011.elis.ugent.be

The 21st International Conference on Field Programmable Logic and Applications (FPL 2011),
5-7 September 2011, Chania, Crete, Greece, http://www.fpl.org

International Symposium on System-on-Chip (SoC 2011),
1-2 November 2011, Tampere, Finland, http://soc.cs.tut.fi/

Conference on Design and Architectures for Signal and Image Processing (DASIP 2011),
2-4 November 2011, Tampere, Finland, http://www.ecsi.org/dasip/

The 7th International Conference on High Performance and Embedded Architectures and Compilers (HiPEAC 2012), 23-25 January 2012 , Paris, France, http://www.hipeac.net/conference

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**

HiPEAC Info is a quarterly newsletter published by the HiPEAC Network of Excellence, funded by the 7th European Framework Programme (FP7) under contract no. IST-217068.
Website: http://www.HiPEAC.net
Subscriptions: http://www.HiPEAC.net/newsletter

Autumn Computing Systems Week, Barcelona, Spain