HiPEAC info 23

COMPILATION ARCHITECTURE

appears quarterly | July 2010

Network of Excellence on High Performance and Embedded Architecture and Compilation

Welcome to ACACES'10, 11-17 July, Terrassa (Barcelona) Spain

www.HiPEAC.net

HiPEAC Computing System Week in Barcelona, Spain, October 19-20 2010

Message from the HiPEAC Coordinator

Koen De Bosschere

Dear friends,
Some of you will read this issue of the HiPEAC newsletter while arriving at the ACACES 2010 summer school. This year again, we succeeded in attracting about 200 students for a full week of courses, keynotes, discussions, and social events. The ACACES summer school remains one of the flagship events organized by the HiPEAC community.

In May, we enjoyed the Innovation Event at the Informatics Forum in Edinburgh. The theme of the event was "Helping European researchers and the EU deliver for SMEs". The event was well attended by over 120 participants, and the appreciation for the event was one of the highest we got in the recent years.

I got particularly interested in the ProspeKT initiative that seems to be very successful in commercializing informatics research in the Edinburgh region. I hope that this example will inspire other universities to set up similar initiatives in the coming years.

For the first time, the recently granted FP7 projects in the computing systems objective had an opportunity to present their plans to the HiPEAC community. I was pleased to see that HiPEAC members are well represented in different projects and hope that this event can be the beginning of a more structural collaboration between the different computing systems projects.

At the end of the event, Panos Tsarchopoulos gave an exclusive preview of the upcoming Call 7. I expect that this presentation, in combination with the presentation of the projects that made it in Call 6, will help our

community to prepare many high quality project proposals for Call 7. I especially hope that HiPEAC members who never tried or have not been successful so far will take advantage of this call to submit a high quality project proposal. The end of the summer school traditionally marks the start of the summer holiday for the HiPEAC community. Let me therefore thank all of you for the fruitful HiPEAC collaboration during the last year, and wish you a pleasant and relaxing summer holiday.

Koen De Bosschere

Release of the Speedup-Test Tool

Numerous code optimization techniques are usually experimented with by doing multiple observations of the initial and the optimized execution times in order to determine a speedup. Even with the input and execution environment being fixed, program execution times vary in general. As a consequence different kinds of speedups may be reported: the speedup of the average execution time, the speedup of the minimal execution time, the speedup of the median, etc. Many published speedups in the literature are observations of a set of experiments. In order to improve the reproducibility of experimental results,

this tool implements a rigorous statistical methodology regarding program performance analysis. We rely on well known statistical tests (Shapiro-wilk's test, Fisher's F-test, Student's t-test, Kolmogorov-Smirnov's test, Wilcoxon-Mann-Whitney's test) to study if the observed speedups are statistically significant or not. By fixing $0<a<1$ a desired risk level, we are able to analyze the statistical significance of the average execution time as well as the median. We can also check if $P[X>Y] > \frac{1}{2}$, the probability that an individual execution of the optimized code is faster than the individual execution of the initial code. Our methodology defines a consistent improvement compared to the usual performance analysis method in high performance

computing. For each situation, we explain the hypothesis that must be checked

Sid-Ahmed-Ali Touati

in order to declare a correct risk level for the statistics. The Speedup-Test protocol certifying the observed speedups with rigorous statistics is implemented and distributed as an open source tool based on R software. This methodology was presented in the form of tutorials at international conferences such as: HIPEAC (Pisa 2010), CGO (Toronto 2010), ICS (Tsukuba 2010), HPCS (Caen 2010).

Sid-Ahmed-Ali Touati and Julien Worms and Sebastien Briais

URL: http://hal.inria.fr/inria-00443839

Panos Tsarchopoulos
Panagiotis.Tsarchopoulos@ec.europa.eu

Message from the Project Officer

The next Call for Proposals in the FP7-ICT research programme is scheduled for publication in July 2010. This Call includes the research objective Computing Systems with a significantly increased budget compared to previous Calls. The significantly increased budget reflects the importance of the formidable research challenges posed by the multicore transition across all computing market segments. The following should outline the main areas of the Computing Systems Call for Proposals. Please consult our Cordis web site to find the full official text of the Call and all related information for preparing and submitting proposals including the deadline and the type of proposals for each area.

a. Parallel and Concurrent Computing

Automatic parallelization, new high-level parallel and concurrent programming languages and/or extensions to existing languages (including their runtime implementation) that provide portable performance taking into consideration that user uptake is a crucial issue. Projects should go beyond on-chip, off-chip boundaries addressing the challenges of programming, testing, verification and debugging, performance monitoring and analysis, low-power and power management especially for large scale parallel systems and data centers, and heterogeneous and accelerator-based multi-core systems. Research priorities include domain-specific languages; concurrent algorithms and transformation of concurrency to parallelism through adaptive compilers and runtime systems; new verification and optimization environments for parallel software; efficient execution exploiting heterogeneous cores; new approaches to scalability of high-performance computing application codes.

b. Virtualization

Virtualization technologies that are ensuring task isolation and optimized resource allocation as well as guaranteeing performance, timing and reliability constraints. The focus is on full virtualization solutions for heterogeneous multicore platforms including the design of virtualization-ready heterogeneous multicore hardware platforms and support for accelerator virtualization.

c. Customization

Unifying hardware design and software development with an emphasis on rapid discovery and production of optimal customizations of heterogeneous single-chip multicore systems and associated tool-chains for particular applications. Research priorities include: reconfigurable, flexible, soft or hybrid architectures and instruction sets; automatic tool-chain generation; system modeling and simulation, including performance predictability; efficient exploration of the customization space; low-power and customization for power efficiency; parallel programming for single-chip multicore architectures; architectural and system-level reliability techniques to counter increasingly probabilistic behavior of transistors in lower geometries.

d. Architecture and Technology

The focus is on the impact of next-generation chip fabrication technology on system architectures, tools and compilers. Research areas include: implications of 3D stacking; alternative (non von Neumann) models of computation. The key challenge is to bridge parallel computing architectures and chip fabrication technology.

e) International Collaboration

The purpose is to analyze international research agendas and to prepare concrete initiatives for international collaboration, in particular with the USA, India, China and Latin America, for all topics of this objective. Separate proposals per geographic area are expected.

Panos Tsarchopoulos

HiPEAC News

Mateo Valero, New Member of the Royal Academy of Science and Arts

On January 21st, 2010, Professor Mateo Valero was received as Elected Academic member at the Royal Academy of Science and Arts of Barcelona. The academy was established on 18 January 1764 as a private literary society, and later became a public consulting body for the king for matters relating to the Principality of Catalonia. In accordance with the Founding Statutes, the Academy is an association with scientific researchers and its applications; membership to which is limited by number and only by nomination. Ordinary member Miguel Angel Lagunas with his kindly overview of Valero's achievements introduced Prof. Valero, from his original Telecommunications School, followed by the Computer Architecture Dept., until the creation of the Barcelona Supercomputing Center, and how, at the same time, he has always been a loyal friend among his colleagues and a well-respected and prolific student advisor. During his speech, Prof. Valero exposed his broad knowledge about how research is done in High Performance Computing and some interesting related success stories.

HiPEAC Innovation Event in Edinburgh

The spring 2010 edition of the HiPEAC Computing Systems Week took place in Edinburgh, UK from 3rd to 5th May. This time, the organizers decided to try a different format. The primary focus was to encourage, start and foster interaction between different clusters and attendees from academia and industry. Despite some intervention by the Icelandic volcano Eyjafjallajökull, a record number of researchers were able to attend. In total more than 120 people from 18 different countries came to Edinburgh.

The first day started with a keynote by Mario Nemirovsky about the challenges faced by high-tech entrepreneurs. The rest of the day was reserved for inter cluster meetings. These meetings discussed problems that involved topics relevant to several clusters. "Cloud Computing Software and Hardware Challenges" addressed issues that arise from the need to become more energy efficient, a new challenge in areas that were traditionally unconcerned about energy consumption. This meeting involved the multicore, interconnects, reconfigurable, virtualization and compilation clusters. The second meeting took the form of a panel discussion to address the topic: "Dusk of general-purpose many-cores, dawn of heterogeneous multi-cores: What are the design and programming issues?" It brought together members from the programming, multicore, compilation, reconfigurable, virtualization and design clusters.

The main event was on the second day. It started with a high profile keynote by Uri Weiser from Technion (Israel) on multi-core processors. The following presentations focused on commercializing academic research. Colin Adams gave some insights into the schemes the School of Informatics at the University of Edinburgh runs in order to help researchers to commercialize their ideas. Robert Jelski from Capital-E presented the view of the venture capitalists and what they are looking for when they decide to fund a project. The afternoon was reserved for interactions between SME and academic members. It started with short presentations by each of the 10 SMEs (Movidius unfortunately fell victim to Eyjafjallajökull), followed by a networking event in the atrium of the Informatics Forum. During this event, the SMEs presented their products in small stands. Many SMEs had interactive presentations. The academic participants presented their work in 19 posters. The atrium of the Informatics Forum proved to be an ideal venue for this event, preventing people from dispersing too much while offering more than enough space to breath. Finally, the industrial participants were invited to cast a ballot to determine the three best posters:

1. Multicore Architecture for Critical Real-Time Embedded Systems by Marco Paolieri, Eduardo Quiñones, Francisco J. Cazorla and Mateo Valero, UPC

2. EnCore microprocessor by Igor Böhm, University of Edinburgh

3. Explicit Interprocessor Communication in the SARC Architecture by Christoforos Kachris, FORTH

Congratulations to the winners.

The final day focused on the 7th Framework Programme of the European Union. The day started with presentations on the following FP7 projects: 2PARMA, Advance, Encore, EuroCloud, HEAP, PEPPHER, PlanetHPC, PRO3D, REFLECT and TERAFLUX. As a conclusion to the day, project officer Panagiotis Tsarchopoulos gave a brief glimpse into the upcoming systems call by the European Commission.

The Innovation Event also featured another innovation: small social events that allowed participants to get in touch with each other in a very informal setting. On Monday, everybody was invited for a short walk up to the summit of Arthur's Seat - Edinburgh's own volcano (extinguished); followed by dinner to regain strength.
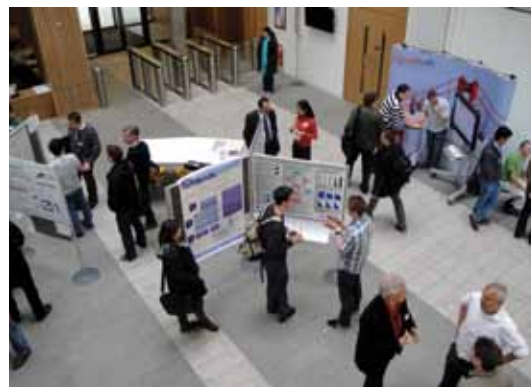
On Tuesday, the social event was a Haunted Underground Tour. Guide William demonstrated with great attention to detail what Edinburgh's citizens considered entertainment before television was invented. The tour continued into abandoned vaults under South Bridge, where some participants swear that they heard some screams of agony - while others

claim they didn't hear anything at all. Overall, the organizers were extremely pleased by the huge interest in the social events. More than 30 people each attended both events.

Chris Fensch
HiPEAC Innovation Event Organizer

Towards hipeac.pl: German-Polish ICT Workshop in Warsaw

Integration of more people and expertise from new EU member states is an important vision of HiPEAC. During the fall 2009 Computing Systems Week in Wrocław, a new informal activity called "hipeac.pl" was founded, coordinated by Zbigniew Chamski (Infrasoft, Poland) and Rainer Leupers (RWTH Aachen University). Recently, also Dr. Bartłomiej Świercz (TU Łódź) has engaged in this activity. Its mission is to create a local HiPEAC portal in Poland that provides an information platform for relevant Polish players from academia and industry and thereby reduces the entry barrier into the "traditional" HiPEAC community. A further step towards this direction was taken at the German-Polish ICT Workshop, held in Warsaw on May 31. The event was kindly organized by Dr. Monika Schidorowitz, head of the science department at the German embassy in Poland, with help of TU Warsaw and the German Education

and Research Ministry (BMBF). It attracted around 80 participants, among them representatives from all leading Polish academic institutions, and was accompanied by a Fraunhofer science truck hands-on exhibition in various spots of Warsaw city. During the workshop, statistics presented by the Polish National Contact Point showed that, in spite of already many existing successful collaborations, higher participation of Polish institutions to European research frameworks is still desirable. To help kick-start new collaborations, the BMBF offers several seed funding programs. In a plenary meeting, various scientists from German and Polish academic institutions presented their expertise and cooperation offerings. Rainer Leupers highlighted opportunities in Embedded System design and pointed out the benefits of a HiPEAC membership. After the plenary talks, the audience split into thematic groups,

Rainer Leupers presenting cooperation opportunities and HiPEAC in the Hall of the Senate of TU Warsaw

enabling more focused discussions. The ICT workshop was very well received by all participants, and the major outcome is a long list of newly established contacts and mutual invitations, which will hopefully result in tangible joint projects in the long term. The next steps of hipeac.pl are to set up a web portal and to present HiPEAC in a special session of the MIXDES conference, held in Wrocław during June 24-26.

Rainer Leupers

Joint Seminar: Imperial College London and RWTH Aachen University

During the last week of May, the Institute for Communication Technologies and Embedded Systems (ICE) of RWTH Aachen University visited the Department of Computing at Imperial College in London for a joint seminar. The event began with three introductory speeches by Profs. Peter Cheung, Rainer Leupers, and Wayne Luk, giving an overview of research activities carried out at their institutions. The meeting continued with a technical part during which Aachen was first introduced to the FPGA design concerns that Imperial is handling. Those included modeling of FPGA degradation, and energy-aware optimization for run-time reconfiguration, with an optimization algorithm driven by an energy model including the overhead of the reconfiguration process itself. The domain specific language Contessa suited for financial applications with wide usage of Monte Carlo simulations was also presented. By means of a high-level synthesis tool, an application modeled with Contessa can be directly compiled to reconfigurable logic. Finally, a method to compute the optimal precision for an algorithm that is to be executed on an FPGA was shown to the audience.

The morning session was concluded with a guided tour of Imperial College buildings and surrounding areas kindly organized by Prof. Cheung and Prof. Luk. This excursion gave some insights into Ph.D. and undergraduate student's life at Imperial. Besides auditoriums and classrooms the Aachen team had a chance to see some extraordinary student projects. For example, a Budweiser bottle organ was used to play MIDI files directly from a PC. The sound quality by the way could compete with real instruments of that kind!

In the afternoon the team from Aachen presented a number of System Design techniques. First, a processor design tool chain (LISATek) was introduced, along with an example of Application Specific Instruction Set Processor design for cryptographic pairings. A presentation on reconfigurable ASIPs followed, a novel concept to make use of both, FPGAs and ASIPs at the same time. Next, an overview of the Nucleus project was given, whose main objective is to estab-

Recore Systems Receives the Van den Kroonenberg Prize 2010

Recore Systems has received the Van den Kroonenberg Prize for young entrepreneurship. The award is presented annually to innovative entrepreneurs at the close of the University of Twente Innovation Lecture.

The prize was awarded to Recore's founders: Paul Heysters, Lodewijk Smit and Gerard Rauwerda. The company was founded as a spin-off of the University of Twente in 2005. Since then, the company has been growing healthily and developed innovative products based on scientific knowledge. Recore Systems' products enable highly efficient reconfigurable multi-core systems for applications such as broadcasting, multimedia, wireless telecommunication and digital beamforming. The company will soon release a chip for receiving digital radio and TV, targeting consumer electronics such as portable media players and smart phones.

Recore Systems and the University of Twente continue to tackle advanced research challenges in several joint projects. In the CRISP project, a Europe-wide consortium is using Recore's new Xentium® technology to demonstrate a highly scalable, reconfigurable system concept for use in a wide range of applications (for more information see www.crisp-project.eu).

Flanders ExaScience Lab

lish a novel design methodology and a tool flow for Software Defined Radio. A presentation of the MPSoC Application Programming Studio (MAPS) followed, giving an overview of one exemplary multi-core programming flow for data streaming applications. The seminar was concluded by a talk about different techniques to achieve high simulation speeds for SoCs that are extremely important for hardware and software developers in the embedded community.

High interest was sparked on both sides, proving that the presentations have served their purpose. A lively exchange of experience was notable, both during the seminar and over a social event that followed. By dinner-time in the exclusive rooms of Imperial College the participants got to know each other, what led to spirited conversations over the excellent food. The positive feedback from both participating sides might be encouraging for further HiPEAC members to foster such kinds of knowledge-exchanging events.

About the Van den Kroonenberg Prize

The Van den Kroonenberg Prize is awarded to an entrepreneur with an innovative product or progressive business model, who maintains a close association with the University of Twente. The annual prize, which is being awarded this year for the 27th time, is named after the former Rector of the UT Harry van den Kroonenberg.

Intel ExaScience team.

On June 8, 2010 – Intel Corporation, IMEC and 5 Flemish universities officially opened the Flanders ExaScience Lab at the IMEC research facilities in Leuven, Belgium. The ExaScience Lab is the latest member of Intel's European research network – Intel Labs Europe – that consists of 21 labs employing more than 900 R&D professionals. Flanders ExaScience Lab will develop software to run on Intel-based future exascale computer systems delivering 1,000 times the performance of today's fastest super-computers, using up to 1 million cores and 1 billion processes to do so.

Designing exascale computers using current technology and design methodologies would mean the systems would become extremely hot and require a power plant to deliver the power needed to run them. When building a system consisting of millions of cores, getting all of them to work together for an extended period of time also represents a challenge. Hence, completely new computer programming methods and software will be required to bring power consumption to acceptable levels and to make the system fault tolerant. Power and reliability will be the key challenges that need to be understood to turn the vision of exascale computing into reality.

In the Flanders ExaScience Lab, Intel will collaborate with IMEC and all Flemish universities – University of Antwerp, Ghent University, Hasselt University, Katholieke Universiteit Leuven and Vrije Universiteit Brussel.

The Flanders ExaScience Lab kicks off with close to two dozen researchers and will add another dozen or so by 2012. The Flanders ExaScience Lab will be focused at enabling scientific applications, beginning with the simulation and prediction of "space weather", i.e., electromagnetic activity in the space surrounding the Earth's atmosphere. Solar flares – large explosions in the Sun's atmosphere – can cause direct damage to the Earth. Damage can be to electric power networks, pipeline systems and the quality of wireless communication, to name just a few examples. To accurately predict and understand such effects, exascale computing power is needed. Chosen for its extremely complex nature, the software findings are expected to be used and extended to address many other problems.

Two long-standing HiPEAC partners, Ghent University and IMEC, are deeply involved in the activities of the lab. They will jointly develop scalable simulation technology, analytical performance models, and reliability models for exascale computing systems.

Website: http://www.exascience.com/ For more information on the simulation technology, contact Prof. Lieven Eeckhout, Ghent University. Lieven.Eeckhout@elis.UGent.be

FP6 hArtes Project: the Gateway to Heterogeneous, Multi-Core Platforms

Partners:
Atmel Roma (coordinator), Scaleo, Thales, Segula, Delft University of Technology (scientific coordinator), Politecnico di Milano, Imperial College, Inria, Leaf, University of Ferrara, Faital, Universita della Marche, Fraunhofer Darmstadt, UAPV, TCC, Politecnico di Bari.

Embedded systems applications require increasingly more processing power for which single processor platforms are no longer sufficient. On the other hand, multi-core platforms not only find their way into the desktop and server markets but also in the embedded systems domain. Such platforms can contain any number of computing nodes, ranging from RISC processors to DSP's and FPGA's. hArtes is targeting such a platform containing an ARM processor combined with a powerful floating point unit, the Diopsis.

Use a Familiar Programming Paradigm

One of the key challenges when adopting such platforms is that one is forced to use programming tools and languages that are very platform specific and require substantial code rewriting to port an existing application. In addition, the learning process for using these tools is long and needs to be repeated each time when adopting another technology. hArtes proposes a familiar programming paradigm that is compatible with widely used programming practice, irrespective of the target platform.

In essence: conventional source code such as ANSI-C, will be annotated after some (semi) automatic transformations have been applied. These transformations are driven by performance requirements. The hArtes tool chain can automatically make the necessary transformations, mappings and subsequent code annotations. It also allows the developer to make simple annotations to the source code indicating which parts of the code will be accelerated.

See Multi-Cores as a Single Processor

One of the distinguishing features of the hArtes approach is that it abstracts away the heterogeneity as well as the multi-core aspect of the underlying hardware. The developer can view the platform as consisting of a single general-purpose processor. This view is completely consistent with the programming paradigm that was presented above. The hArtes toolboxes take care of the mapping of parts of the application on the respective hardware components.

The hArtes Toolboxes

The entry point of the tool chain is either hand written C-code or C-code generated by tools such as NU-Tech and SciLab. In principle, any tool that produces ANSI-C will be able to connect to the tool chain.

The hArtes tool flow

There are 3 main tool boxes: a parallelization tool, called Zebu, a mapping tool called Harmonic, and a system tool box consisting of a modified GCC compiler, a licensed Target compiler, and the DWARV hardware compiler.

The hArmonic toolset contains a transformation engine, a data representation optimiser, and a cost estimator. hArmonic investigates different possible mappings and applies clustering, mapping and scheduling in one integrated step.

The system toolbox consists of a modified GCC compiler, the DWARV hardware compiler and a licensed version of the Target Compiler (r).
The well known and widely used open source compiler GCC is being used as the software compiler. GCC version 4.2 has been extended to generate appropriate codes for the pragma annotations that have been introduced by the tool chain, targeting the ARM processor that is the GPP in the Diopsis platform. The parts of the application that have been mapped on the FPGA are used by the DWARV compiler to generate synthesizable VHDL that can then be sent through any proprietary synthesis tool kit.
In the end, a linker stitches everything together to produce an ELF executable file.

The tool chain has been extensively tested on a wide variety of applications in the streaming domain such as beamforming and wavefield synthesis (Fraunhofer), noise filters (Thales), and in-car audio applications (Univ. Dellamarche). Also, various hardware platforms have been used such as the hArtes platform comprising of the Arm, Diopsis and Xilinx 4, as well as the Scaleo Chip that used the tool chain to map applications on their
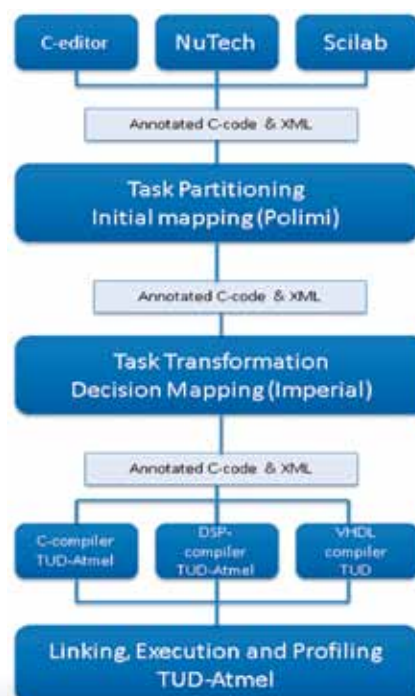
emulation board that runs on an Altera FPGA.

The project ended in February 2010 but the tool chain is made available through a foundation, which is open for companies, research institutes as well as individuals. A book will soon be published by Springer Verlag describing in detail the results obtained by the project. For more information consult the website www.hartes.org or email k.l.m.bertels@tudelft.nl.

CAPS, Global Solutions Provider for Manycore Applications Deployment

Founded in April 2002 and based in France, CAPS has been developing manycore programming tools for more than 7 years. CAPS is now a leading global provider of compiler technologies and engineering services for parallel hybrid computing.

Now boarding to... Shanghai, China

Bolstered by its success in Europe with major companies in energy, oil and gas, defense and research, CAPS has started its internationalization in 2009 by signing new partnerships with American and Asian high performance computing actors.

Early April this year, CAPS expands into Asia Pacific with the opening of new offices in Shanghai in China, thus reinforcing its local presence in APAC first initiated in 2009 by new reselling agreements with supercomputing specialists in China, Japan and Taïwan. The opening of this office in Shanghaï is the first step towards the company's internationalization strategy. CAPS is currently working out how to settle down in the United States. Beyond the opening of new offices, this geographical expansion has already expressed itself by prestigious references on main geographical areas using numerical computation. Among our last references, we count ORNL in the US, HLRS in Germany, Tokyo Institute of Technology in Japan. This success was also possible thanks to our local resellers: Paratools in the US, JCC Gimmick, Aravision and CHPC in Asia.

The opening of new offices in Shanghai arises at a significant time in CAPS expansion to APAC. Thanks to its recent reseller network relying on supercomputing specialists, more and more organizations, computing centers, universities and industrial companies from APAC are now using HMPP™, CAPS hybrid compiler.

HMPP 2.3: a mature hybrid compiler with CUDA and OpenCL back-ends

This international success comes with a matured CAPS offer. On June 1st at ISC'10 in Hamburg, the company announced the availability of an OpenCL code generator within the just released 2.3 version of its HMPP directive-based hybrid compiler. In addition, the CUDA back-end generator has been enhanced with Fermi capabilities. This new release brings support for even more native compilers like Intel ifort/icc, GNU gcc/gfortran and PGI pgcc/pgfort, enabling developers to freely use their favorite compiler with HMPP 2.3.

Based on GPU programming and tuning directives, HMPP offers an incremental programming model that allows developers with different levels of expertise to fully exploit GPU hardware accelerators in their legacy code.

As an emerging open programming standard the OpenCL back-end expands the portfolio of targets supported by HMPP to AMD ATI GPUs. The OpenCL version of HMPP fully supports AMD and NVIDIA GPU compute processors, bringing to users a wider set of hybrid platforms they can execute their applications on. Recently released, the NVIDIA Tesla 200-series GPUs based on the 'Fermi' codenamed new CUDA architecture is also supported by HMPP 2.3.

The addition of this OpenCL back-end to the existing NVIDIA CUDA back-end is a major milestone in HMPP development that gives users another powerful standard programming option.

Benoît Raoult, CAPS APAC VP and General Manager

FP7 NaNoC Project: Nanoscale Silicon-Aware Network-on-Chip Design Platform

Project coordinator:
José Flich
Universidad Politécnica de Valencia
jflich@disca.upv.es
Project website:
www.nanoc-project.eu
Partners:
Universidad Politécnica de Valencia (Spain)
Università degli Studi di Ferrara (Italy)
Infineon Technologies AG (Germany)
Simula Research Laboratory (Norway)
Teklatech A/S (Denmark)
iNoCs SàRL (Switzerland)
Lantiq (Germany)

Main objective

Multi-core Systems-on-Chip (SoCs) are becoming ubiquitous in multiple industrial domains, from consumer electronics to automotive, from telecommunications to industrial automation. However, numerous challenges lie ahead, especially regarding the design complexity of such platforms and the physical-level issues as fabrication is further miniaturized. On the other hand, there is today wide consensus on the inherent performance scalability limitations of state-of-the-art interconnect fabrics, ranging from shared busses to bridged busses, all the way to the latest multi-layer communication architectures. Networks-on-chip (NoCs) are currently advocated as an alternative interconnect fabric for effective system integration. With NoCs, performance scalability becomes more a matter of instantiation and connectivity capability rather than architecture complexity. The key-limiting factor for widespread industrial adoption of NoCs is not the maturity of the architecture but rather

the design technology support. The NaNoC project aims at developing an innovative design platform for future NoC-based multi-core systems. The ultimate objective is to master the design complexity of advanced microelectronic systems by enabling a strictly component-oriented architecture design. A compositional approach to NoC design in future multi-core chips is out of the reach of current design methods and tools due to emerging challenges.

On one hand, NoC co-design with high-level platform management frameworks raises the need for enhanced dynamism and flexibility in NoC composition (e.g., virtualization, power management, thermal management, application management). On the other hand, a higher degree of uncertainty originating from nanoscale IC fabrication technologies makes the design of reliable systems out of unreliable components a challenging task. The NaNoC design platform will provide design methods and prototype tools to cope with both challenges and make NoCs a mainstream interconnect backbone for effective system integration in the landscape of 2015 multi-core platforms.

A key to the success of the NaNoC design methods and tools, and to the industrial uptake and practical exploitation of the outputs of the project, is their integration into a coherent design platform. In this direction, the NaNoC platform not only provides a cross-layer approach to tackle future

SoC design challenges (e.g., physical design techniques for enhanced reliability combined with architecture-level techniques for fault containment), but also defines an open standard for communicating design intents across layers of the design hierarchy. This way, tool interoperability in a cooperative design environment will be promoted, and silicon-aware decision-making at each design step will be enabled. A key concern of the NaNoC project is to make developed tools interoperable with and sometimes even integrated into mainstream industrial tool flows. The Figure illustrates the vertical integration pursued by the NaNoC design platform, showing how it intends to bridge an abstract design specification with a physical implementation.

The NaNoC design platform aims at a vertical integration of NoC technologies across layers of design abstractions.

Vertical integration provided by the NaNoC design platform.
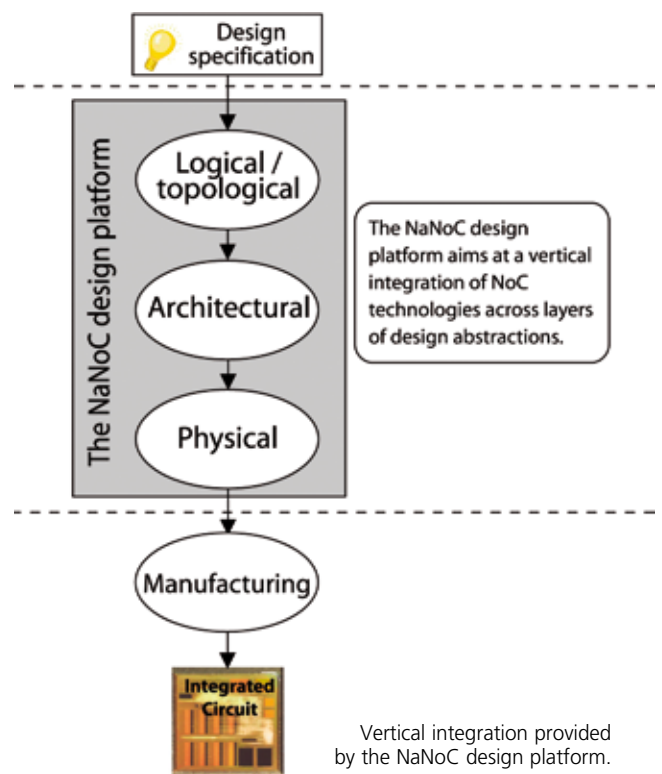
Partner Contributions

Vertical integration has been conceived from the ground up by careful selection of consortium partners based on their key expertise.

From the design technology perspective, iNoCs will provide its unique expertise in design tools for automatic instantiation of NoCs matching application requirements. Universidad Politécnica de Valencia, Simula, Infineon, and Lantiq will be the architecture-level players of the project, and will provide novel design methods able to cope with the emerging reliability, reconfigurability and

3D integration requirements of multi-core systems. When it comes to the physical-layer, Teklatech will provide its expertise on the backend design tool flow to address the challenges of designing with nanoscale technologies. Finally, the interdisciplinary expertise of Universitá degli Studi di Ferrara makes it suitable for the task of platform-level integration of developed design methods and tools.

Project Coordinator José Flich

How to Teach Introductory Architecture & Programming: Videotaped Pisa Tutorial

The first tutorial organized by the HiPEAC Task Force on Education and Training was held in Pisa, Italy, on 24 January, just before the HiPEAC 2010 Conference started. It dealt with Teaching Introductory Computer Architecture and Programming, and addressed the questions of What, When, and How. Yale Patt and Sean Halle lectured, followed by a discussion led by Avi Mendelson. The full lectures and discussion were videotaped, and the recorded files are available from the web addresses given on the task force page, below.

The first part, on "The Art of Teaching", reminded us of valuable advice that we usually have vaguely in our minds, but all too often forget to apply in practice -among others: Teach fundamentals, not tools, fads, etc; teach critical thinking, not memorization (Bill Gates was once asked "what to teach", and answered "Knuth Volume 1"). Classical teaching tools are the basis - blackboard is the best; animations can help, but don't overdo them; email helps, especially if you reply quickly and at odd hours. Exams are best with closed books, but

allow each student to have one page of personal, hand-written notes.

The second part advocated the "Motivated Bottom-Up Approach": repeated cycles of motivating (top-down), then explaining how it works (bottom-up) - be careful to distinguish Design (a top-down activity) from Learning (you can only design after you know the components, from bottom-up). Understanding memory and addresses must be a first concern; also, understand calls and activation records, then understand recursion. Information hiding is good, but first the students need to have some information before they have something to hide.

The third part talked about "The Top-Down Approach", motivated by the need for code portability and by the "most radical change" of all times: parallelism. It advocated building on primitives such as: memories and their content; transformations to the state in memory - procedures, or tasks as "processor specifications"; the clock as an animator; and dependencies as communication in memory. Three

parallel teaching tracks are important: Software (applications, dependencies, scheduling), Hardware (sequential, parallel, networks, systems), and Tools (compilers, etc).

During the fourth part, there were extensive discussion between the audience and the lecturers. Among others, one question was whether the first course in programming should be or contain parallel programming. The recommendation was to start simple and concrete: the first few programs should be sequential; tell the students that parallel is the next step, and do NOT say it is difficult; quickly move on to a parallel program, within the same course.

For more information, and for pointers to the 4 video files, see www.hipeac.net/TF_education

FP7 ERA Project: Embedded Reconfigurable Architectures

Project coordinator:
Dr. Stephan Wong, TU Delft, The Netherlands
Email: J.S.S.M.Wong@tudelft.nl
Website:
www.era-project.eu
Partners:
Delft University of Technology (The Netherlands), Industrial Systems Institute (Greece), University of Siena (Italy), Chalmers University of Technology (Sweden), University of Edinburgh (United Kingdom), Evidence s.r.l. (Italy), ST Microelectronics (Italy), IBM Research Laboratory (Israel), Universidade Federal do Rio Grande do Sul (UFRGS) (Brazil)
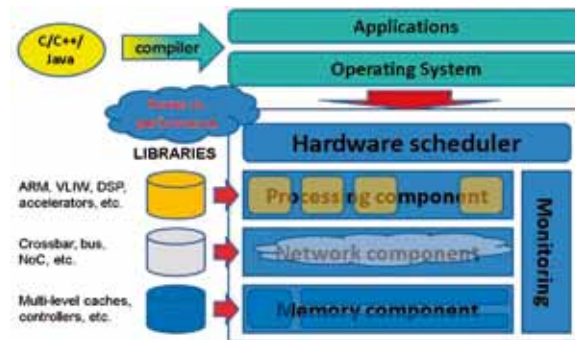**Start:** 2010.01.01
Duration: 36 months

Abstract overview of the ERA platform

ERA aims at investigating and developing new methodologies in both tools and hardware designs to break through current power and memory walls for the next-generation embedded systems platforms. The proposed strategy is to utilize adaptive hardware to provide the highest possible performance for a given power budget. The envisioned ERA platform is adaptive and employs a structured design to integrate the necessary computing, networking, and memory elements.

Main Objectives

The envisioned adaptive ERA platform employs a structured design approach that allows integration of varying computing elements, networking elements, and memory elements. For computing elements, we will utilize a mixture of commercially available off-the-shelf processor cores, industry-owned IP cores, and application-specific/dedicated cores, and we will dynamically adapt their composition, organization, and even instruction-set architectures to provide the best possible performance/power trade-offs. Similarly, the choice of the most-suited network elements and topology as well as the adaptation of the hierarchy and organization of the memory elements can be determined at design- time or at run-time. Furthermore, the envisioned adaptive platform must be supported by and/or made visible to the application(s), run-time system, operating system, and compiler exploiting the synchronicities between software and hardware. We strongly believe that having complete freedom to flexibly tune hardware elements will allow for a much higher level of efficiency (e.g., riding the trade-off curve between performance and power). Finally, an additional goal of the adaptive platform is to serve as a quick prototyping platform in embedded systems design.

Technical Approach

In the ERA project, we identified four key areas to pursue innovations in order to achieve our objectives:
• Definition and characterization of application benchmarks for embedded systems employing reconfigurable architectures.
• Definition of a reconfigurable and parameterized processor architecture.
• Definition of a reconfigurable memory subsystem.
• Definition of software/compiler tools and OS support for the ERA platform.

The applications exhibit behavior that can be exploited for more efficient processing (at given power budgets) by adapting the hardware (processor and memory) to them. This paradigm shift requires new approaches in compiler algorithms and tools and advanced (embedded) OS-level support. All partners have expertise in one or several of the mentioned areas.

Expected Impact

The industrial partners clearly identified the benefits of the ERA project expressed by their involvement in and their commitment to the project. All the proposed solutions in this project will be combined in a demonstrator platform that is expected to allow industrial partners fast access to new products developed on top of it. The intended platform will serve several purposes:
• Quick development platform for industry: the clear interfaces defined in this project should allow industrial partners to take from the platform everything they need while still being able to incorporate their own IPs. Moreover, for low volumes even the prototype can be used as a commercially viable product, since the consortium will use available FPGA technology to validate its contribution.
• Academic purposes: the ERA platform can be easily used to build different instances of embedded processing solutions and we foresee and will actively pursue the possibility of incorporating the ERA platform as a teaching tool in embedded courses or labs.

Project
Coordinator
Stephan Wong

FP7 PROARTIS Project: Probabilistically Analyzable Real-Time Systems

Project Coordinator
Francisco J. Cazorla
Barcelona Supercomputing
Center (Spain)
Project Website:
www.proartis-project.eu
Partners
Barcelona Supercomputing
Center (Spain)
Rapita Systems Ltd (UK)
INRIA (France)
University of Padua (Italy)
Airbus (France)
Start: 2010.02.01
Duration: 36 months

The Critical Real-Time Embedded (CRTE) Systems industry demands new functionality and ever-higher levels of performance together with reduced cost, weight and power consumption. This can only be delivered by advanced hardware features. However, the timing behavior of systems using these advanced hardware features is not analyzable with current analysis techniques and paradigms. New hardware/software design paradigms, developed together with novel analysis techniques, are required to enable the analysis of these systems with high levels of confidence in their temporal correctness.

Systems cannot be fault-free because, even if extremely low, the probability of incurring an unexpected fatal or harmful event is never null (for example, a meteorite that hits the system and destroys it). In fact, system reliability is expressed in terms of probabilities for hardware failures, memory failures, and software failures and for the system as a whole. PROARTIS extends this probabilistic approach to timing correctness.

The underlying objective of the PROARTIS project is to enable a probabilistic timing analysis that will prove that pathological timing cases can only arise with negligible probability, rather than trying to eliminate them (which is arguably not possible and could be detrimental to performance). This is a key difference from previous approaches that seek analyzability by trying to predict with cycle accuracy the state of hardware and software through analysis. To summarize:

The central hypothesis of PROARTIS is that new advanced hardware/software features can be used and analyzed effectively in CRTE systems with designs that provide truly randomized behavior. This shift will enable probabilistic analysis techniques that can be used effectively in arguments of verification of these systems, demonstrating that the probability of pathological execution times is negligible.

The techniques developed in PROARTIS will enable probabilistic guarantees of timing correctness to be derived. For example, if the requirements placed on the reliability of a sub-system indicate that the probability of a timing failure must be less than 10-9 per hour of operation, then the analysis techniques developed in PROARTIS aim to translate this reliability requirement into a probabilistic worst-case execution time for the sub-system. Probabilistic analysis effectively provides a continuum of worst-case execution times (WCETs) for different confidence levels. Thus, a sub-system may have a probability of 10-8 per hour of exceeding an execution time of 1.5 ms, and probabilities of 10-9, and 10-10 per hour of exceeding 1.55 ms and 1.59 ms, respectively. The main idea of PROARTIS is that for future CRTE systems, such probabilistic guarantees offer significant advantages over deterministic approaches attempting to make absolute guarantees, severely limiting the use of advanced hardware features and inevitably attaining considerably lower performance guarantees.

The project involves Academic institutions: Barcelona Supercomputing Center (Spain) as project coordinator lead by Francisco J. Cazorla with expertise on hardware architecture and simulation, INRIA (France) led by Liliana Cucu with expertise on probabilistic analysis, University of Padua (Italy) led by Tullio Vardanega with expertise on real-time systems modeling and analysis; SME: Rapita Systems Ltd (UK) led by Guillem Bernat with expertise on Worst-Case Execution Timing Analysis, probabilistic timing analysis and Tools (RapiTime); and end customer: AIRBUS (France) led by Benoit Triquet with expertise on analysis and development of safety critical systems for avionics systems. The project also incorporates an industrial advisory board to bring industrial expertise into the project and to promote exploitation and dissemination. Members of the IAB are: Hardware manufacturers: IBM (Israel), Infineon (UK) and NXP (Netherlands); Tool vendors: SYSGO (France) and AdaCore (France); OEMs: AUDI, BMW and European Space Agency; Academic institutions: City University (UK). The project also includes the University of York (UK) and the University of Massachuse as affiliated members.

Project Coordinator
Francisco J. Cazorla

FP7 TERAFLUX Project: Exploiting Dataflow Parallelism in Teradevice Computing

Project Coordinator
Università degli Studi di Siena
Prof. Roberto Giorgi
giorgi@dii.unisi.it
Project Website:
www.teraflux.eu
Partners:
Università degli Studi di Siena
(Italy)
Barcelona Supercomputing
Center (Spain)
CAPS (France)
Hewlett Packard Labs (Spain)
INRIA (France)
MICROSOFT R&D (Israel)
THALES (France)
University of Augsburg
(Germany)
University of Cyprus (Cyprus)
University of Manchester (UK)
Start Date: 2010.01.01
Duration: 48 months

Future teradevice systems will expose a large amount of parallelism (1000+ cores) that cannot be exploited efficiently by current applications and programming models. The aim of this project is to propose a complete solution that is able to harness the large-scale parallelism in an efficient way. The main objectives of the project are the programming model, compiler analysis, and a scalable, reliable, architecture based mostly on commodity components. Data-flow principles are exploited at all levels in order to overcome the current limitations.

Main Objectives

Technology trends indicate that by 2020 chips will accommodate teradevice systems of 1000+ cores. The success of these future architectures depends on addressing important challenges such as programming applications to use such large-scale systems, developing compiler analy-

sis and optimizations required for the generation of code, developing appropriate execution models, which can take on both performance and reliability. In addition, the architecture cannot be reinvented from scratch each time. Therefore, it should be composed of commodity modules such as the execution cores and the interconnection network.
TERAFLUX focus is on developing an infrastructure for programming future multicore systems

One of the key aspects of this project is the proposal of a new programming and execution model based on data-flow instead of traditional control-flow. Data-flow is known to overcome the limitations of the traditional control-flow model by exploring the maximum parallelism and reducing the synchronization overhead. Although its benefits are well known and have been presented a long time ago, this model has not yet been fully exploited for commercial systems.

This project represents a unique opportunity to integrate complementary essential aspects from applications through the whole tool chain, encompassing reliability, an appropriate architecture and resource management (that accounts for power, temperature, faults), and to test research ideas in a simulated teradevice system.

Expected Impact

We expect to develop a coarse grain dataflow model (or fine grain multi-threaded model) that will encompass fine grain transactional isolation, scalable to many cores and distributed memory, with built-in application-unaware resilience, and with novel hardware support structures as needed.
Moreover we will provide an open

evaluation platform based on an x86 simulator based on COTSon by TERAFLUX partner HPLabs (http://cotson.sourceforge.net/) that enables leveraging the large software body out there (OS, middleware, libraries, applications).

Technical Approach & Key Issues

• Applications are becoming more and more complex, demand for higher degrees of accuracy, and process larger amounts of data. As such, in this project, we will look at current and emerging demanding applications to be executed on teradevice systems. Their evaluation will allow us to study the limits of such large-scale future systems.
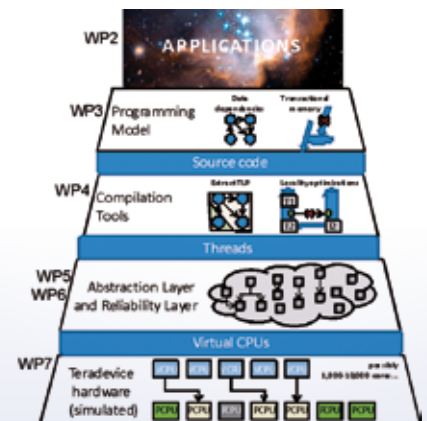
• The programming challenge is how to make the parallel resources easily available to programmers for such large-scale systems. In TERAFLUX, we propose a two-level parallel programming (efficient programming + performance programming) approach. Furthermore, we explore programming models that combine the benefits of data-flow with transactional memory principles.

• Compiler optimizations are required to coarsen the grain of concurrency, allocate memory statically, convert streams into

shared memory buffers, overlap communication and computation, and instrument code with resource and power management probes/actions.

- In modern architectures, reliability is a major aspect for system designers and users. New process technology exposes us to new challenges such as aging, process variability, and soft errors. Defects and errors will dramatically increase in the near future. Therefore, building a reliable system out of unreliable components becomes a major problem for future systems and needs special attention

- Simple technology improvements will allow current designs to scale to 1000+ cores on the same chip. One concern is to keep the system within the required power budget. In order to satisfy that goal and at the same time provide a large degree of parallelism, the TERAFLUX architecture will be composed of heterogeneous multi-cores supporting the same instruction set. Simpler more power efficient cores will provide the parallelism while more complex and less power efficient cores will be used to execute codes requiring Instruction Level Parallelism (ILP).

- An existing infrastructure for full system simulation (COTSon by HP labs) using a MIT license has been chosen as the simulation infrastructure able to provide fast and accurate evaluation of current and future computing systems, covering the full software stack and complete hardware models.

Project Coordinator
Roberto Giorgi

Modaë Technologies Aims at Promoting Dataflow and Process Network Methodologies and Technologies for Application Modeling

The huge challenge of parallel programming in embedded systems engineering that results from the wide adoption of multicore architectures or FPGAs causes problems to the industry because of a lack of competences. Consequently, there is a real need for new methodologies and tools allowing the description of the system at a higher level of abstraction. Actually, parallelism is foremost a matter of application modeling styles. A good concurrent model of the application may profitably help automatic parallelization and hardware/software co-design. However, there are complex relationships between applications and different types of modeling techniques used to describe them that are still not fully understood (Berkeley's Ptolemy project is still alive), and it is

difficult to find a good formalism for catching applicative algorithms with the right model of computation.
Modaë Technologies is a French start-up that is developing methodologies and tools for modeling and exploiting application level parallelism.
The following describes Modaë Technologies innovative technology in a nutshell.
Model Driven Engineering: Basically, the technology aims at creating domain specific abstraction models of concurrent applications.
Algorithm: Nevertheless, unlike traditional modeling approaches, algorithmic and computational descriptions are embedded in the models. This allows for full simulation, analysis and transformation of the model up to hardware and software code synthesis.
Agility and fast prototyping: In order to be useful, modeling must be much faster than lower level programming. The Modaë solution is based on

interpreted languages which provide good properties for the implementation of agile methodologies and speed-up the design phase.
Models of Computation: The monoprocessor Von Neuman model is not central in today's parallel architectures. Other models of computation have to be handled in order to capture the real parallelism of applications. In particular, dataflow and communicating process networks are well suited for signal and image processing.
Applications: Finally, Modaë not only provides tools but also services and intellectual property in the domain of multimedia and telecommunication, one of the team's core areas of expertise.

IETR-INSA Image Group Laboratory of Rennes
The Modaë project was originated by two former engineers from Thomson, in association with the IETR-INSA Image Group Laboratory of Rennes.

The laboratory is working on both automatic mapping and scheduling of Static/Synchronous Data Flow (SDF) application models to multicore architectures and the modeling, analysis and compilation of Dynamic Data Flow (DDF) application models written in Berkeley's actor-oriented language named CAL. The lab is a member of MPEG-RVC, an ISO consortium for reconfigurable video coding standardization that has chosen CAL as the standard language for MPEG video codec's specification. The laboratory has developed open-source tools that the Modaë Technologies team will be able to support and enhance.

Pierre-Laurent Lagalaye, Modaë Technologies and IETR-INSA

Before joining IETR-INSA laboratory as a research engineer, Pierre-Laurent Lagalaye, president and co-founder

of Modaë Technologies, has been working for 8 years in the industry of consumer, mobile and professional multimedia embedded systems.

After having graduated with a degree in engineering at the IFSIC institute of INRIA-IRISA of Rennes in 2001, he has joined the multimedia team of Mitsubishi Mobile Communication Europe R&D center. The team has developed embedded multimedia stacks for audio/video processing,

display and record, video-telephony and RTP streaming on both advanced bi-CPU platforms and ARM with NVIDIA hardware acceleration commercial products. In 2005, he moved to Thomson Silicon Components, a Thomson business unit that aimed at developing ICs for both consumer and professional digital TV equipments. He has lead the simulation & modeling team, achieving architecture models in CABA C, then TLM SystemC. Such ICs typically consist of multiple RISC CPUs, SIMDs and pure hardware accelerators for real-time HD and multi-standard video encoding and decoding. Finally, as software architect and project leader, he has managed a team distributed between Beijing and Rennes Thomson sites while leading the architecture definition, design and verification of a low cost, low power, multi-standard video decoding IP.

Application Profiling and Instruction-Set Customization for Application Specific Instruction-Set Processor Design

By Kingshuk Karuri
(karuri@iss.rwth-aachen.de)
Advisor: Prof. Dr. Rainer Leupers
RWTH Aachen University, Germany
June 2010

Application Specific Instruction-set Processors (ASIPs) constitute a new breed of processing engines that incorporate application specific hardware customizations inside a programmable core. Although ASIPs are attractive processing elements for embedded SoC designs due to their unique blend of performance and flexibility, their widespread acceptance has been greatly hindered because of the high design effort

involved in the initial development of a complete processor architecture for a small set of applications. This work presents a design-flow that attempts to raise designer productivity and consequently, lower the design effort by providing detailed analysis tools for mapping a given set of applications to an initial ASIP architecture. The design-flow is centered around a fine grained application profiler called μ-Profiler which can be used for computational bottleneck identification and micro-architectural analysis, and an Instruction-set Architecture (ISA) customization tool which can identify promising special instructions from a target application's source code.

μ-Profiler is designed to collect various dynamic computational characteristics - such as usage frequencies of different arithmetic/logic/control transfer operations, dynamic bit-widths of various integral data-types, data-cache access behavior, memory access and branching patterns etc. - for a given set of applications in early ASIP development phases. This information can be utilized while making various initial micro-architectural design decisions.

Application specific special instructions, also known as Instruction-Set Extensions (ISEs), are often the primary source of hardware acceleration

in many ASIP cores. The ISA customization tool automatically extracts a set of ISEs by clustering multiple operations from a target application's source code into large instruction data-paths. The customization tool also generates hardware implementations of the identified ISEs that can be easily linked to a variety of existing processor design/customization frameworks creating a seamless application to architecture flow.

The combined tool-chain provides a generic flow for deriving the initial micro-architecture and ISA specification for an ASIP core. Case studies with computationally intensive embedded benchmarks show that the design flow can significantly reduce the ASIP design effort and produce highly optimized initial architectural prototypes.

Early Design Space Exploration of Multi-Processor System-on-Chip Platforms

By Torsten Kempf
(kempf@iss.rwth-aachen.de)
Advisor: Prof. Dr. Gerd Ascheid
RWTH Aachen University, Germany
June 2010

Over the past 20 years, advances in digital wireless communication technologies altered everyone's day-to-day life. Parallel to these achievements, user devices have evolved at an incredible pace over the last years. The main driver behind this trend were technology advances in the semiconductor industry, that have led to supercomputers in the form factor of a mobile terminal. As a result, latest-generation smartphones are no longer limited solely to pure voice communication, but support a wide range of applications from the domains of multimedia, entertainment and infotainment. In turn, these applications have had a particularly strong impact on connectivity requirements, requiring the support for multiple wireless communication standards.

These requirements have created one of the most challenging assignments in engineering today. Looking purely at the necessary computational performance shows an approximate demand of 10 to 80 GOPS peak performance for the execution of today's communication standards. In addition, upcoming standards will further increase the demands, e.g. the upcoming Long Term Evolution (LTE) standard extension. The demand to support the mobility of battery-powered devices makes high energy efficiency one of the key elements for business success within the anticipated market. This demand together with the requirements of low cost, short time-to-market and the extremely short lifecycles put particular pressure on system architects when designing such terminals.

In order to address the design issues of future multi- and many-processor core architectures, with particular attention to platforms in the domain of wireless communication, this thesis outlines a unique early design space exploration framework. Its major contribution is a joint environment that covers several abstraction layers for the purpose of the exploration and evaluation of heterogeneous MPSoC platforms. The framework introduces the main concepts and techniques of an analytical implementation model as well as an abstract simulation model. The analytical model is built on fundamentals of statistical processes and graph theory that help to easily identify whether a system complies with the necessary requirements. Following an effective design process a smooth transition between the models exists, so that the abstract simulation model supports an evaluation of the anticipated system. This paradigm has culminated in the Virtual Processing Unit (VPU) including several extensions for practical use. Already today, major parts of the environment have been included into commercialized tools and have been successfully applied to industrial projects.

Superoptimization: Provably Optimal Code Generation using Answer Set Programming

By Tom Crick (tcrick@uwic.ac.uk)
Advisors: Professor John Fitch
and Dr Marina De Vos
University of Bath, UK
August 2009

Code optimization in modern compilers is an accepted misnomer for performance improvement some of the time. The code that compilers generate is often significantly improved, but it is unlikely to produce optimal sequences of instructions; and if it does, it will not be possible to determine that they are indeed optimal. None of the existing approaches, or

techniques for creating new optimizations, is likely to change this state of play.

Superoptimization is a radical approach to generating provably optimal code that performs searches over the space of all possible instructions. Rather than starting with naively generated code and improving it, a superoptimiser starts with the specification of a function and performs a directed search for an optimal sequence of instructions that fulfils this specification.

In this thesis, we present TOAST, the Total Optimization using Answer Set Technology system, a provably optimal code generation system that applies superoptimizing techniques to optimize acyclic integer-based code for modern microprocessor architectures. TOAST utilizes Answer Set Programming (ASP), a declarative logic programming language, as an expressive modeling and efficient computational framework to solve the optimal code generation problem.

We demonstrate the validity of the approach of superoptimization using Answer Set Programming by optimizing code sequences for two 32-bit RISC architectures, the MIPS R2000 and the SPARC V8. We also present an application of the TOAST system as a peephole optimiser, by generating libraries of equivalence classes of all optimal instruction sequences of a given length for a specific microprocessor architecture.

While this is a computationally expensive process, it only ever needs to be performed once per architecture. We also provide significant benchmarks for the performance of state of the art domain solver tools, further demonstrating the applicability of Answer Set Programming in modeling complex real-world problems.

MARTE Based Model Driven Design Methodology for Targeting Dynamically Reconfigurable FPGA Based SoCs

By Imran Rafiq Quadri
(imran.quadri@lifl.fr)
Advisors:
Professor Jean-Luc Dekeyser,
Assistant Professor Samy Meftali
University of Lille 1, France
April 2010

In our work, we present a novel model-driven design methodology that moves from high abstraction level UML models to dynamically reconfigurable FPGA based SoCs. The abstraction levels offered by Model-Driven Engineering (MDE) permit to decrease system complexity, while allowing to model SoC co-design aspects and enabling their maximum re-utilization. Moreover, the models are not only used for specification purposes. By using model transformations, it is possible to automatically generate execution models or code from these models for final implementation.

We integrate the UML MARTE (Modeling and Analysis of Real-Time and Embedded Systems) profile in our design flow, which is rapidly becoming the de-facto standard by SoC designers both in academia and industry. Moreover, regarding dynamically reconfigurable SoCs, we present an application oriented design flow, as there are already numerous works that focus on optimizing the architectural details at the electronic register transfer level (RTL). The drawback of these approaches is that targeted applications are generally simplistic or nonexistent in nature.

Regarding a reconfigurable system, we focus on two key components, a dynamically reconfigurable area and a reconfiguration controller that enables switching between different available implementations of the reconfigurable area. In order to express these concepts with UML models compliant to the MARTE profile, we extended MARTE to express control semantics and permit Intellectual Property (IP) integration.

At the UML level, we can model a complex data intensive parallel computation application and its several configurations, which are associated with mode automata control aspects. Afterwards, via MDE model transformations and intermediate enriched models present in our design flow, we are able to automatically generate RTL code. More specifically, the modeled application is converted into a dynamically reconfigurable hardware accelerator with several implementations. The implementations correspond to the modeled configurations. Additionally, control aspects permit code generation for a reconfiguration controller that manages the dynamic hardware accelerator and its different implementations.

Our design flow was validated in a case study related to an anti-Collision radar detection system. A key component of the system, a Delay Estimation Correlation Module (DECM) was modeled with the UML MARTE profile along with related configurations and control aspects. The model transformations were able to successfully generate RTL code, which was then taken as input by Xilinx commercial tools, such as ISE and PlanAhead for implementation using the Early Access Partial Reconfiguration (EAPR) flow. Finally, our design flow has been integrated in our research team's GASPARD2 SoC Co-Design environment, which is available at www.gaspard2.org

Runahead Threads

By Tanausú Ramírez García
(tramirez@ac.upc.edu)
Advisors: Mateo Valero Cortes,
Oliverio Santana, Alex Pajuelo
Universitat Politècnica de Catalina
(UPC), Spain
April 2010.

Research on multithreading topics has gained a lot of interest in the computer architecture community due to new commercial multithreaded and multicore processors. Simultaneous Multithreading (SMT) is one of these relatively new paradigms, combining multiple instruction issue features of superscalar processors with the ability of multithreaded architectures to exploit Thread Level Parallelism (TLP). Shared resources are the key of simultaneous multithreading, making the technique worthwhile.

SMT also entails important challenges to deal with because threads also compete for resources in the processor core, thereby hindering overall system performance.

The main goal of this thesis is to alleviate these shortcomings on SMT scenarios. The key contribution is the application of the runahead execution paradigm in the design of multithreaded processors by Runahead Threads (RaT).

The idea of RaT is to transform a memory intensive thread into a light-consumer resource thread by allowing that thread to progress speculatively. Therefore, as soon as a thread undergoes a long latency load, RaT transforms the thread to a runahead thread while it has that long latency miss outstanding.

The main benefits of this simple action performed by RaT are twofold. While being a runahead thread, this thread uses different shared resources without monopolizing or limiting the available resources for other threads. At the same time, this fast speculative thread issues prefetches that overlap other memory accesses with the main miss, thereby exploiting memory level parallelism.

Regarding implementation issues, RaT adds very little extra hardware cost and complexity to an existing SMT processor. Therefore, by means of runahead threads, we contribute to alleviate simultaneously the two shortcomings in the context of a SMT processor, thus improving the performance. RaT-based mechanisms are promising options providing a better performance and energy balance than previous proposals in the field.

Improving the Scalability of Multicore Systems - With a Focus on H.264 Video Decoding

By Cor Meenderinck
cor@ce.et.tudelft.nl
Advisor: Prof. Ben Juurlink
Delft University of Technology,
The Netherlands
June, 2010

In pursuit of ever increasing performance, more and more processor architectures have become multicore processors. As clock frequency was no longer increasing rapidly and ILP techniques showed diminishing results, increasing the number of cores per chip was the natural choice. The transistor budget is still increasing and thus it is expected that within ten years chips can contain hundreds of high performance cores. Scaling the number of cores, however, does not necessarily translate into an equal scaling of performance. In this thesis, we propose several techniques to improve the performance scalability of multicore systems. With those tech-

niques we address several key challenges of the multicore area.

First, we investigate the effect of the power wall on future multicore architecture. Our model includes predictions of technology improvements, analysis of symmetric and asymmetric multicores, as well as the influence of Amdahl's Law.

Second, we investigate the parallelization of the H.264 video decoding application, thereby addressing application scalability. Existing parallelization strategies are discussed and a novel strategy is proposed. Analysis shows that using the new parallelization strategy the amount of available parallelism is in the order of thousands. Several implementations of the strategy are discussed, which show the difficulty and the possibility of actually exploiting the available parallelism.

Third, we propose an Application Specific Instruction Set (ASIP) processor for H.264 decoding, based on the Cell SPE. ASIPs are energy efficient and allow performance scaling in systems that are limited by the power budget.

Finally, we propose hardware support for task management, of which the benefits are two-fold. First, it supports the SARC programming model, which is a task-based dataflow programming model based on StarSS. By providing hardware support for the most time-consuming part of the runtime system, it improves the scalability. Second, it reduces the parallelization overhead, such as synchronization, by providing fast hardware primitives.

The International Symposium on Low Power Electronics and Design 2010 (ISLPED 2010),
18 -20 August 2010,Austin, USA, http://www.islped.org/

The 16th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA 2010),
23 - 25 August 2010 Macau SAR, China, http://conference.cs.cityu.edu.hk/rtcsa2010/

International Conference on Field Programmable Logic and Applications 2010 (FPL 2010),
31 August – 2 September 2010, Milan, Italy, http://www.fpl.org/

The 13th EUROMICRO Conference on Digital System Design (DSD 2010),
1-3 September, 2010, Lille, France, http://www.iuma.ulpgc.es/dsd10/

The 19th International Conference on Parallel Architectures and Compilation Techniques (PACT 2010),
11-15 September 2010, Vienna, Austria, http://www.pactconf.org/

The IEEE International Conference On Cluster Computing (CLUSTER 2010),
20 - 24 September 2010, Heraklion, Greece, http://www.cluster2010.org/

The International Symposium on System-on-Chip 2010 (SoC 2010),
29-30 September 2010, Tampere, Finland, http://soc.cs.tut.fi/

The 28th IEEE International Conference on Computer Design (ICCD 2010),
3-6 October 2010, Amsterdam, the Netherlands, http://www.iccd-conference.com/

The 6th Embedded Systems Week (ESWeek 2010),
24 - 29 October 2010, Scottsdale, AZ, USA, http://www.esweek.org/

The International Conference on Computer-Aided Design (ICCAD 2010),
7-11 November 2010, San Jose, CA, USA, http://www.iccad.com/

The 28th Norchip Conference (NORCHIP 2010),
15-16 November 2010, Tampere, Finland, http://www.norchip.org/

The 43rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO-43),
4–8 December 2010, Atlanta, Georgia http://www.microarch.org/micro43/

The 17th IEEE International Symposium on High-Performance Computer Architecture (HPCA-17),
12-16 February 2011, San Antonio, Texas, USA http://hpca17.ac.upc.edu/

The 10th IASTED International Conference on Parallel and Distributed Computing and Networks (PDCN 2011),
15 - 17 February 2011, Innsbruck, Austria, http://www.iasted.org/conferences/home-719.html

The Design, Automation, and Test in Europe Conference (DATE 2011),
14-18 March 2011, Grenoble, France, http://www.date-conference.com/

HiPEAC Spring Computing Systems Week,
7-8 April, 2011, Chamonix, France, http://www.hipeac.net/

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**

**HiPEAC'11 Conference, January 24-26 2011, Heraklion,
Greece http://www.hipeac.net/hipeac2011**