HiPEAC COMPILATION ARCHITECTURE INFO 49
APPEARS QUARTERLY | JANUARY 2017

Welcome to the future: HiPEAC Vision 2017

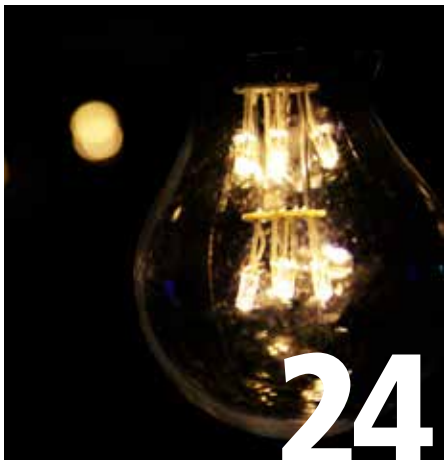
Kathryn S. McKinley on data centre tail latency

Sarita Adve on memory coherence and consistency

contents

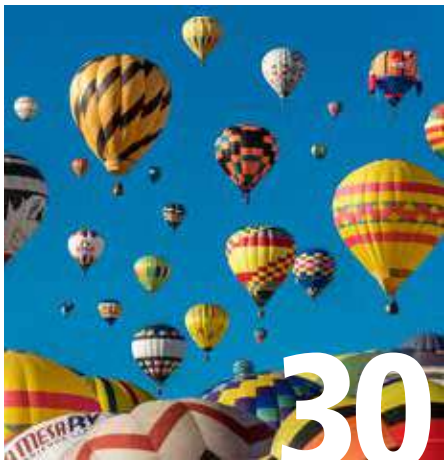
Computing Systems Week in Dublin: making inroads in tech-friendly Ireland

HiPEAC Vision 2017



Innovation Europe

Architecting hyperscale scientific systems



Startup success: Techmo

Career talk: Alessandra Bagnato

HiPEAC is the European network on high performance and embedded architecture and compilation.

hipeac.net

@hipeac hipeac.net/linkedin

HiPEAC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 687698.

Design: www.magelaan.be
Photo credits: image in HLPP 2017 item in News: originally posted by lugarzen in Flickr, www.flickr.com/photos/luisangel/179038600, used under cc-by-sa-2.0 license
Editor: Catherine Roderick
Email: communication@hipeac.net

First of all, I would like to wish you a healthy and prosperous 2017, personally as well as professionally.

Most of us will remember 2016 as the year of Brexit and of the election of Donald Trump as the 45th President of the United States. No matter how disruptive these events seem at this moment, only the future will be able to tell us whether they will be a turning point in history and whether they will have an impact on the computing industry.

The new HiPEAC Vision, which is summarized in this magazine, starts with a quote from 1958 by Giuseppe Tomasi di Lampedusa: *If we want things to stay as they are, things will have to change.* These prophetic words seem to be very appropriate these days. If we want to keep our standard of living in a safe society, we will have to care more about the victims of conflicts, of globalization and of climate change all over the world. If we fail to do so, unsolved problems in distant places will increasingly impact our lives because we are all sharing the same planet, and the resources of the planet are finite. If Europe wants to lead in the era of digitization, industry in Europe will have to innovate faster and it will have to compete with an increasing number of Asian competitors in order to keep its position in the global market. Europe will need to retrain its workforce, it will have to become more entrepreneurial and to invest more in future technologies like data analytics, machine learning, cybersecurity, additive manufacturing and the Internet of Things. And finally, if we want scaling to continue, the electronics industry and research institutions will have to find creative technical solutions to continue scaling for computing, for storage and for communication. If that turns out to be impossible, we better prepare for a future in which scaling ends, and for when the innovations that are enabled by scaling slow down too. If you want to know more about the future after tomorrow, I definitely encourage you to read the HiPEAC Vision 2017.

Many of you will read this newsletter at the HiPEAC conference in Stockholm. The conference is the flagship event for the HiPEAC community. I am thankful to the many volunteers who work very hard to make this a successful and well-attended event, and a sign of a thriving community in computing systems in Europe.

Take care!

Koen De Bosschere, HiPEAC coordinator

"Digitising European Industry" initiative – what's going on, and why does it matter for HiPEAC?

The European Commission's measures to digitize European industry aim to support one of the key pillars of the European economy. Sandro D'Elia, Programme Officer in the Technologies and Systems for Digitising Industry unit at the Commission, explains how the research community can contribute to the new industrial revolution.

I would like to update you on some aspects of the Digitising European Industry initiative (or DEI, for short).

A clearly-stated policy objective of the European Commission for the coming years is to support the introduction of digital technologies in industry. This is needed: European industry cannot grow if it does not make the best possible use of digital technologies to provide innovative and high-quality products and services. In Europe many companies are already at the forefront of technology, but this is not always true for the many SMEs that are behind most of the jobs and value creation across the EU; DEI wants to reach out to them.

Ok, but what's in it for the HiPEAC community? The answer is simple: DEI supports the creation of digital platforms for industry through Horizon 2020 funding.

Using digital tech in industry should mean building on something that already exists and not having to reinvent the wheel for each new problem to be solved. Today, if you want to develop a smartphone app you do not also need to build Android, but if you want to develop an application for a smart factory you first have to build the 'factory operating system', or write a million and one different interfaces for all the different machines that your application needs to control. This is because there are many platforms for consumer apps, but not enough digital platforms for industrial applications.

"There are many platforms for consumer apps, but not enough digital platforms for industrial applications"

The competence and knowledge of the HiPEAC community is needed to build the digital platforms for the systems of tomorrow: maybe not Android, but the equivalent 'operating system' for creating applications in industrial automation, avionics, automotive, energy distribution and public transport. This is therefore a field of research with significant potential for long-term and wide-ranging impact on people's lives.

The ICT-05-2017 call, which will close on 25 April 2017, addresses these issues: we hope to fund a few successful projects to create the software development tools for the industrial platforms of tomorrow, and one hardware project to kick-start the development of next generation processors. Last but not least, the support action which will take over from HiPEAC 4 will have the objective of supporting the building of cross-sectorial industrial platforms.

If you are interested, check the official call information at bit.ly/ICT-05-2017

See also https://ec.europa.eu/digital-single-market/en/digitising-european-industry for a full explanation of the DEI.

Good luck!

Computing Systems Week in Dublin: making inroads in tech-friendly Ireland

In November, HiPEAC took its Computing Systems Week to Dublin. Organized by David Moloney of Movidius and David Gregg of Trinity College Dublin, the event aimed to take advantage of Dublin's blooming tech scene to promote and encourage Irish participation in the network, as well as to provide the usual high-quality forum for discussion and networking.

'While being well known for having a wealth of multinational companies located in Dublin, Ireland has been very much on the periphery of HiPEAC. With a view to getting many of the Irish technology SMEs involved in the HiPEAC network we decided to pitch Dublin as a location for CSW and thankfully, the HiPEAC Steering Committee was very receptive to the idea.' David Moloney commented. 'We put a lot of effort into emailing people in our network, particularly in Ireland and the UK and the effort really paid off in terms of attendance. We also had two excellent speakers in the form of Andrew Fitzgibbon of Microsoft and Ravi Iyer of Intel which also helped to get people to attend. In order to get some additional breadth, we added a session from some local venture capitalists to explain how the funding process works along with some real-life stories which helped illustrate the issues in the process.'

The programme of CSW Autumn 2016 was indeed a varied one, with Thematic Sessions including Technological Challenges to IoT Security, Embedded Deep-Learning and Student Heterogeneous Programming Challenge. On the Deep-Learning session, David Moloney noted that: 'the session was extremely well attended and led to lots of lively debate and questions from across the board, from university researchers to SMEs and multinationals such as Xilinx.'

Putting students to the test: supporting student participation through a programming challenge

Talking about the Student Heterogeneous Programming Challenge, co-organizers Marisa Gil of UPC and BSC and Chris Fensch of Heriot-Watt University explained their motivations for running the session: 'Along with our colleague Georgios Goumas of NTUA, we wanted to offer an activity to help get the next generation of computer scientists and researchers involved in the HiPEAC community,' said Marisa. 'Our main goal is to encourage students to proactively get involved in HiPEAC events so as to access the many opportunities for collaboration, receiving feedback and ideas and sharing their expertise.'

'With this in mind, we wanted to arrange a session which was not only good fun but also offered solid technical training in the form of hands-on experience of trying new approaches to tackle old problems,' added Chris Fensch. 'By opening the session up to student teams from around the world and allowing participation by video link, we ended up with a good mix of participants who we are hoping will encourage others to take part if we are able to offer this activity again, which we really hope to do given the wide-ranging and really interesting ideas that the student teams came up with!'

These sentiments were echoed by Simon Donné of the University of Ghent: 'the Student Challenge at the HiPEAC CSW in Dublin consisted of an isolated, well-defined computing problem that challenged us to push our programming to the limit. The task was to perform physical n-body simulation as fast as possible - no questions asked. There are several different methods, and students were encouraged to discover for themselves which approach they preferred. This led to a wide range of programs, ranging from multi-core CPU calculations to a centralized distributed system incorporating cell phones and tablets - I do believe each of us was introduced to high-throughput computing systems we were not familiar with. Discussion was a big part of the presentation session, which really let us voice considerations and ideas about each other's techniques. I know I took home a lot of new ideas!"

Teams from five universities participated: University of Ghent (Belgium), Rice University (US), TU Dresden (Germany), University of Edinburgh (UK) and Budapest University of Technology and Economics (Hungary).

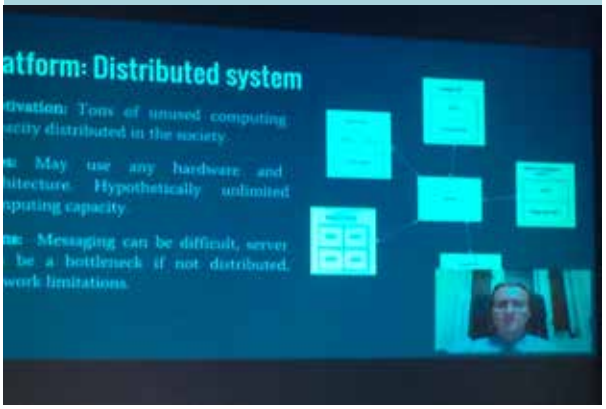


Photo: Marisa Gil

Award: Best Paper of PACT 2016

Scalable Task Parallelism for NUMA: A Uniform Abstraction for Coordinated Scheduling and Memory Management

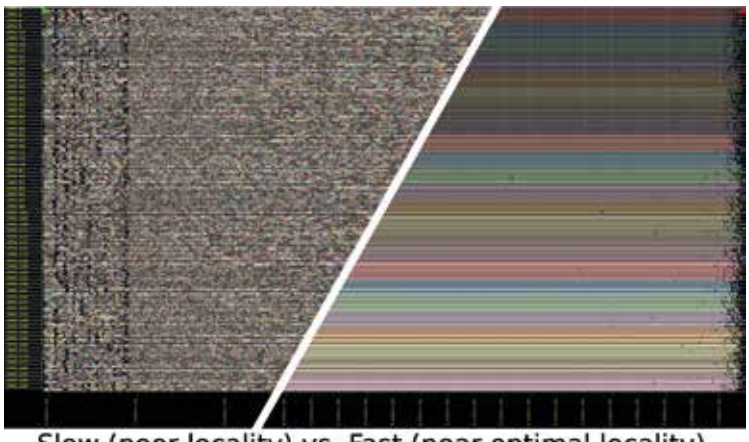
- Authors: Andi Drebes, Antoniu Pop, Karine Heydemann, Albert Cohen, Nathalie Drach

The promise that computing devices would become faster and consume less energy every year has mostly been delivered at the hardware level. From mobile devices to laptops and desktops to servers and data centres, today's computers are capable of more computations and handle more data than ever before. However, this capability is often not translated into faster and more efficient applications for users. This is due to the difficulty of writing programs that take advantage of such additional compute power and memory resources. Indeed, current systems exhibit significant static and dynamic non-uniformity which cannot be managed by programmers.

To tackle this issue, a team of researchers from The University of Manchester's School of Computer Science and the Université Pierre et Marie Curie and INRIA in Paris, has developed a dynamic performance optimization system that allows software engineers to harness the ever increasing power of contemporary computing systems without being exposed to their complexity. In particular, they show that it is possible to preserve a uniform hardware abstraction for programmers, leaving the work of managing non-uniformity to an automated run-time system. The system is automatic, application-independent, performance-portable across machines, and adapts to dynamic changes.

This work received the Best Paper Award of the 2016 IEEE/ACM International Conference on Parallel Architectures and Compiler Techniques in Haifa, Israel. It is anticipated that it will influence the key trade-offs that determine how computing systems are designed and, eventually, lead to significant changes in the way software is deployed and optimized in future computing systems.

Link: http://dl.acm.org/citation.cfm?id=2967946
Contact: Andi Drebes andi.drebes@manchester.ac.uk



Slow (poor locality) vs. Fast (near-optimal locality)
Colors represent the origin (memory node) of memory accessed by each core over time

Collective knowledge on deep learning

A suite of tools for collaboratively optimizing deep learning

$$\frac{d\vec{v}}{dt}$$

Deep learning is the hottest research topic today thanks to many artificial intelligence and computer vision applications. Still, designing and optimizing deep learning applications is a dark art, mainly because of lack of effective mechanisms for knowledge sharing: in spite of millions of individuals active in this area, the community is not necessarily becoming any wiser! This is because, like the blind men touching the elephant in the famous parable, individual observations do not easily add up to a coherent and comprehensive view of deep learning.

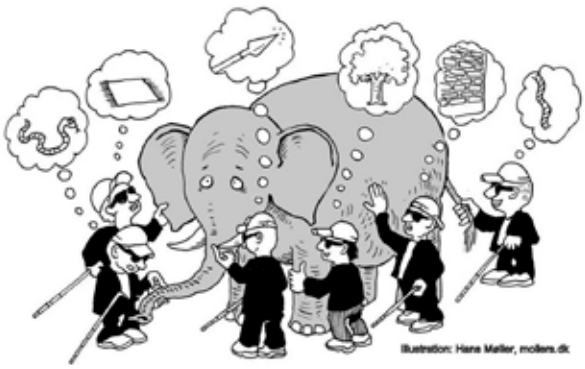
The cTuning foundation and dividiti are introducing a suite of open-source tools for crowd-benchmarking and crowd-tuning of deep learning to meet the performance, prediction accuracy and cost requirements for a wide range of applications and for deployment on a wide range of form factors – from sensors to self-driving cars.

The first such tool is CK-Caffe, developed in collaboration with General Motors and other partners (http://github.com/dividiti/ck-caffe). CK-Caffe is based on the Collective Knowledge framework (CK) from the cTuning foundation and the Caffe framework from the Berkeley Vision and Learning Center (BVLC). CK-Caffe leverages the key capabilities of CK to crowd-source experimentation and perform multi-objective autotuning across diverse platforms, trained models, optimization options, and so on; exchange experimental data in a flexible JSON-based format; and apply leading-edge predictive analytics to extract valuable insights from the experimental data.

Another such tool is an engaging Android app (search for "crowdsourcing video experiments" at http://play.google.com/store/apps). The user can apply different engines (trained models, math libraries, etc.) to classify objects in images and report misclassifications along with a correct category. At the same time, fine-grain execution information gets aggregated in a common public repository (select the "crowd-benchmark DNN libraries using mobile devices" scenario at http://cknowledge.org/repo), and can be viewed by models, platforms, processors and so on.

We are inviting the HiPEAC community to contribute knowledge on optimizing deep learning – from low-level building blocks such as kernels and libraries to high-level blocks such as layer and network designs!

http://cknowledge.org/caffe
Please get in touch via the cTuning group https://groups.google.com/forum/#!forum/ctuning-discussions or by email info@dividiti.com



Flying high in mixed-criticality systems

Modern embedded applications integrate into a single system an increasing number of functionalities with different criticality levels. This trend is expected to increase with the shift that the European real-time industry is facing towards multicores that allow the consolidation of more applications per hardware unit. The potential to reduce SWaP (Size, Weight and Power) costs and increase system efficiency of multicores in real-time systems is challenged by the complexity they bring for validation and verification.

The EU projects CONTREX, DREAMS and PROXIMA, which form the European Mixed Criticality Cluster, collaborate on this *multicore mixed-criticality systems* challenge. These teams' research is being used in wide-ranging domains including automotive, avionics, space and rail transport.

One example of industry take-up of research results is LEOPARD, a four-core LEON-based processor, suitable for measurement-based timing analysis, developed by BSC and Cobham Gaisler within PROXIMA, coordinated by BSC. LEOPARD uses randomization techniques to naturally expose the timing behaviour of jittery resources, such as caches, across runs of the same program, thus releasing end-users from the burden of designing complex testing scenarios that often lack evidence on their coverage of the worst possible execution time.

LEOPARD is suited mainly for the space domain, where LEON processors are widely used, and its capabilities have been satisfactorily evaluated with early use cases from the European Space Agency and Airbus Defence and Space. Cobham Gaisler, already advertising this technology on its website, plans to include it in some of its future processors.

Read more about the Cluster:
http://bit.ly/2goOkqi

Download the Cluster impact brochure:
http://bit.ly/2gaOgYn

MEMSYS 2016

A new kind of conference, the International Symposium on Memory Systems, took place in October 2016, in Alexandria, Virginia

MEMSYS 2016 attendees

The memory system has become extremely important. Memory is slow, and this is the primary reason that computers don't run significantly faster than they do. In large-scale computer installations such as the building-sized systems powering Google.com, Amazon.com, and the financial sector, memory is often the largest dollar cost as well as the largest consumer of energy. Consequently, improvements in the memory system can have significant impact on the real world, improving power and energy consumption, performance, and/or dollar cost. Moreover, many of the problems we see in the memory system are cross-disciplinary in nature – their solution would likely require work at all levels, from applications to circuits.

MEMSYS is a new kind of conference, focusing on the field of memory systems, founded to showcase interesting ideas that spark conversations between disparate groups – to get applications people and operating systems people and compiler people and system architecture people and interconnect people and circuits people to talk to each other about the challenges facing our field.

We just had our second successful meeting in Old Town, Alexandria near Washington DC, with 110 attendees versus 95 last year, so we are happy to see it growing. The conference was organized in one single track including 20-minute paper presentations for both short and full research papers. We enjoyed two and a half days of excellent paper presentations, lively panel discussions, a Hardware Keynote by J. Thomas Pawlowski (Micron) and a Software Keynote by Richard Vuduc (Georgia Tech), and finished up with a spectacular Awards Ceremony.. See below for details of some of the award winners.

More than 25% of the attendees were from Europe and Asia, and because we had multiple requests to expand there, we have decided to hold our inaugural European meeting in Frankfurt am Main this June, immediately following ISC (21–23 June 2017). If you are interested in either MEMSYS Europe 2017 (June) or MEMSYS 2017 (October), please visit the conference website at www.memsys.io for more information.

We look forward to seeing you in Frankfurt!

Bruce Jacob & Kathy Smiley, Conference Organizers

UPMEM RECEIVES THE AWARD OF "STARTUP OF THE YEAR" AT MEMSYS CONFERENCE 2016

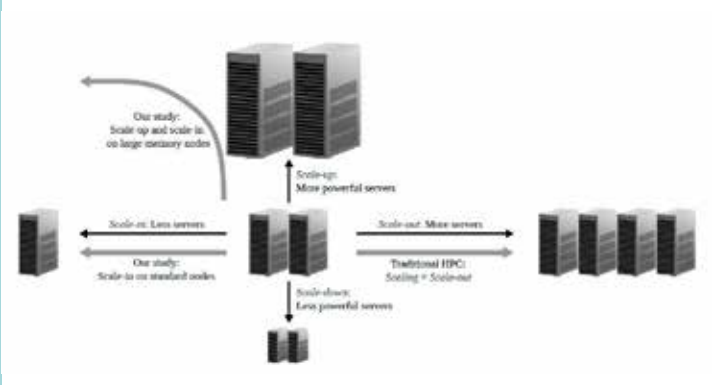
UPMEM is a young startup based in Grenoble, France building an innovative solution that drastically improves the performance of data-intensive algorithms in the data centre. The Processing-In-Memory solution developed by the company is combining DRAM and hundreds of processing units, to enable massively parallel processing of data with unbeatable bandwidth and latency. The technology has been evaluated for multiple use cases such as genomics, column-oriented databases, artificial intelligence and web analytics, showing double-digit performance improvements over conventional solutions.

www.upmem.com

Jef Roy, UPMEM COO and co-founder (centre) with Kathy Smiley and Bruce Jacob, conference organizers

BEST PAPER AT MEMSYS 2016: LARGE-MEMORY NODES FOR ENERGY EFFICIENT HPC

BSC researchers received the award for Best Paper of the MEMSYS 2016 conference with the paper entitled "Large-memory nodes for energy efficient High-Performance Computing" in which they show that a practical energy-saving approach is to scale-in the HPC applications on large-memory nodes. The proposal was evaluated using a set of HPC applications running on BSC's MareNostrum supercomputer. Achieving energy savings of up to 52% means that the investment in upgrading the hardware would be recovered in a system lifetime of less than five years. The study is an important finding of the BSC Memory Systems team using Samsung DRAM memory with the technical support of Samsung Electronics Co., Ltd.



System scaling can be horizontal (scale-in or -out) and vertical (scale-up or -down). Traditionally, the HPC community focuses mainly on scale-out, referring to it simply as scaling. The BSC study analyzes scale-in on standard nodes, and a combined scale-up and scale-in approach on large-memory nodes

Scaling-in not only reduces the energy consumption of HPC jobs, it also decreases the overall experimentation cost, improves system throughput and reduces the execution time for batches of jobs. 'We hope that the study will motivate the community to consider the trade-offs between horizontal and vertical scaling when provisioning and using HPC clusters. Maybe we could start this journey with some second thoughts about the way that we use the word scalability,' says Petar Radojkovic, Memory Systems team leader at BSC.

See the paper at http://dl.acm.org/citation.cfm?id=2989083

MEMSYS 2016: UNIVERSITY OF KAISERSLAUTERN RECEIVES MULTIPLE AWARDS

Matthias Jung, Carl Rheinländer, Christian Weis and Norbert Wehn won the Chair's Choice Best Paper Award with their paper "Reverse Engineering of DRAMs: Row Hammer with Crosshair". In this paper they present a technique that reconstructs the physical location of memory cells in a Dynamic Random Access Memory (DRAM) without opening the device package and microscoping the device. The method consists of a retention error analysis while a temperature gradient is applied to the DRAM device. This enables the extraction of the exact neighbourhood relation of each single DRAM cell, which can be used to accomplish Row Hammer attacks in a very targeted way. However, this information can also be used to enhance current DRAM retention error models.

Furthermore, the presentation given by Matthias Jung, Irene Heinrich and Marco Natale on "ConGen: An Application Specific DRAM Memory Controller Generator" won the award for the most creative presentation. This interdisciplinary work, co-authored by Deepak Mathew and Christian Weis, has been accomplished in a joint cooperation between the Microelectronics Systems Design Research Group of Prof. Norbert Wehn and the Optimization Research Group of Prof. Sven Krumke.

Marco Natale, Irene Heinrich, Bruce Jacob, Kathy Smiley, Matthias Jung

2nd PANDORA Summer School on Progression and Diversity of Reconfigurable Architectures and Tools

"TEAChER" (TEach AdvanCEd Reconfigurable architectures and tools) is a collaborative project between Karlsruhe Institute of Technology (KIT) and National Technical University of Athens (NTUA) and is funded by the German Academic Exchange Service (DAAD) with funds from the Federal Foreign Ministry (AA) through the programme "Hochschulpartnerschaften mit Griechenland". Within this project, we develop educational material for topics related to 2D and 3D reconfigurable architectures. We particularly address, amongst other things, architectural-oriented issues, CAD algorithms and efficient ways for digital design with the usage of advanced programming languages. Due to the varying demands of this domain, the developed material gets appropriately tuned for the target audience, spanning from undergraduate to PhD studies.

The 2nd PANDORA Summer School, which is part of the TEAChER project, was held at KIT from 19-23 September 2016. A selection of 20 German and Greek participants from the partner universities KIT and NTUA attended lectures, talks and hands-on labs led by experts in the fields of reconfigurable architectures, respective CAD algorithms and virtual prototyping tools. The scope of this summer school was not only to handle established and commercial technologies, but also to incorporate recent research results and trends with emphasis on 3D integration and their benefits/opportunities in the reconfigurable domain. A virtual laboratory (http://proteas.microlab.ntua.gr) which is unique to the TEAChER framework was used for lab exercises, as it serves as a perfect platform for quickly developing and evaluating custom FPGAs without the effort and expense of actual physical implementation. The summer school programme included design contests between all the participating students. The winners were six students from KIT and six students from NTUA, who were awarded prizes from Intel and Texas Instruments. Additionally, the organizers would like to thank CADENCE and Microsemi for sponsoring social activities.

Currently, we are working on a concept for the 3rd PANDORA Summer School, which will be open to a larger international target group, either as part of a conference workshop or as a self-contained event.

See http://proteas.microlab.ntua.gr/teacher for more info.

Spin-off success at Silexica: the SLX Tool Suite

Silexica, a spin-off of HiPEAC partner RWTH Aachen University has obtained a US$8 million Series A round of financing. Merus Capital led the investment round with participation from new investor Paua Ventures as well as current investors Seed Fonds Aachen and DSA Invest. The funds will allow the company to expand its capacity and adapt its technologies to additional market segments.

Although Silexica is only two years old, it has enjoyed rapid adoption of its SLX Tool Suite. The software tools use a compiler paradigm to analyze code coupled with a deep understanding of how hardware platforms behave with actual production workload. Global customers like Huawei and Fujitsu are already using these solutions and companies across multiple industries are now embracing the SLX Tool Suite. It has been broadly adopted to automate the distribution of code onto large, multicore platforms and an updated version was released in late November.

You can learn more about the SLX Tool Suite at the *Heterogeneous multicore design automation: current and future* tutorial at HiPEAC 2017 on Monday 23 January at 10.00. Multicore systems with complex architectural features such as heterogeneous processors and sophisticated power management are widely deployed in virtually all computing segments. Not only processors but also memories and interconnects have become heterogeneous. This makes designing software for such systems extremely difficult, considering the huge design space to consider for all kinds of trade-offs such as performance and area/power. The tutorial presents state-of-the-art multicore software design tools (the SLX Tool Suite) in order to demonstrate the current industrial capabilities to address such a challenge. An industrial perspective will also be provided by a leading wireless company, giving a preview of the challenges that will come in the coming five to ten years with the next wave of multicore design. As innovation often comes initially from academia, this tutorial will also discuss research results to give another angle on the same multicore design challenges.

www.silexica.com

SILEXICA

Ghent University hosts FPL 2017

The 27th International Conference on Field-Programmable Logic and Applications (FPL) comes to Ghent, Belgium, 4-8 September 2017.

Reconfigurable logic accelerators currently find their way in data centres to boost performance and power efficiency. More and more data centres employ specialized hardware such as graphics processing units (GPUs), field-programmable gate arrays (FPGAs) and custom application-specific integration circuits (ASICs). For example, the current Project Catapult sees an FPGA integrated into nearly every new Microsoft data centre server. If you want to learn more about how FPGAs are built, how applications can be implemented on them and what the latest evolutions are in the tools that make this possible, you should consider joining us at FPL 2017 in Ghent.

The International Conference on Field-Programmable Logic and Applications (FPL) was the first and remains the largest conference covering the rapidly growing area of field-programmable logic and reconfigurable computing. During the past 26 years, many of the advances in reconfigurable system architectures, applications, embedded processors, design automation methods and tools were first published in the proceedings of the FPL conference series. The conference objective is to bring together researchers and practitioners from both academia and industry and from around the world.

The 27th edition will showcase papers on FPGA architectures and technology, applications and benchmarks, design methods and tools, self-aware and adaptive systems and surveys, trends and education on FPGA design.

Paper submissions are due 26 March 2017 (with abstract and title to be sent by 19 March). Submission instructions can be found at www.fpl2017.org

With keynote speeches by Ivo Bolsens (CTO of Xilinx), Viktor Prassana (USC), David Merodio Codinachs (ESA) and an FPGA specialist at Intel (to be confirmed), this year's FPL will again be an opportunity to learn about FPGAs, and is not to be missed!

The general co-chairs, Dirk Stroobandt from Ghent University and Nele Mentens from KULeuven, and the programme co-chairs, Marco Santambrogio from Politecnico di Milano and Diana Göhringer from Ruhr University Bochum, look forward to welcoming you!

Image: LLVM2017/Apple

European LLVM Developers' Conference 2017

The LLVM Foundation and Saarland University are pleased to announce the 7th European LLVM Developers' Conference on 27-28 March 2017 in Saarbrücken, Germany.

LLVM is an open source compiler framework that is widely used in industry and in academia. The LLVM Developers' Conference is a forum for developers, researchers and users of LLVM and the various tools that have been developed around the LLVM compiler infrastructure. The two days of the programme include technical talks, Birds of a Feather sessions (BOFs), hackers' lab, tutorials, a poster session and a social event.

For the first time, this year's conference has a student research competition! It offers students who do LLVM-related research a non-academic platform to announce and advertise their work, and to discuss it with other researchers, developers and users of LLVM.

The conference targets a broad audience that includes:

- active developers of projects in the LLVM Umbrella;
- anyone interested in using these as part of another project;
- compiler, programming language, and runtime enthusiasts;
- those interested in using compiler and toolchain technology in novel and interesting ways.

It is important to us that the conference is open to everyone, whether from industry or academia, a professional or enthusiast, and it is not restricted to those from Europe. Attendees from all regions are welcome.

For information about the venue, the call for papers, the student research competition, and travel grants as well as other news please visit: http://llvm.org/devmtg/2017-03

If you have any questions or concerns, feel free to email:

- *Arnaud De Grandmaison arnaud.degrandmaison@llvm.org or*
- *the local organizers eurollvm17-orga@cs.uni-saarland.de*

We are looking forward to seeing you in Saarbrücken!
The EuroLLVM 2017 organization team

HLPP 2017

10th International Symposium on High-Level Parallel Programming and Applications
10-11 July 2017, Valladolid, Spain)

lugarzen/Flickr

As processor and system manufacturers increase the amount of both inter- and intra-chip parallelism, it becomes crucial to provide the software industry with high-level, clean and efficient tools for parallel programming. Parallel and distributed programming methodologies are currently dominated by low-level techniques such as send/receive message passing, or equivalently unstructured shared memory mechanisms.

Higher-level, structured approaches offer many possible advantages and have a key role to play in the scalable exploitation of ubiquitous parallelism.

Since 2001 the HLPP series of workshops/symposia has been a forum for researchers developing state-of-the-art concepts, tools and applications for high-level parallel programming. The general emphasis is on software quality, programming productivity and high-level performance models. The 10th Symposium on High-Level Parallel Programming and Applications will be held on 10-11 July in the cultural city of Valladolid, Spain.

Find more information about specific topics of interest, submission details, venue, and organization, on https://hlpp2017.infor.uva.es/

Matryx Computers: Pre-integrated FPGA-based computers and OS for connected devices

MATRYX computers

Embedded Computing Specialists (www.ecspec.com) is pleased to launch its new branch of activities, Matryx Computers, at the HIPEAC 2017 Conference. Matryx Computers will provide pre-integrated (hardware and fully-featured OS) computer platforms based on FPGA. The first product released by Matryx Computers is Matryx OS, a complete Linux-based distribution for Xilinx Zynq FPGAs.

Matryx OS is a complete turnkey solution for connected devices such as those in Machine-to-Machine and Internet-of-Things applications. The OS integrates connectivity services and middlewares which are usable right out-of-the-box on a set of predefined Matryx development boards. It also eases the remote management of distributed devices as it offers solutions to monitor their status, to control their configurations, or even to perform automatic Over-The-Air updates. Matryx OS comes with many other advanced features including fully integrated remote platform management which offers a centralized solution for platform power and thermal monitoring, devices management and FPGA configuration. Such features have become essential for the building of reliable connected solutions.

Matryx OS provides a powerful solution enabling teams to focus on their applications and avoid losing time on the maintenance of custom software stacks, which can be tricky to maintain, update or secure. Indeed, Matryx OS will provide updates on a regular basis, through minor and major releases, which will offer new features, bug fixes and security updates.

Development boards will be available at the end of the first quarter of 2017 for prototyping and evaluation. The Matryx development boards are complete computing platforms built around the Zynq System-on-Chip, with pre-compiled binaries and packages to provide a rapidly deployable solution to development teams.

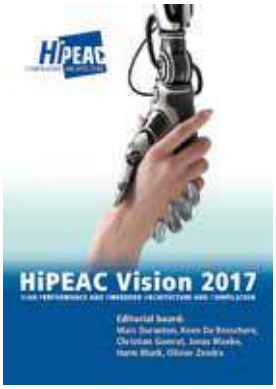
In the future, Matryx Computers will release fully integrated computer modules with extensions (e.g. storage solutions, communication devices, high-throughput digital interfaces, etc.). Those extensions will come with full support from the OS, which will provide well-maintained, readymade solutions.

If you are interested in a rich-feature set OS for Zynq and FPGA-based computers, please contact us: contact@matryx.io

Embedded Computing Specialists

Every two years, HiPEAC produces a technology roadmap, the Vision, which summarizes the role that technology play in our lives, societies and economies, and recommends to the European research community what we need to do in order to address challenges and to keep Europe competitive on the global scene. Here, Editor of the Vision, Marc Duranton, sums up the key findings of the roadmap. We then meet Kathryn McKinley and Sarita Adve, keynote speakers at HiPEAC 2017, before hearing about the future of some key areas of computing systems from three very different perspectives.

HiPEAC Vision 2017



Information technology is one of the cornerstones of modern society and it is evolving rapidly: while the main challenges identified in the HiPEAC Vision 2015 remain valid and have even increased in importance, new challenges are ahead of us.

Computers are disappearing from view. They are taking on new forms, such as cars, smart meters, thermostats, and so on. They communicate with their users using voice, sound, pictures and video, closely resembling human interaction. We are entering the Artificial Intelligence era. This will not only change how we interact with machines, but it will also redefine how we instruct a machine what to do: less programming and more learning.

The function of the computer is shifting from carrying out computational tasks to provide answers to numerical problems, to working together with humans (what we call the beginning of the Centaur Era), augmenting reality to assist us, or even creating virtual worlds for us to explore: the **cyber-physical entanglement between the physical and virtual world**.

Computers will increasingly interact with the physical world, leading to a shift from security to safety. Humans need to trust both the machines and the information that they keep about us, and therefore enforcement of security and privacy is of paramount importance.

For compute-intensive tasks, we will continue to use the cloud; this means that connectivity is crucial, yet local processing is becoming increasingly important. The increasing computational requirements are making computer system architects look for accelerators for specialized tasks, diverting in many cases from the traditional Von Neumann architecture.

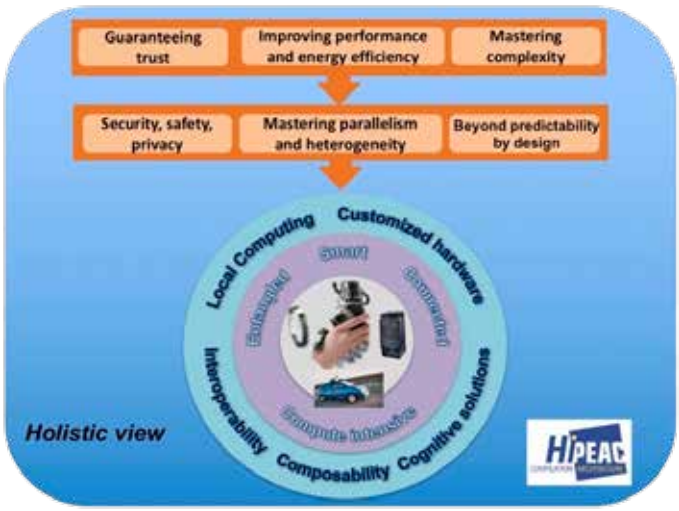
Energy efficiency of computing systems remains a major challenge for the coming years.

As the cost per transistor is no longer decreasing, we might see diversified paths for using silicon technology: many designs will not use the latest technology node, but the more mature (and cheaper) one. It is also the right time to revisit the basic assumptions in order to open new tracks and approaches and to eventually **reinvent computing**.

With the flood of new systems and new system architectures, increasing attention must be paid to composability and interoperability between systems. The complexity of the new systems will be so high that human designers will only be able to master it with the help of computers using AI-based techniques. Innovative approaches will be required to ensure that the systems will do what they are supposed to do, both at the functional and at the non-functional level (e.g. timing requirement or reliability). We need to develop design techniques that go beyond predictability by design and allow the building of reliable systems from unreliable parts.

Finally, holistic approaches, implying **multi-disciplinary techniques, will be needed in order to meet all the requirements of trustability, efficiency and cost**.

Download the HiPEAC Vision 2017 at www.hipeac.net/vision



Dr Kathryn McKinley, Principal Researcher at Microsoft: Taking Tail Latency Beyond the Limits of Queuing Theory

Before joining Microsoft, Dr McKinley was an Endowed Professor of Computer Science at the University of Texas at Austin. She is interested in creating systems (programming languages, compilers, runtimes, and architectures) that make programming easy and the resulting programs correct and efficient. She has graduated 22 PhD students. Her research awards include the ACM SIGPLAN Programming Languages Software Award and Best & Test-of-Time awards from ASPLOS, OOPSLA, ICS, SIGMETRICS, IEEE Top Picks, SIGPLAN Research Highlights, and CACM Research Highlights. She is an IEEE and ACM Fellow.

What attracted you to study computer science initially? What then drew you to your specific field of research?

I took my first programming class as a sophomore at Rice University, and I loved the problem solving and experimentation aspects. You could try out different solutions and see what worked best. The summer of my junior year was my first research experience – building a simulator for local area networks, which hooked me on research!

What achievements are you most proud of in your career to date?

One of the things that makes my job so satisfying is working with great colleagues to produce excellent technology.

With respect to technical accomplishments, our research in memory management continues to influence industry and academia. The Immix mark-region garbage collector remains the fastest garbage collector in the literature and is increasingly being adapted by industry (e.g., Haxe and Rubinius). The Hoard memory manager introduced provably scalable explicit memory management and is widely adopted by industry. Our performance analysis and DaCapo benchmarking work has influenced how researchers and industry evaluates managed runtimes.

With respect to collaborators, I am proud of helping scores of undergraduates, graduate students, other collaborators, and my 22 PhD students achieve their professional goals. As a CRA-W Board Member, I have mentored over a thousand women and minority graduate students to succeed and stay in computing research.

For me, the most exciting part of working in industry is that good ideas can quickly go into practice, and already I impacted Microsoft software that millions of people use.

The gender balance in computing is poor in both the US and Europe. Do you have any advice for girls and young women thinking of entering the field of computer science? What could schools do better to help existing efforts such as ACM-Women and the Anita Borg Institute redress this imbalance?

Computer science is great choice of career with lots of opportunities to impact the world. Every day there are new problems emerging that require creative solutions that make my job interesting, challenging, and rewarding.

Every high school should require a high quality computing series. Equal access would be a great way expose lots of people to the joy and power of computing. I am on the CRA-W board and we have great programs for undergraduates and graduate students to give them the skills and experiences they need to succeed in research. These programmes have been proven to help retention and inspire participation in computing research. I recommend that Europe and Asia start some similar programmes!

Your keynote speech at HiPEAC 2017 looks at improving response times to user requests in large data centres. What do you consider to be the main challenges for this aspect of data centre operation in the coming years?

Efficiency on heterogeneous hardware! In the current power-limited era for designing new hardware, hardware must better match the needs of software to create added value. This technology push requires hardware software cooperation at deep levels. The software and hardware communities have not needed this level of integration historically, which is part of the reason the field thrived. Both could innovate independently and yet still drive and take advantage of each other's advances. Software is still changing really quickly which may make programmable FPGAs economically viable in datacentres and beyond. Microsoft is investing heavily in this direction. However, the challenges of creating a tool chain to automatically configure hardware for software is still daunting. My talk will focus on the shorter view though. Given what architectures we have now, what can software do to exploit it to optimize the 99 percentile response time. We have spent decades optimizing average latencies, but tail latencies require different techniques.

In your opinion, which three key technology trends or research areas which will have a major influence on your work in the next 5-10 years?

Data centre workloads are and will continue to be dominated by interactive web software. Their requirements will drive innovations in hardware and software, spanning programming systems, runtimes, and hardware. My keynote will focus on why I think this topic is important now and will continue to provide a lot of research challenges in the future.

The other major trend my work is focused on is programming with estimates. Software increasingly consumes, reasons about, and produces estimates, e.g., sensors, machine learning, and telemetry to name a few such applications. Current programming systems encourage developers to treat all data as facts. These applications require new programming abstractions, compilers, and runtimes to help developers correctly reason and manage estimates. I think we have to expand over Knuth's definition of a program as an algorithm and data structure to define programs as data, algorithms, and data structures, and treat data in software as first class.

You testified to the House Science Committee in 2013. Can you give us a brief overview of the topics you talked about? Do you think governments and policymakers pay enough attention to scientists and researchers?

The US has a rich history of investing in science, but investment is not keeping pace with inflation or societal needs. My testimony focused on how government investments, University research, and corporate advanced development and research create a virtuous cycle of economic activity and billion dollar new business to encourage continued and increased investment in science.

What benefits do you think the HiPEAC conference offers attendees?

Scientific excellence and new ideas!

"Datacentre workloads are and will continue to be dominated by interactive web software. Their requirements will drive innovations in hardware and software, spanning programming systems, runtimes and hardware"

Kathryn's speech on Taking Tail Latency Beyond the Limits of Queuing Theory takes place as part of the HiPEAC 2017 conference on Monday 23 January at 09.00 in Conference Hall C1

Sarita V. Adve, Richard T. Cheng Professor at the Siebel Center for Computer Science at the University of Illinois at Urbana-Champaign: Coherence, Consistency & Déjà vu: Memory Hierarchies in the Era of Specialization

Professor Sarita Adve is Richard T. Cheng Professor at the Siebel Center for Computer Science at the University of Illinois at Urbana-Champaign. Prior to that, she was on the faculty at Rice University. She is a Fellow of the ACM and of the IEEE and currently serves on the board of directors of the Computing Research Association (CRA) and as the chair of ACM SIGARCH. Her research interests are in computer architecture and systems, parallel computing, and power and reliability-aware systems.

What made you embark on a career in computer science? How did you end up in your specific field of research?

Growing up in India, I always had a keen interest in mathematics and physics. I applied to study engineering at university as it seemed a popular subject at the time for those interests. I didn't really know what I was getting into but it turned out really well! Within my electrical and computing engineering degree programme, I particularly enjoyed the computing courses and it all went from there. I then did my graduate studies in computer science and now, all this time later, I am still working in the field that my undergraduate research project looked at.

What achievements are you most proud of in your career to date?

I think it would have to be my technical contributions in the area of memory consistency models that I have been working on for over 25 years now. The memory consistency model, or memory model, presents some really tricky challenges as it affects both hardware and software. My background is really in hardware but I realised that this is a hardware-software issue and so looking at the situation from a software point of view gave me an edge. So I guess one of the things I am proudest of is having brought the hardware and software communities together, and then getting people to adopt the approach I argued for in my PhD thesis, the data-race-free-model. The latter took over twenty years, and I was really pleased to have been able to get my approach adopted in the programming language community, which wasn't initially my area.

On a personal note, I felt proud – and a little overwhelmed – when I was made a Fellow of the ACM. Although a huge honour, it didn't initially feel like something particularly remarkable, given that I know a lot of ACM Fellows but then an old college friend emailed me to congratulate me and pointed out that I was the first woman of South Asian origin to receive this award. This was a real eye-opener. Although a lot of women in India pursue computing at undergraduate level, remarkably few are recognized at senior levels. Women in computing are still very much in the minority worldwide and it is high time this changed.

On that subject, do you have any advice for girls and young women thinking of entering the field of computer science? What could universities, for example, do better to help existing efforts redress this imbalance?

The first thing I'd tell girls is that this is an exciting field. These days, computer scientists can do such amazing things and the breadth of the field is enormous. Plus, it doesn't

have to be a strictly scientific discipline: it can be combined with practically any other field of study, even art or history. The applications of computing are huge, varied and ever-increasing. Secondly, I'd tell young women not to let anybody tell them that they can't do it. Women can be tough on themselves when they make mistakes or things don't turn out as planned, but you shouldn't let your failures get you down. Computer science is like anything else; you need to work hard and get over your failures but, with enough passion for the subject, you can do amazing things.

In terms of getting more women into computing in the first place and then keeping them there, there are organizations, such as ACM-Women, CRA-W and the Anita Borg Institute, already doing a great job at national and international levels. A lot can also be done at the local levels. For example, my own Computer Science department here at the University of Illinois at Urbana-Champaign has made some real progress. In fact, the recent incoming freshman year has almost gender parity which is something we're really proud of. The department runs summer camps for middle- and high-school girls where they get the chance to meet and chat with our students. The camps are good fun and really help bust the stereotype of the male computer scientist! The department's Women in Computer Science group is really active and does a lot of outreach activities with schools and new freshman year students. Even local-level efforts like these make a big difference.

Moving on to your keynote speech at HiPEAC 2017, which looks at memory hierarchies in the era of specialization, what do you consider to be the main challenges in this respect in the coming years?

The Moore's Law era is coming to an end so we can no longer expect the exponential performance improvements that we've had over the last few decades unless we do something very interesting with architecture.

One way to ensure we continue to obtain better efficiency and performance is to adopt specialization or heterogeneous computing. In this respect, lots of attention has been paid to compute, but little to memory hierarchies. In the future, I believe the action will be in reducing data movement and making memory systems and communication far more efficient than they are today. Doing this effectively will not only require innovation in hardware but will also require ensuring that the systems we design are easy to program. Hardware-software co-design has been a buzzword for a long time, but now these communities truly need to work together to ensure we develop the appropriate solutions. In my group, we have been working on the DeNovo project where we have developed software-inspired solutions for coherence and consistency that can provide better performance and energy efficiency than the state-of-the-art employed today in homogeneous and heterogeneous systems, while retaining design simplicity. We have also developed architectures that enable the energy benefits of specialized memory structures such as scratchpads while giving the programmability benefits of a global address space as seen in the more general-purpose caches. The message of my talk is going to be that hardware designers need to be cognizant of software requirements. By doing so, we can have our cake and eat it too: we can build systems that provide high performance and energy efficiency, and reduce design complexity, all while being easy to program.

"Our recent incoming freshman year has almost gender parity … we run summer camps for middle- and high-school girls where they get the chance to meet and chat to our students. The camps really help bust the stereotype of the male computer scientist!"

Sarita's speech on *Coherence, Consistency & Déjà vu: Memory Hierarchies in the Era of Specialization* takes place as part of the HiPEAC 2017 conference on Tuesday 24 January at 09.00 in Conference Hall C1

Optimization of Big Data applications

Vicent Sanz Marco, Lancaster University

Over the last decade there has been an exponential increase in the volume of digital data produced, leading to what is known today as 'Big Data'. The enormous number of bank transactions and all the information generated on social networks, blogs and websites every day are examples, among many others, of the vast quantities of information and data being collected. Studies on this confirm that the amount of digital information generated every day is equivalent to all of the information that has ever been written on paper. Therefore, the ability to process and analyze large data sets is essential. Building software and hardware systems that can process Big Data with high throughput is a non-trivial task. As a result, large-scale data processing frameworks, such as Apache Spark or Hadoop, are emerging to help with the storage, processing and analysis of this vast volume of data. However, effective scheduling of application tasks that use these frameworks remains an outstanding challenge.

We are seeing an ever greater number of conferences and workshops dealing with Big Data resource matters. One example of this is the recent thematic session "Systems Support for Big Data Applications" at HiPEAC Computing Systems Week, Dublin 2016. This session focused on up-to-date strategies for the creation of systems supports to help Big Data frameworks improving their performance; presentations tackled ways to improve the Big Data utilization of hardware and software within a system so as to improve information analysis.

Within this field, a large body of works focused on the resource optimization of Big Data applications using domain-specific knowledge has been developed. Prior work in this area includes query optimization, graph or data flow optimization and compiler-based optimization. Nevertheless, numerous techniques have been proposed to manage memory resources of Big Data applications independently of the domain-specific knowledge.
For instance, at Lancaster University, where I am currently working as a Research Associate, we are working on the amelioration of memory resource for Big Data applications. Common practice often requests the maximum resource allocation from a host for each application; however, this is often far beyond what is actually needed, wasting capacity that could be used for other tasks. As a part of the UK government-funded SANDeRs project (Smart, Adaptive Compilation for Dark Silicon (grant no. EP/M01567X/1)), we are aiming to develop a machine learning-based approach to optimize Big Data applications for performance and energy consumption on heterogeneous systems. For this reason, we created a novel approach to co-locate Big Data applications on a computing host to improve throughput without compromising application performance. This is achieved by training a predictive model on a range of typical applications; we combine this with feature-based clustering to then automatically determine the memory requirements of new applications. By accurately predicting the resource level that is actually needed, we can determine how many applications can be co-located on the same host to improve the system throughput.

In conclusion, Big Data has produced a unique moment in the history of data analysis. The huge growth in Big Data utilization means that companies need to analyze huge volumes of information quickly and cost-effectively. For this reason, more and more research is focusing on this topic, aiming to optimize Big Data applications in terms of efficiency, productivity and profitability.

Automatic tuning of applications using Machine Learning: a possible near future

Amir H. Ashouri, Politecnico di Milano

Moore's Law states that transistor density doubles every two years. However, the rate compilers which are a key tool to exploit that potential improvement have not been able to improve by more than a few percentage points each year due to the range of as yet unsolved research challenges that we face. Entering the 'post Moore's Law' era, compilers struggle to keep up with the increasing development pace of ever-expanding hardware (e.g. CPUs, GPUs, FPGAs) and software programming paradigms (OpenMP, MPI, OpenCL, and OpenACC). Additionally, the growing complexity of modern compilers and increasing concerns over security are among the most serious issues that the compilation research community faces.

There has been something missing within this process: intelligent tuning of applications. This would be a way to automatically understand the application and choose the right set of modifications it needs to perform better under different scenarios. This brings us to the field of autotuning and the possibility of using machine learning based techniques, such as Bayesian networks or deep learning, to intelligently fill the aforementioned gap. In general, autotuning is the application of automatic optimizations by using different scenarios and architectures. It constructs techniques for automatic optimization of different parameters in order to maximize or minimize the satisfiability of an objective. Recent software systems have many different parameters. Think about these parameters like software knobs on your radio app. You can turn them right and left to tune the sound quality or simply switch to a different channel. What if your intelligent tuner could rapidly find all available channels with crystal-clear sound in a matter of seconds? That would indeed be joyful to have when you develop new software and you target a new architecture to run your code on. A dream come true!

Fast forwarding to the near future, I can see such intelligent systems capable of tuning your code, suggesting a few modifications to you, changing some parameters within the compiler you are using and mapping them all to the right computer hardware so that it runs smoothly, fast and securely, all in a matter of seconds. Deep learning algorithms will come into play in a much broader way to find further similar patterns of interest and full-stack suggestions of code-snippets that you might want to add or remove so as to get even better results. It will be like dreaming of having something and then boom!, it is there ready for you to use.

The ACM and HiPEAC communities: increasing diversity and getting better results

ACM Women in Computing Europe will host a Birds of a Feather session at the HiPEAC 2017 conference. Bev Bachmayer and Virginia Grande lay out some of the challenges facing the move towards a better gender balance in our field and what we can all do to benefit from a more diverse and strengthened community.

Increasing diversity in the technical world is a lofty goal which keeps many of us busy both weekdays and weekends. As a team of committed volunteers, we seek to raise visibility of technical women and improve diversity levels in STEM (science, technology, engineering and mathematics) fields in Europe. In some European countries, diversity in Computer Science or Informatics programmes at universities seems to be improving, whereas in other countries diversification is stagnant[1]. Across Europe we all need to invest in encouraging women to explore technical fields and, at the same time, invite our male counterparts to join us as a community to contribute to effecting the change. Four years ago, when the ACM-W Europe executive committee was chartered, we had the lofty goal to have an impact for an improvement by 2020. Now in the first months of 2017, we fear that time is running out and so we need your support to help our daughters, friends and colleagues to gain access to this exciting, stimulating and in-demand field of work.

What is ACM-W Europe and what are the challenges?

The vision of ACM-W Europe is:

"a transformed European professional and scholarly landscape where women are supported and inspired to pursue their dreams and ambitions to find fulfillment in the computing field."

The questions we ask are: are we making progress? How can we make a bigger impact? What else must we do to really make a difference for the young women entering the field today?

ACM-W Europe drives many activities to improve the situation: from advocating diversity, to promoting visibility, to raising awareness, we collaborate with partner organizations and provide opportunities for technical women of all ages to network, collaborate, and encourage each other. These efforts are dependent upon our male colleagues coming to the table and working with us to remove the barriers. It takes a collaboration among all community participants to bring awareness, raise visibility and alleviate barriers to equality.

One effort specifically designed to address the diversity problem is ACM-W Europe working together with the HPC community in many ways. Possible steps include:

- recognizing the value of diverse teams,
- addressing challenges that members of underrepresented groups face in the community,
- increasing visibility of the accomplishments of women,
- and providing networking and collaboration opportunities

It has been shown that increased diversity improves the bottom line, makes innovative projects more innovative and teams more successful[2]. With the number of tech jobs set to grow by 17% between 2014 and 2024, we will fall short in the number of eligible candidates unless we initiate a strong cultural interest in STEM, like in China, where 40% of the STEM workforce is women. We also know that diverse teams file 40% more patents than all-male teams[3], therefore demonstrating more innovation. In Europe, we must start changing the culture, encouraging STEM studies at a younger age and therefore providing enough workforce for the increasing number of STEM jobs.

Addressing the challenges that face women once they enter the workforce in a STEM position is a critical action that every company and academic institution must address to maintain a good level of diversity. Harvard Business Review[4] reports that 45% of women in science and technology jobs tend to leave within one year because of a reported hostile environment. What we observe is that 'men and women need to be on a level playing field in their work environment in order for men and women to work together productively'[2]. When people feel uncomfortable or undervalued by others in the workplace, they tend to remove themselves, reducing productivity and leaving gaps in the working pipeline.

"Diverse teams file 40% more patents than all-male teams, therefore demonstrating more innovation"

To bring awareness to the challenges underrepresented people face in a male-dominated environment, we need to enter a discussion where we address the unconscious biased actions that occur in the community whether online or in person.

Addressing diversity through Birds of a Feather sessions

One such example of a discussion about how to increase diversity occurred during the EuroBSD conference in Belgrade, in September 2016. As with the HiPEAC community, the BSD community is male-dominated and almost all interactions between community members occur online. The members only see each other face to face perhaps once a year. As in any open source community, all members work together on specific portions of the operating system, but even in this environment unconscious bias sometimes places blockades in the productive work environment. Using a teaching-by-cases approach, ACM-W Europe members worked with a full room of engineers to determine solutions to the scenarios presented in the cases. The discussion was enlightening for both the women and the men participating in the session. This open discussion with an honest exchange of ideas and thoughts on solutions. Several of the male attendees stated that discussing these issues openly and in a trusted environment helped their understanding. The female attendees felt the discussion created an environment to listen to each other and discuss solutions. This discussion enabled us to take strides in reducing the number of challenges that underrepresented members must overcome.

Celebration of Women in Computing events

In December of 2012 when ACM-W Europe was formed, the committee created a plan for a new networking event for Women in Computing in Europe. That idea, after much work and effort, became the highly successful ACM Celebration of Women in Computing: womENcourage, which is now in planning for the fourth annual edition which will have

an HPC focus and will be held in Barcelona Spain in September 2017. Networking is the top priority during a Celebration of Women in Computing, the role models give inspiration and seeing the multitude of technical women is a highly-valued benefit from the event. Many technical women work for years as the only woman in a team of twenty or thirty men. When gathered together in one room, women are astonished at the number of other technical women who are out there! Through networking, exchanging of ideas, discussing projects and sharing their work on posters, they are inspired and become encouraged to continue their work. This effect of inspiration and encouragement occurs not only at the large Europe-wide Celebration of Women in Computing: ACM also encourages smaller Celebrations to support women at a local level.

ACM Celebrations are often referred to as a 'regional' within ACM circles. These are Celebrations of Women in Computing events held in a geographical location or country. It is common that one university – or, to be more precise, one school of a computing-related field – hosts the event, and invitations are open to students in that geographical area, regardless of their affiliation. The number of attendees ranges from 50 to 300. Nowadays, we hold Celebrations in which professionals and students mingle and present their work, sometimes using the local languages other than English. This is particularly attractive in Europe, where events in Greek, Spanish, and Azerbaijani (to name a few) celebrate not only the work of women in different European countries but also the richness of the variety of cultures in the continent.

"When gathered together in one room, women are astonished at the number of other technical women who are out there!"

Both womENcourage and regional celebrations have affordable (when not free) registration fees for students making it much easier for them to attend the celebration. Another initiative with the same goal are the travel scholarships to help fund the trip expenses, thanks to ACM-W and the supporters of ACM-W Europe. ACM-W Europe and the travel scholarship recipients are very thankful to our supporters who provided funds for travel scholarships for womENcourage 2016 in Linz Austria: SIGPLAN, Google, Oracle Academy, Bloomberg, Informatics Europe, Intel GmbH and FreeBSD.

Regional celebrations provide plenty of opportunities to network, something that the attendees constantly express as being one of the best things about these events. They can also be stepping stones, giving women a platform to showcase their work and adapting it for bigger events by using the valuable feedback from attendees.

These Celebrations not only encourage the attendees but also members of the organizing team. Important roles, such as Conference Co-Chair or Program Committee Co-Chair are given to PhD students or young researchers.

Nowadays many institutions around the globe are desperately looking for strategies to broaden participation in computing and improve retention. Having women in computing attend ACM Celebrations is a small expense for a university to fund and, in return, they receive a more confident and inspired coworker, who has built a network of like-minded colleagues.

Importance of chapters: networking

While ACM Celebrations give attendees the opportunity to broaden their network in terms of peers and role models over one or two days, the ACM-W Chapters extend this at a local level. Chapters are created by enthusiast volunteers who would like to have an impact in their communities[5].

A chapter is linked to an institution (for student chapters) or a regional area (for professional chapters). Both kinds of chapters have a Chapter Board that includes the Chair, Vice Chair and Secretary/Treasurer. Other roles are usually added, such as Webmaster, Communications Officer or Mentorship Coordinator. For student chapters, there is also a faculty advisor: a faculty member who supports the students and ensures the continuity of the chapter by helping bring in new members (as students graduate).

Chapters organize a range of different activities: visiting high schools to encourage girls to pursue careers in computing, and running mentoring programmes for students and professionals, career fairs, social events such as pizza coding nights, or gatherings of women researchers, to name a few examples. Women in industry and academia are invited to chapter events to talk about their work and inspire others, and panel discussions on gender issues are held to raise awareness. Overall, chapters support the ACM-W mission in their communities. This involvement in the organization greatly develops not only essential skills, such as leadership and planning, but also the volunteer's network.

This network is not only local: it has an international component as well. ACM Europe organizes the ACM Europe Chapters Workshop, where chapters of all kinds from all over Europe get together to share best practices and establish collaborations. The picture shows the participants of the last workshop, held in Prague this August. Chapter representatives learned how different chapters work and they met ACM volunteers and staff (such as Past President Prof. Alexander Wolf and ACM CEO Bobby Schnabel). These events and meetings show the potential found in collaborations between professionals and students in our field. Across Europe, all kinds of volunteers work together on the ACM vision: advancing computing as a science and profession.

So, what's next?

Advancing computing as a science and profession, and particularly the role that women have in this endeavour, requires the collaboration of all of us. Not only women, but also men are crucial. Regardless of your gender, there is something you can do: celebrate women in computing (by attending an ACM Celebration) and raise awareness about gender issues in computing (through your local student or professional chapter).

A great way to get started is to attend our Birds of a Feather session on diversity at the HiPEAC 2017 conference on 25 January in Stockholm, where we will have a lively discussion on improving diversity in the HiPEAC community. ACM also invites you to attend our European Celebration of Women in Computing: womENcourage 2017 in September in Barcelona, Spain, where we will have keynotes, technical talks, a hackathon, workshops, posters and sessions for networking.

ACM-W Europe (ACM Women in Computing Europe) is a standing committee of ACM Europe. It is setup to carry out the ACM-W mission within the scope of the ACM Europe strategy, in ways that effectively address idiosyncrasies and complexities of supporting women in the highly diverse and multicultural European professional and scholarly environment. More information about ACM-W Europe can be found at http://europe.acm.org/acm-w-europe.html

The ACM-W Europe BoF will take place at HiPEAC 2017 on Wednesday 25 January

1. Informatics Education in Europe: Institutions, Degrees, Students, Positions, Salaries – Key Data 2009-2014 http://www.informatics-europe.org/publications.html
2. https://www.wired.com/brandlab/2015/05/5-numbers-explain-stem-diversity-matters-us/
3. https://prezi.com/5264hlygqvtj/gender-and-diversity-in-stem-the-current-status/
4. https://hbr.org/2014/03/whats-holding-women-back-in-science-and-technology-industries
5. http://dl.acm.org/citation.cfm?id=2544049.2538690&coll=portal&dl=ACM

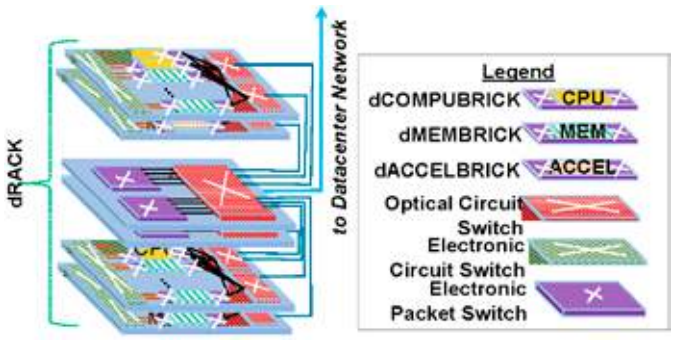
This issue's round-up of news and results from EU-funded projects sees an update on the VINEYARD project, new data centre architectures and I/O software stack of extreme-scale systems. We also hear about a novel toolchain to assess GPU performance and power in a vendor-agnostic way.

Innovation Europe

DISAGGREGATED RECURSIVE DATACENTRE-IN-A-BOX

dReDBox, a three-year project funded by the EC H2020 programme aspires to change the way in which data centres are built, proposing the deployment of pooled, disaggregated resources rather than monolithic and tightly integrated components. The goal is to improve utilization, scalability, reliability, and therefore power efficiency in cloud and edge data centres.

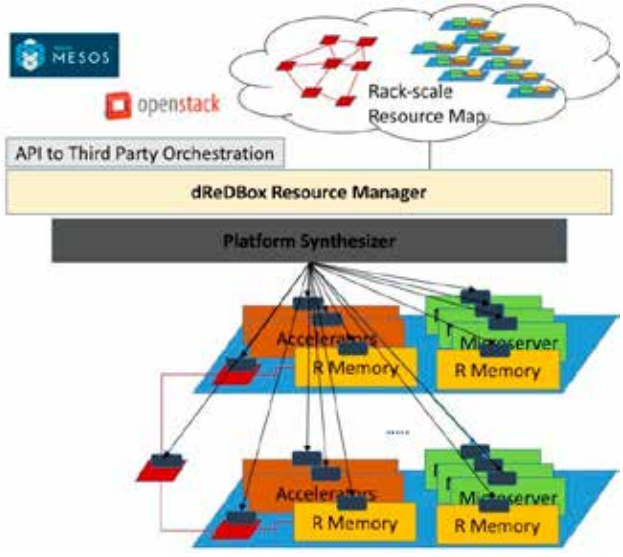
Contemporary cloud data centres use individual server units for processing, memory, acceleration, and storage. This arrangement incurs significant underutilization of resources due to the inherent inability of the system to closely match the amount of available resources within a server (or set of servers) to the variable user IT requirements. dReDBox aspires to remedy these limitations by moving to a pooled computing approach that enables the flexible pooling of disaggregated resources to match the runtime application requirements.



1: The dRedBox architecture

The project approach (Figure 1) is to disaggregate the traditional tray and overall rack architecture by introducing a set of bricks per tray, i.e. dedicated modules of specific functionality (compute, memory, acceleration, etc.). The main processing block is the Compute Brick, which can host multiple Virtual Machines (VMs). Memory Bricks collectively serve as a large pool of memory that can be (dynamically) redistributed to VMs. Accelerator Bricks host dedicated hardware modules to boost application performance.

Intra-tray communication between bricks is realized via electrical circuit crossbar switches, while an optical network will provide inter-tray connectivity. Therefore, inter-brick communication will be implemented through multiple communication networks of different speeds depending on the brick locations.



2: Orchestration tools high-level architecture

From the software management perspective (Figure 2), the Linux Kernel Virtual Machine (KVM) hypervisor will allow authorized tenants to reserve system resources to efficiently execute applications on hardware. dReDBox revisits the memory management

architecture at the VMM-level, implementing physical/logical attachments of remote memory to VMs.
Collectively, dRedBox will deliver a prototype of a fully-fledged, vertically-integrated data centre-in-a-box to demonstrate dReDBox in three use-cases: security, network analytics, and telecoms.
Contact: katrinisk@ie.ibm.com

NAME: Disaggregated Recursive Datacentre-in-a-Box (dReDBox)
START/END DATE: 01/01/2016 – 31/12/2018
KEYWORDS: Advanced computing, Embedded computing, Computing for servers, data centres
PARTNERS: IBM Research Ltd (Ireland), University of Thessaly (Greece), University of Bristol (UK), Barcelona Supercomputing Center (Spain), Sintecs BV (Netherlands), Foundation for Research and Technology - Hellas (Greece), Telefonica Investigation y Desarrollo SA (Spain), Kinesense Ltd (Ireland), Naudit High Performance, Computing and Networking SL (Spain), Virtual Open Systems SARL (France), Polatis Ltd (UK)
BUDGET: €6.45M
WEBSITE: www.dredbox.eu
The dRedBox project has received funding from the European Union's Horizon 2020 Programme under grant agreement no. 687632.

VERSATILE INTEGRATED ACCELERATOR-BASED HETEROGENEOUS DATA CENTRES

VINEYARD's goal is to develop the technology and the ecosystem that will enable the efficient and seamless integration of the hardware acceleration into data centre applications. The deployment of energy-efficient hardware accelerators will permit the significant improvement in the performance of cloud computing applications and the reduction of energy consumption in data centres.

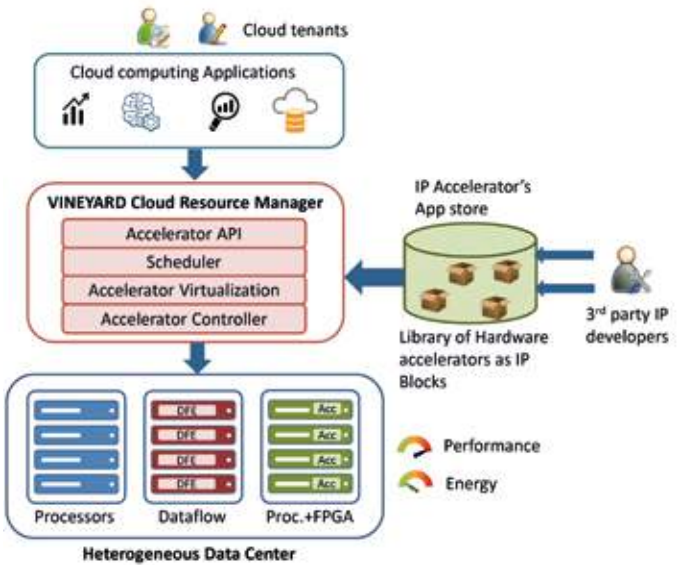
VINEYARD is developing an integrated framework for energy-efficient data centres based on programmable hardware accelerators. It is working towards a high-level programming framework that allows end-users to seamlessly utilize these accelerators in heterogeneous computing systems by using typical data centre cluster frameworks (i.e. Spark). The VINEYARD framework and the required system software hides the programming complexity of the heterogeneous computing system based on hardware accelerators. This programming framework also allows the hardware accelerators to be swapped in and out of the heterogeneous infrastructure so as to offer both efficient energy use and flexibility. To allow the efficient utilization of the accelerators by several applications, a novel VM appliance model for provisioning of data to shared accelerators has been developed. The enhanced VINEYARD middleware augments the functionality of the resource manager, by enabling more informed allocation of tasks to accelerators.

VINEYARD is also developing two types of novel energy-efficient servers integrating two kinds of hardware accelerator: programmable dataflow-based accelerators and FPGA-based accelerators. The servers coupled with dataflow-based accelerators are suitable for cloud computing applications that can be represented in dataflow graphs while the latter will be used for accelerating applications that need close communication between the processor and the hardware accelerators.

VINEYARD also fosters the establishment of an ecosystem that will empower open innovation based on hardware accelerators as data centre plugins, thereby facilitating innovative enterprises (large industries, SMEs, and creative startups) to develop novel solutions using VINEYARDS's leading edge developments. The ecosystem will bring together existing communities from all relevant stakeholders including providers of hardware intellectual-property (IP) technologies, data centre developers, data centre operators and more. This ecosystem will allow the promotion of open pluggable custom hardware accelerators (i.e. a hardware accelerator Application Store) that can be used in data centres in the same way that software libraries are currently being utilized.

VINEYARD plans to demonstrate the advantages of its approach in three real use-cases:
a) a bioinformatics application for high-performance brain simulations
b) two critical financial applications
c) a big-data analysis application

The VINEYARD framework will also be integrated with widely-used cluster computing frameworks, like Spark, that will allow the utilization of the accelerators for typical cloud computing applications like machine learning and graph computations.



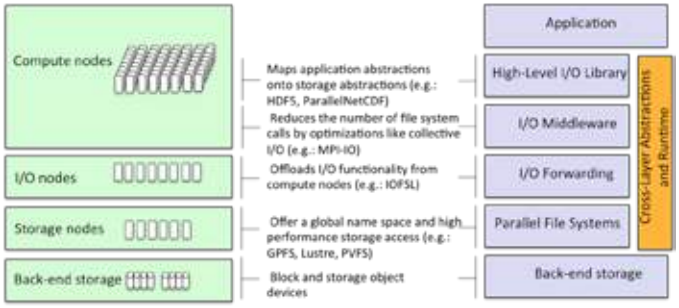
Contact: info@vineyard-h2020.eu

NAME: Versatile Integrated Accelerator-based Heterogeneous Data Centres (VINEYARD)
START/END DATE: 01/02/2016 – 31/01/2019
KEYWORDS: Accelerator-based, heterogenous, data centers
PARTNERS: Institute of Communication and Computer Systems (ICCS) (Greece), Maxeler Technologies (UK), Bull SAS (France), Queen's University of Belfast (UK), Foundation for Research and Technology-Hellas (FORTH) (Greece), Science and Technology Facilities Council (UK), Neurasmus BV (Netherlands), Neurocom Luxembourg (Luxembourg), Athens Exchange (ATHEX) (Greece), Leanxcale SL (Spain), Globaz SA (Portugal)
BUDGET: €6.28M
WEBSITE: www.vineyard-h2020.eu
The VINEYARD project has received funding from the European Union's Horizon 2020 Programme under grant agreement no. 687628.

TRANSATLANTIC COOPERATION COMES TO FRUITION

The recently concluded project CLARISSE (Cross-Layer Abstractions and Run-time for I/O Software Stack of Extreme-scale systems) focused its investigations on methods to increase the performance, scalability, programmability and robustness of data management of parallel scientific applications in order to actively support the current interest in designing large-scale parallel computing infrastructures that are two orders of magnitude faster than today's supercomputers (Exascale). University Carlos III of Madrid (UC3M) coordinated the project in cooperation with Argonne National Laboratory (USA), one of the key actors in research and development of system software for large-scale parallel supercomputers. The technologies developed in this project could be applied for significantly advancing the performance and scalability of parallel scientific and engineering applications such as climate modeling, new material design, astrophysics, genetics and bioengineering.

The key challenge is that the software for managing today's supercomputers has, historically, been designed by several actors in an uncoordinated manner; today, this approach is a major obstacle to increasing the scale of current systems. The CLARISSE project has been exploring a new solution to this problem by designing novel mechanisms and abstractions for coordinating data management at different system layers.



The CLARISSE vision appears in a work presented in May 2016, appearing in the Proceedings of IEEE/ACM CCGrid 2016 and awarded best paper at the conference (http://arcos.inf.uc3m.es/~florin/Publications/2016_ccgrid.pdf). In this paper, we proposed a novel approach for designing cross-layer optimizations for the software I/O stack of large-scale parallel supercomputers. After investigating various options for the CLARISSE software architecture, we opted for a design on three separate layers: a control backplane, a data backplane and a policy layer. The CLARISSE control backplane is a coordination framework that aims to support the global improvement of key aspects of data staging including load-balance, I/O scheduling and resilience. The data plane manages the distributed data flow and exposes control points to the control backplane, which coordinates the data staging. On top of the data and control planes, we have illustrated the capabilities of CLARISSE through two policy case studies: an elastic collective I/O and a parallel I/O scheduling policy. Furthermore, in work not included in the paper, we followed a co-design approach for showing that by trading off data locality and computational load balance in scientific workflow execution we can substantially reduce the I/O traffic and improve the performance and scalability over existing practice.

Contact: Florin Isaila fisaila@inf.uc3m.es

NAME: Cross-Layer Abstractions and Run-time for I/O Software Stack of Extreme-scale systems (CLARISSE)
START/END DATE: 30/09/2013 - 29/09/2016
KEYWORDS: HPC, storage, system software I/O stack, scientific workflows
PARTNERS: University Carlos III de Madrid (Spain) and Argonne National Laboratory (USA)
BUDGET: 352,823.70 Euro
WEBSITE: http://arcos.inf.uc3m.es/~florin/clarisse
SOFTWARE: https://bitbucket.org/fisaila/clarisse
The CLARISSE project received funding from the European Union's Seventh Framework Programme under grant agreement no. 328582.

VENDOR-AGNOSTIC TOOLS FOR ASSESSING GPU PERFORMANCE/POWER

The toolchain delivered by the LPGPU2 project will support all state-of-the-art APIs: OpenGL ES, Vulkan, OpenCL, SyCL and HSA and will enhance the European innovation capability in mobile GPUs by delivering a novel toolchain to assess GPU performance and power in a vendor-agnostic way.

The LPGPU2 EU-funded project consists of a consortium of graphics experts from TU Berlin, Samsung Research UK, Codeplay, Think Silicon and Spin Digital.

The toolchain will be tested using contemporary VR/AR/video and image processing applications in today's most popular graphics and compute APIs. The consortium's efforts are of particular interest to members of the Khronos group, because of its ambitious goal to enrich (via certain extensions) Khronos APIs with new methods to access power efficiency metrics and related performance data.

The project builds upon technology developed within the first LPGPU project and the CodeXL infrastructure (released recently as an open-source project by AMD). This infrastructure is enhanced in order to increase its visualization capabilities, its data acquisition functions, and its GPU power modelling and power analytics capacity.

'Our customers want high fidelity and high performance graphics from their mobile devices, but at the same time they want the maximum possible battery life to enjoy games and graphical apps on the move. We are constantly looking for ways that can help application developers to meet these requirements, and the LPGPU2 project is providing valuable research so that we can create standards and tools that will support this effort,' says Graham Mudd, Senior Technology Manager at Samsung Research UK.

'We have performance monitoring hardware and a corresponding API for tuning the live measurements. But to associate these measurements, in a smart way, with specific performance and power bottlenecks of real-time graphics applications is a challenging problem. The LPGPU2 project includes the right group of people to solve this problem,' adds Georgios Keramidas, CSO of Think Silicon.

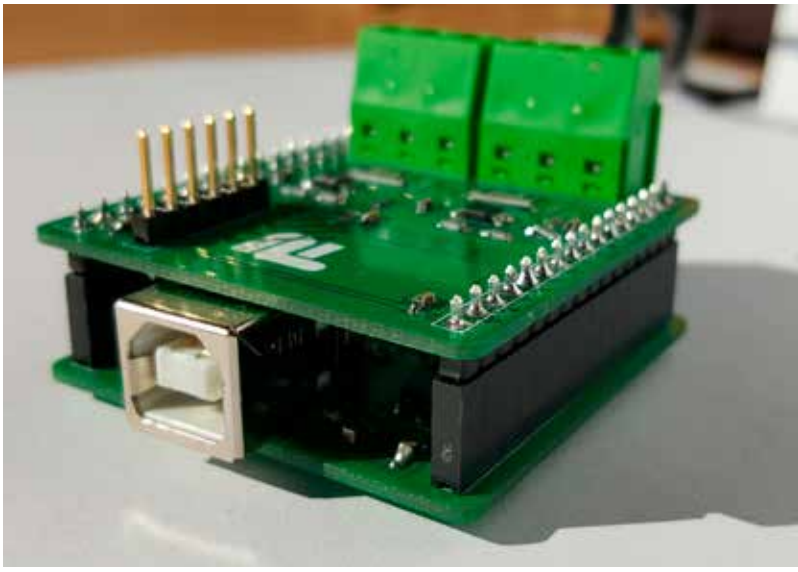
Andrew Richards, CEO of Codeplay concludes that: 'Codeplay is experienced in implementing demanding compute solutions into complex processor systems. These are exciting times for GPU compute. LPGPU2 is a perfect collaboration bringing knowledge and skills together for the benefit of high performance and power efficient solutions.'

The LPGPU2 consortium is working together closely to build live demonstrations of the toolchain using real life mobile applications. These demonstrators will actually provide proof-of-concept scenarios of the capabilities of the toolchain to help developers holistically optimize mobile GPU power consumption at all levels: hardware, algorithmic, and application software while running the most advanced graphics processing.

Contact: Graham Mudd graham.mudd@samsung.com

NAME:Low-Power Parallel Computing on GPUs 2 (LPGPU2)
START/END DATE: 01/01/2016 – 30/06/2018
KEYWORDS: GPU, low power, tools
PARTNERS: TU Berlin (Germany), Codeplay Software (UK), Think Silicon (Greece), Samsung Electronics UK (UK), Spin Digital Video Technologies (Germany)
BUDGET: €3.95M
WEBSITE:www.lpgpu.org

The LPGPU2 project has received funding from the European Union's Horizon 2020 Programme under grant agreement no. 688759.



$\frac{d\vec{v}}{dt}$ dividiti: unlocking the power of collaboration to enable efficient, reliable and cheap computing everywhere

In the latest in our series on small and medium enterprises (SMEs) that are going from strength to strength, Anton Lokhmotov and Grigori Fursin of dividiti tell us about how networking at a HiPEAC event led to them creating a company with a research focus and collaborative ethos.

dividiti is a computer systems startup co-founded in early 2015 by Anton Lokhmotov and Grigori Fursin. Our mission is to accelerate (dv/dt is... acceleration) the advent of efficient, reliable and cheap computing via effective collaboration and knowledge sharing.

We first began discussing collaborative optimization in 2006 (at the HiPEAC ACACES Summer School!), but it was not until 2012 that we started drawing up serious plans for collaboration (at the HiPEAC Conference!) At that time, Anton was leading a ten-person team for the ARM Mali GPU Compute compilers (OpenCL, RenderScript, CARP), constantly on the lookout for disruptive approaches to compilation and programming. Grigori had just finished a stint as the head of the program optimization group at the Intel Exascale Lab in France and returned to his senior tenured position at INRIA. Through the lenses of our R&D experience, we realized that the many outstanding problems in computer systems – such as the ever-growing space of design and optimization choices, lack of representative workloads, lack of common methodology and tools, and so on – can only be practically solved by using a community-driven, collaborative approach.

In 2014, we received a grant from the FP7 609491 TETRACOM Coordination Action to transfer to ARM Grigori's technology for collaborative and reproducible R&D, which became known as Collective Knowledge (CK) and was later released under a permissive open-source licence (https://github.com/ctuning/ck). Using CK, ARM were able to obtain valuable insights into performance of their products in a fraction of the time required by conventional analysis. This engagement confirmed that we were on the right track, and encouraged us to start up dividiti.

Together with a growing, passionate community across industry and academia, we are continuing to develop CK. We envision that CK will evolve into a multi-sided online platform that will bring together all key computer systems players - hardware designers, software developers, tool developers and end-users - and, by doing so, will uncover the tremendous value of collaboration between them.

We believe that very soon there won't be any other way of designing and optimizing computer systems other than by leveraging community effort. To ensure that we can pursue our long-term vision and not put it at risk for short-term goals, we haven't sought any external funding. In fact, most of our revenue comes from visionaries at leading companies who get the business value of collaboration and reproducibility. Our customers include ARM, General Motors (see our announcement on page 7) and even Raspberry Pi. They are all highly competent in what they do, yet they see how using CK adds real value to their business.

We love balancing between industry and academia. On the one hand, we are creating robust solutions that address the real needs of our customers using our unique expertise and technology. On the other hand, we are continuing our research on collaborative performance optimization (from IoT to supercomputing), run-time adaptation, compilation and machine learning which we started with our PhDs. We also feel we do social good by crusading for reproducible research and contributing to artefact evaluation activities for leading conferences such as PPoPP, CGO and PACT (http://ctuning.org/ae). We are always happy to hear from like-minded individuals and to collaborate to maximize the impact of computer systems research on society at large!

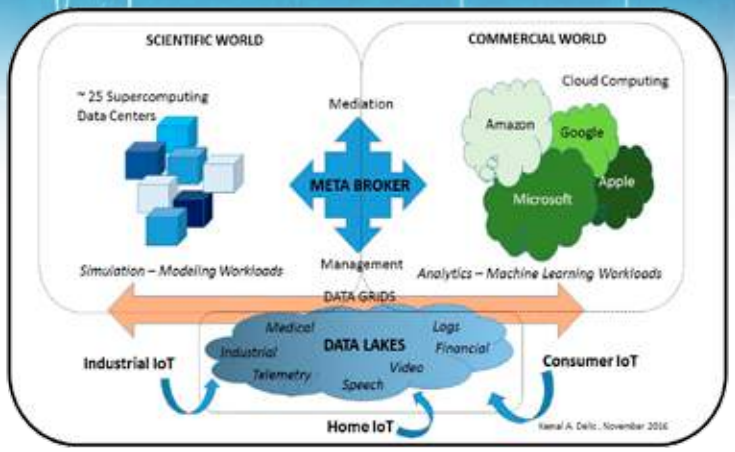
DO YOU WORK FOR AN INNOVATIVE TECHNOLOGY SME?
Contact communication@hipeac.net with your story.

Architecting hyperscale scientific systems

Kemal A. Delic of Hewlett Packard Enterprise looks to the future and tells us about how computing will take scientific systems and research to the next level.

We are witnessing the rise of hyperscale companies dominating business, commercial, government and media ecosystems. It is very likely that the domain of scientific research is the next area to be transformed by the rising tide of so-called 'hyperscalers': Apple, Microsoft, Amazon and Google. In this article, I outline the conceptual architecture of Hyperscale Scientific System (HSS), and describe briefly the most relevant parts of such an architecture (HPC, Cloud, Data Lakes, IoT). Convergence of those components will likely change the nature and substance of our scientific future in the next 10-15 years. It will also resolve long-standing challenges such as understanding climate change and finding the cure for various cancers and other serious diseases, and will eventually explain how our brains function.

Architecting means taking a very wide, long-ranging and overall system view of various subsystems. Here, HPC, Cloud, IoT and Data Lakes are the subsystems whose integration will represent HSS. In an analogy to commercial hyperscale systems, we postulate that HSS will accelerate, augment and automate all activities in many scientific communities. In practical terms, it will mean shorter project duration due to technological acceleration; reach and impact of research will be greatly enlarged and the cost of conducting research will be much lower.



Conceptual architecture of Hyperscale Scientific System

Meta-broker is the key component of HSS facilitating monitoring, management and mediation among competing parties. Here, we will have effective cooperation/competition between the commercial world represented as cloud computing and the scientific world represented as scientific computing (see diagram). It will also enforce standardization and regulation, therefore enabling smooth functioning. It will manage delicate balance among economic factors, performance requirements, legal obligations and criticality dimension for these very different worlds. The HPC world can be represented by some ~25 supercomputing data centres running mainly heavy simulation and modelling scientific workloads. More than half of the big commercial workloads are executed in only three hyperscalers' cloud facilities (Amazon, Microsoft, Google). Many Exabytes of data are stored in Data Lakes – containing a wide variety of different data collections, in different data formats, and being of a very different nature. IoT is the emerging paradigm of internet becoming web and now appearing as the web of things.

Envisioning future HSS might be the technological basis to realize the long-standing dream of scientists cooperating on a global basis and with limitless resources and facilities. Today, scientific research is not really globally unified, and is not really open or highly collaborative – despite all efforts to make this a reality. The ultimate expectation would be that big projects currently taking decades will be shortened to years, costs reaching billions today will be reduced to hundreds or tens of millions, and the number of participants and those who benefit will grow significantly. These effects are already seen in the commercial world of hyperscale companies.

HiPEAC is keen to help European companies large and small to prosper and to facilitate technology transfer from universities to industry. Bartosz Ziółko of Techmo, a successful university spin-off company, gives us some valuable insights and tips for researchers thinking of launching a similar venture.

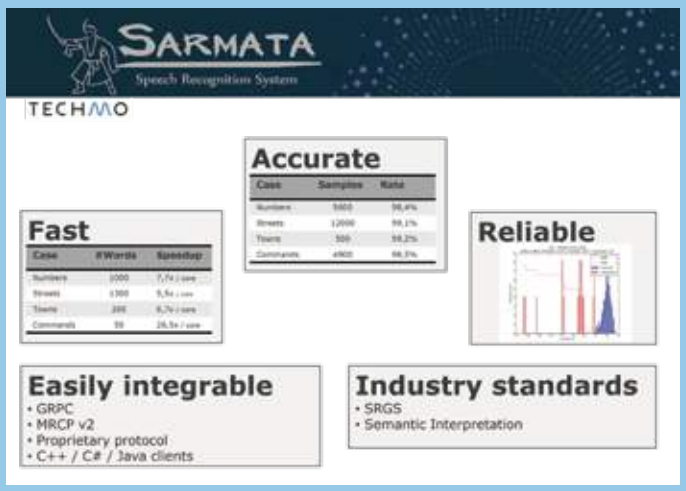
Techmo

Researchers from the Digital Signal Processing Group at AGH University of Science and Technology started their work on automatic speech recognition (ASR) of the Polish language around ten years ago. Our group focused largely on applied research and started to sell its first licences in 2009. Ever since then, setting up a spin-off company was on the cards. The main driver was the need to speed up the licensing process, which had always been too slow. With this business need in mind and with a specialization in speech and language technologies which are not easily transferred between countries due to the differences between languages, Techmo was born in Autumn 2012.

The process of setting up the company was a multi-stage one. The founding group was formed to work on a business model and to define the company's activities and scope. The company was launched in February 2013 and we made the decision to initially start off the company as part of parent company INNOAGH, which is owned by the university. My co-founders were four members of the AGH faculty and a PhD student.
By late March 2013, the company was legally registered as Techmo sp. z o.o. and could begin to work on developing its products. The negotiations with the university were thankfully rather short because there was no hard transfer of any intellectual property. Instead, in the meantime, the university had signed a non-exclusive licence for the know-how which would be needed within Techmo. The licence was based on royalties and involved no advance payment.

Fast-forward to September 2014, and four new team members joined Techmo to support the original founders. Two of them were technical experts, one a lawyer and the last one a businessman experienced in the speech technology market and who runs another company with which Techmo cooperates.

Techmo took an order from Pirios S.A. in September 2015 for help in developing its IVR (Interactive Voice Response) system. Techmo outsourced people to work in Pirios alongside its own engineers. Part of the work was the integration of Pirios IVR with Techmo products. One short month later, in October 2015, Pirios delivered IVR with Techmo ASR software to the Board of Municipal Infrastructure and Transport in Krakow (ZIKiT).



speech recognition system

Techmo offers a variety of speech technologies, however its core product remains ASR. Techmo has also made a pilot version of an automatic system for reporting of failures by customers in Gdansk for the company Energa, and is now in the process of making a working system for customers of Fortum, both with Pirios. We are also working on an Oculus-based simulator for firefighters as part of a consortium with AGH and Anshar Studios.

Our experiences of the last few years have given us a few valuable lessons. Strategies for startup companies are very often based on American experiences and the US market. Techmo operates mainly in Poland and we would like to stress some of the crucial differences between building a startup company in the US and in Poland (or possibly in many other EU countries) which we have observed through our Techmo journey.

The first key difference is that the customer market is much stronger in the US. This is why American investors and, as a result, startups are more often focused on business-to-customer products. In Poland, the customer market is quite weak and building a company based on B2C is very risky. We would not advise that. The business-to-business market is much stronger in Poland, and this is the market that Techmo principally targets. Another important customer in Poland, one which covers a large part of the market, is the public sector. It is difficult to acquire public institutions as customers as there are often lengthy tender processes involved, but, when it happens, it is easier to keep them. It should be also stressed that the majority of large IT companies in Poland were able to grow well as a result of having the public sector as a customer. In the US, while public sector is still important, very few people would advise a startup to focus on it.

The second important difference is that in the EU there are many more public grants. This is an interesting point that American startup founders often make to their European counterparts: that it is easier to build up a company in Europe because public money can be used. Of course, it's not always that easy to access such help and it's worth bearing in mind that, on the other hand, in the US there are far more venture capital operations. As a result, European startups may want to focus on gaining public grants, rather than on searching for an investor.

Looking to the future, Techmo wants to develop two new products. The first is IVR for small enterprises and operations, especially the medical profession. The second is SoundToolkit, a software library for video games allowing soundtracing, spoken dialogues and AI based on sound cues.

www.techmo.pl/index.php/en

Europe is a great place to access support and public funding for small businesses. The European Commission's Startup Europe campaign aims to strengthen the business environment for web and ICT entrepreneurs so that their ideas and businesses can start and grow in the EU.
For inspiration and ideas, see
www.ec.europa.eu/digital-single-market/en/startup-europe
www.startupeuropeclub.eu/eu-funds-and-support
www.startupeuropeweek.eu
www.europa.eu/youreurope/business/start-grow/start-ups/index_en.htm
www.ec.europa.eu/small-business/finance/index_en.htm
www.eu-startups.com
www.twitter.com/euroinvestnews

STARTUP EUROPE

CRAFTERS experience: a Smart Real-Tim

In this issue's feature on transfer of technology from projects and universities to industry, we hear both sides of the story: the CRAFTERS and EUROSERVER projects tell us about how their results and innovations have been applied and led to the creation of new companies, while Eta Scale AB, set up by university researchers, reveals its foundations in three European projects.

Center of Excellence DEWS, University of L'Aquila, Italy

Electronic components and systems are a pervasive key enabling technology which have an impact on all industrial sectors and almost all aspects of life. A smartphone, a smart card, a smart energy grid, a smart city, even smart governance; every “smart” thing is based on integrated chips running embedded software. The main political decision-makers in the areas of micro- and nano-electronics, smart integrated systems and embedded/ cyber-physical systems are the European Union (through the Commission), and the member states and associated countries of the research funding framework programme Horizon 2020 and of the associations EPoSS, AENEAS and ARTEMIS Industry Association. The Joint Undertaking (JU) ARTEMIS was formed by a decision of the European Council on 20 December 2007 and merged into the Electronic Components and Systems for European Leadership (ECSEL-JU) in 2014. All ongoing ARTEMIS projects have been continued under the ECSEL-JU. CRAFTERS was an ARTEMIS project proposed under the ARTEMIS-2011-1 call. The name "CRAFTERS" stands for "ConstRaint and Application driven Framework for Tailoring Embedded Real-time Systems". The project started on 01 June 2012, with a duration of 36 months, and the final review meeting was held on 18 September 2015 in Brussels. ARTEMIS subprogrammes relevant to CRAFTERS were the realization of computing platforms and the security and critical infrastructures protection for embedded systems.

Marco Santic, Luigi Pomante, Giacomo Valente, Vittoriano Muttillo of DEWS Center of Excellence, University of L'Aquila, Italy

In this scenario, our group, which is part of the Center of Excellence DEWS (Design Methodologies of Embedded Controllers Wireless Interconnect and Systems-on-chip) based at the University of L’Aquila (Italy), participated to CRAFTERS in close collaboration with Thales Italy S.p.A. DEWS started its operations in 2001 and its research activities involved Networked Embedded (Control) Systems with a focus, in particular, on Wireless Sensor Networks and Electronic Design Automation (EDA) and working on Model-Driven HW/SW Co-Design methodologies at System-level.

In CRAFTERS, DEWS acted as a provider of “methodologies and HW/SW platforms for FPGA-based prototyping”, by contributing to the realization of different multi-core platforms for FPGA, the definition of model driven approaches and profiling methodologies, the development of middleware (MW) and the definition of metrics to evaluate use cases. DEWS also collaborated with other partners (Rapita Systems LTD in the first instance) on the integration of tools into current industrial development flow and their application in a Thales Italy case study to profile the behaviour of the system during application execution.

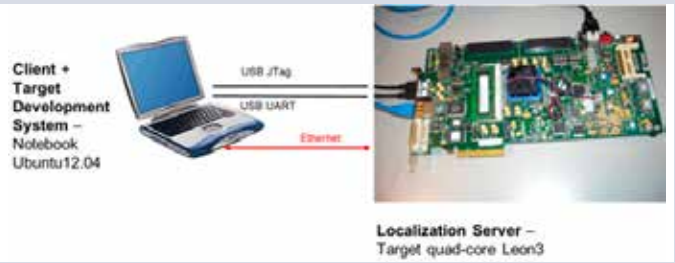
With respect to the HW/SW platform, DEWS has provided multi-core systems based on MicroBlaze and Leon3 soft-processors,

ne Location System for Indoor Scenarios

with operating systems running on top of them. Then, DEWS has provided a specific MW framework and related design guidelines starting from the analysis of general project requirements, use cases, and existing reference MW classes. Moreover, a MW has been designed for supporting run-time monitoring of resource usage in cooperation involving Rapita Systems and Thales Italy. Specifically, the run-time monitoring was composed of a hardware profiling system, tailored to measure response time and bandwidth, and a middleware layer.

The cooperation with Rapita Systems on tool integration targeted the integration of the hardware profiling mechanisms into the advanced profiling suite RVS (Rapita Verification Suite): this collaboration defined interfaces for RVS profiling data collection, according to the operating mode, and defined middleware component able to propagate data to RVS independently of OS and hardware.

In the context of the CRAFTERS project, three main industrial use-case contexts were targeted: Image and Video processing, Industrial and Intelligent Transportation Systems and Ultra-Wide Band Positioning and Communications. The Thales Italy use-case related to the third one: they worked on the problem of indoor localization of mobile network nodes using UWB-based ranging to measure the inter-nodal distances. The main application was a software that localized the nodes in the scenario, starting from a first estimation of inter-nodal distances. It ran on a quad-core processor based on Leon3, implemented on FPGA.



The prototype

At the beginning of 2016, after the decision of Thales Italy S.p.A. to close down its business operations and the industrial area of Chieti Scalo, TEKNE srl, a company that works in the automotive area from both electrical and mechanical points of view, started a project that involved the acquisition of the remaining human resources of the Thales site. This was due to its interest in inheriting Thales' ICT resources, in order to join the Internet of Things trend in the automotive context and expand internal know-how by strengthening their skills and moving them from automotive to ICT. Specifically, the results of the CRAFTERS project are in line with the overall strategies of the TEKNE business plan, into which relevant activities of the former Thales site will gradually be absorbed, to ensure that such results will be used as quickly as possible to have a positive impact on the products and solutions already developed. Furthermore, the realization of real devices based on the technologies arising from the CRAFTERS experience, and their utilization in a real environment, will be exploited by TEKNE industry in different domains, from the armed forces to the fire service and the police, in order to support the improvement of their actions in emergency and dangerous situations (from war, to fires and earthquakes). TEKNE is developing solutions for environmental protection (sensor networks, dealing also with chemical, biological, radiological threats), special vehicles to be used during emergencies (cooperation with unmanned systems and dismounted crews), indoor or GPS denied localization and tracking of vehicles, robots, personnel, including rescue.

Also Rapita is examining CRAFTERS results to improve their software products: following initial collaboration in the project itself, there is now work being carried out on the concept of timing analysis. The DEWS research centre is involved with this project via a Master thesis.

Finally, from the DEWS research centre side, we are building upon CRAFTERS results to initiate new research activities on HW/SW profiling mechanisms for multi-core and network-on-chip platforms on FPGA. New teaching modules have been introduced into University of L'Aquila courses, and new research projects (e.g. Artemis-JU EMC^2) have been set up as a follow up of CRAFTERS.

Contact: Luigi Pomante luigi.pomante@univaq.it

Eta Scale AB: a company grown from European projects

Eta Scale

Eta Scale AB is a Swedish company which has been created by Uppsala University researchers to develop software technology for High Performance Computing and Big Data applications.

Eta Scale's technologies have their roots in three FP7 projects: ERA, HEAP, and LPGPU.

The company offers advanced solutions for distributed computing (ArgoDSM) and for Decoupled Access-Execute (DAE) compilation:

- ArgoDSM offers scalable shared memory at a cluster scale and is especially well-suited for fast development of shared-memory programs. Data movement is handled automatically and optimizations such as caching and prefetching are managed entirely by the system. ArgoDSM's proprietary coherence protocol and advanced hierarchical scheme for distributed locking offer high-performance, scalability, and ease of use.

- Our DAE compiler tools optimize memory-bound applications for energy (Daedal) and performance (Clairvoyance).

Both technologies are offered as open source for evaluation, academic and research purposes, while for commercial use we offer custom licensing, support, and joint evaluation and development. In addition, Eta Scale develops and licenses IP related to multicore cache coherence.

ARGODSM

ArgoDSM is a modern approach to software distributed shared memory (Software DSM) that allows the well-known paradigm of shared memory programming to be applied on the scale of distributed systems (clusters). It transcends the difficulties imposed on the programmers by other paradigms such as message passing or partitioned global address space (PGAS) approaches, which require significant effort from the programmer to optimize.

Unlike standard PGAS approaches we offer high performance shared memory programming without the need for explicit awareness of data placement: Coherent caching takes efficient care of data communication automatically. In addition, we extend all the standard shared memory synchronization that is available today in a single server and standard operating systems to work efficiently at cluster scale.

Our approach makes use of today's technological landscape, where network bandwidth in compute clusters has grown much faster than processing speeds or access latencies, both for memory and for network accesses. Today, instead of bandwidth limitations, the miss latency is the limiting factor for large data caches. ArgoDSM includes solutions that trade off latency for bandwidth. It relies on fast remote direct memory access (RDMA, readily available in modern network interconnects), combined with a scalable, low-latency, distributed coherence protocol, and cutting-edge approach to distributed locking.

ArgoDSM is intended both for HPC and Big Data applications by providing different flavours optimized to the respective needs of these workloads. For example, Argo-BigMem, intended for Big Data applications, decouples the allocation of computing resources from the allocation of memory resources, and allows a program in one node to use the memory of many other nodes without taxing their compute resources.

DECOUPLED ACCESS-EXECUTE COMPILER INFRASTRUCTURE

Our Decoupled Access-Execute (DAE) LLVM compilation infrastructure optimizes applications for energy efficiency or for memory performance. In our publications we have shown that our tools can improve the Energy-Delay-Product (EDP) on Intel Platforms by an impressive 30%, and can speed up memory-bound programs on ARMv8-64 platforms by 30%. User programs are transformed for software decoupled access-execute (DAE), where memory accesses are decoupled from computation. This separation allows optimizations, such as selective Voltage-Frequency scaling (DVFS), high-bandwidth prefetching, and customized acceleration of the computation. Our compilation infrastructure for DVFS on Intel platforms is called Daedal and is offered under a dual license scheme: open source license for evaluation, academic, and research purposes and under a custom licence for commercial purposes. The compilation infrastructure for memory performance on ARMv8-64 platforms is called Clairvoyance and is offered under custom licence.

For more information refer to www.etascale.com
or contact@etascale.com

EUROSERVER: transferring and using power efficiency know-how

MORE COMPUTATIONS FOR LESS ENERGY

The EU-funded project EUROSERVER has developed a new approach for ARM-based technology to halve the cost of powering data centres. Based on the concept of chiplets, where multiple silicon subsystems are mounted in an integrated device, along with an associated new groundbreaking system architecture, the project has enabled more energy-efficient servers and has even inspired startups motivated by the new technology.

Delivering the potential to reduce energy consumption in data centres by at least half, the breakthrough leads to substantial reduction in the total cost of acquisition and operation (total cost of ownership (TCO)).

EUROSERVER created system architecture and runtime software innovations: the sharing of peripheral devices, access to system wide memory, compression of data to better utilize memory and lightweight hypervisor capabilities. These are being demonstrated with applications across the data center and telecommunications domain. All reduce energy consumption and have been developed for systems built around energy-efficient 64-bit ARM®v8-based architecture.

Two startups have already been launched, inspired by the technology created during the project:
KALEAO Ltd., with headquarters in the UK and labs in Crete and Italy, has introduced a unique new generation of web-scale, true-converged server appliance that features physicalized resource sharing, OpenStack virtualization services and extreme core density, leading to low energy consumption and significant computing capabilities.

ZeroPoint, in Sweden, commercializes the memory compression innovations.

Data centres account for a huge consumption of power. If all the data centres in the USA alone were a country, their energy consumption would rank 12th in the world. As the capacity and number of data centres increase, so do the financial and environmental impacts of their vast energy use. EUROSERVER's innovation will take efficient and scalable ARM processors and use the flexibility of a System-on-Chip (SoC) design at the system level to create a new type of server.

This evolution is comparable to the transition from mainframe computers to mass-produced personal computers (PCs) in the 1980s, which in turn evolved into modern servers. With smartphones the contemporary equivalent of 1980s PCs, now is the time to take advantage of their architectural flexibility and evolve them to create the energy-efficient servers of the future.

GREEN COMPUTING NODE FOR EUROPEAN MICRO-SERVERS

At the time of writing, the consortium is in the final stages of results evaluation that will be released in the project workshop at the HiPEAC 2017 Conference in Stockholm (25 January 2017). These results compare Intel XeonD, ARM Juno, Gigabyte MP30AR0 ARM X-Gene1 and the EUROSERVER prototype platforms.

The the EUROSERVER platform is based on a Xilinx UltraScale+ FPGA that demonstrates novel technologies built in the project. A number of relevant workloads are being evaluated including web-server hosting (LAMP/WAMP), distributed databases (HADOOP) and network communications (C-RAN).
ARM based processors currently dominate the mobile market and EUROSERVER is attempting to integrate the full stack as a power-efficient alternative to the Intel-based designs that currently dominate the data centre.

Please find the most up-to-date set of results on the EUROSERVER webpage.

EUROSERVER has received funding from the European Union's FP7 programme under grant agreement no.610456. For more information, visit www.euroserver-project.eu or contact Ms. Isabelle Dor - isabelle.dor@cea.fr
About KALEAO: visit www.kaleao.com
About ZeroPoint: visit www.zptcorp.com

Career talk: Alessandra Bagnato, Research Scientist and Project Manager at SOFTEAM R&D Department

Tell us a little about your professional background. How did you come to be working at SOFTEAM and where have you worked in the past?

I studied at University of Genoa, Italy, obtaining a MSc in Computer Science with a focus on distributed computing before getting my PhD in Computer Science from TELECOM SudParis and Université Evry Val d'Essonne in France. In Italy, I have also been Adjunct Professor at the University of Milano Bicocca for a few years.

My studies and work in universities set me on the path of research activities in close collaboration with companies and inspired me to pursue my career as a research scientist in industry.

Just after obtaining my MSc in April 1999, I started working as research scientist for TXT e-solutions Corporate Research Division headquartered in Milan in several areas of technology research related to embedded systems design, software/service development and security. There, I participated in a number of publicly-funded research projects involving academic and industry partners. Among them, in 2007, I was the project coordinator of MOMOCS (Model driven Modernisation of Complex Systems) and in 2009 I was the coordinator of MADES (Model-based methods and tools for Avionics and surveillance embedded SystEmS) (www.mades-project.org). I met my Softeam colleagues whilst coordinator of the MOMOCS and MADES projects and enjoyed so very much working with the team there and using the Modelio modeling tool. I happened to move to Paris and, just after the MADES project ended in October 2012, I was very happy to start working at Softeam following up on the research on model-based methods and tools for cyber-physical systems carried out with them in the preceding years. In particular, I really liked to pursue my research on CPS modelling as a very active and interesting domain in which modelling is seen as one of the foundational approaches for dealing with natural heterogeneity in CPS, effectively capturing domain-specific concepts and requirements.

Around the time that the MADES project started, I was also very happy to become a member of the European Network on High Performance and Embedded Architecture and Compilation (HiPEAC).

What is a typical day at work? What are your department's main current priorities?

My day-to-day work mostly involves developing new ideas and discussing with colleagues and partners, prioritizing action items, taking decisions (both administrative and technological) on the research project we are coordinating, following up on activities, and supporting project partners and group members in the standardization activities of the Object Management Group.

The aspects which I find most enjoyable are the enthusiasm within our team in fostering R&D in Europe, including developing new ideas and discussing with research colleagues and partners in different countries and research areas on different issues and, above all, learning new technologies, contexts and industry needs to which to apply them. I will soon start working as Softeam representative in the H2020 CPSwarm project to extend the SysML and MARTE representations and methodologies to better address design of groups of CPSs showing behaviours emerging from the aggregate of simpler local behaviours. Such a scenario requires specific extension and design techniques to address the peculiarities of these swarms of interacting CPS. This will be extremely enjoyable work since, until now, Modelio mainly targets single CPS systems and currently no support is provided for modelling and deploying swarms of CPS.

What are your department's main current priorities?

Softeam's main R&D priorities are to develop technology surveys, innovative usage scenarios and case studies within the current research projects and to deliver them to the whole company.

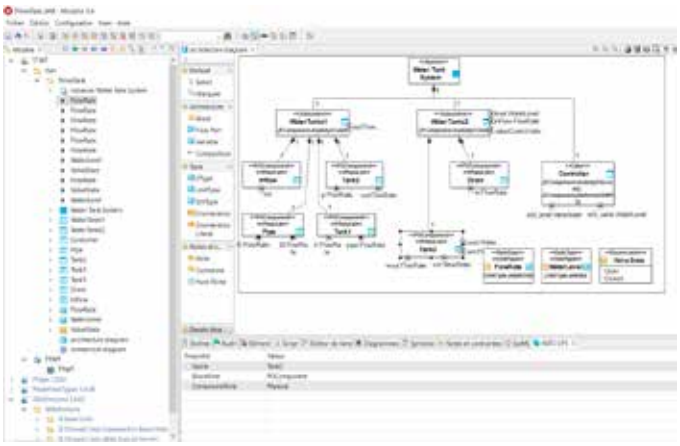
We are currently involved in the H2020 INTO-CPS project. The aim of INTO-CPS is to create an integrated 'tool chain' for comprehensive model-based design (MBD) of cyber-physical systems (CPSs). The tool chain will support the multidisciplinary, collaborative modelling of CPSs from requirements, through design, down to realization in hardware and software. Its results will be extremely interesting for HiPEAC community members who design embedded systems / cyber-physical systems and simulate, test or integrate them into larger systems.

We have also started working on the EUROSTARS MODELS project to improve system modelling and design exploration of applications for heterogeneous and parallel platforms (models.epfl.ch).

Moreover, we are coordinating the ITEA 3 MEASURE project on measuring software engineering issues. Although it is not currently focusing on cyber-physical systems, its results could be potentially of great interest in the domain as well for measuring

"Modelio will be defined as the "host" environment in which the Softeam's CPS projects outcomes will live and evolve, bringing this asset to our customers, to research projects and to the research community"

purposes in a standardized way (http://measure.softeam-rd.eu/). It will also release the Modelio modelling tool enabled with the SMM module to allow the specification of metrics through the implementation of the OMG's Structured Metrics Meta-Model (SMM) profile.



Screenshot from Softeam's Modelio INTO-CPS Project Open Source CPS Modelling Extension

What do you see as being SOFTEAM's main priorities or objectives in topics relevant to HiPEAC in the coming years?

Softeam will continue its current research in cyber-physical systems modelling to push its work within the Modelio modelling tool forward, exploiting results to enlarge the functionalities of the tool and, more in general, its services and consultancy offerings. Many modelling techniques have been investigated in the CPS design community and Softeam with its Modelio (Modelio.org) modelling environment will deliver a broad-focused range of standards-based functionalities for software developers, analysts, designers, business architects and system architects.

Softeam has been a contributing member and voter of the Object Management Group since 1994, and is deeply involved in its work and, in particular, the standardization of UML, the UML Testing Profile, the OMG Systems Modeling Language (OMG SysML) and MARTE (Modeling and Analysis of Real-time and Embedded systems). We will continue to pursue these activities in the coming years. Modelio will be defined as the 'host' environment in which the Softeam's CPS projects outcomes will live and evolve, bringing this asset to our customers, to research projects and to the research community.

What's the best part of your job? What keeps you motivated?

The constant challenge of learning new things and to apply research to real industrial contexts in the most helpful way possible. I like very much the potential for technology transfer and dissemination of the work that's been done to create awareness of the advances made by research.

The gender balance in computing in Europe remains poor. What were your experiences of being one of few women during your studies and early career? What advice would you give to girls and young women who are interested in studying and working in this field?

I was very passionate about studying and learning and was very happy to prove that women as well as men can succeed in computer science studies and in the profession.

I would say to keep themselves motivated and always follow their passions and what they really like. That goes for men too of course and for any sector or decision we have to make in life.

I suppose that the tech community can promote the access of women to computing and computer science in general by showing 'working examples' in the field more and more. Exactly as the broad distribution of the HiPEAC newsletter is doing.

WHERE WILL YOUR CAREER TAKE YOU NEXT?

If you're passionate about your career and would like to share it with the HiPEAC community, we'd love to hear from you. Email communication@hipeac.net with your story.

HiPEAC internships are a great way to get a funded work experience placement at a leading technology company. They provide an excellent opportunity to build long-lasting networks with experts in your field working outside academia. Interested in applying? Further information is available on the HiPEAC website: www.hipeac.net/mobility/internships

HiPEAC internships: making the most of fresh new talent

NAME: Mariano Benito
RESEARCH CENTRE: University of Cantabria
HOST COMPANY: Recore Systems, Netherlands
DATES OF INTERNSHIP:
13/06/2016 – 15/09/2016

MULTI-PORT 10G ETHERNET TRAFFIC GENERATION AND PERFORMANCE BENCHMARKING

I am a third year PhD student at the University of Cantabria. My research interests focus on HPC system networks (topologies, routing, Ethernet technology, and so on). Thanks to a HiPEAC industrial PhD internship grant, I spent three months at Recore Systems in the Netherlands.

Recore is currently focused on a new network device belonging to its FlexaWare® family. This device is a real-time streaming analytics embedded platform which receives large amounts of streaming data in parallel, processes it, and turns the data into actions. Recore will verify its performance and functionality and therefore created this internship project to research and develop a test environment to verify and benchmark Ethernet-enabled devices with high bandwidth data streams. During my internship, I worked largely on the following activities.

Firstly, in order to select an appropriate tool, we made a packet generators' background study in which we summarized the principal options available in the market. The best alternative was MoonGen (presented on IMC'15), a packet generator based on the DPDK framework from Intel.

Secondly, in order to design the traffic patterns to inject into the device, several works were studied. These papers are not very recent; therefore, we made a traffic distribution of 16 different kinds by combining information from the sources studied. The representativeness of each type of traffic on the internet carrier characterization developed is determined by its associated injection rate.

Thirdly, we designed the current testing architecture thinking of the future test requirements for this product. The proposed architecture can be largely divided, from a logical point of view, into an injection thread for generating and injecting traffic and a reception thread for capturing and analyzing packets generated by the system under test (SUT). From a physical point of view, the elements described above can be in the same or in different machines.

Finally, we generated tests for testing functionality and performance. With these tests, for instance, we can inject the traffic distribution referred to above with a desired injection rate or print and analyze the correctness of information generated by the SUT. These tests are a useful tool for the Recore team for its daily work and a good starting point for continuing with the improvement of its integration testing system.

Looking back, this internship was a great experience. I would like to express my sincere gratitude to HiPEAC organization for giving me this opportunity and to Recore Systems for embracing me in its work team and for its hospitality.

Gerard Rauwerda, CTO of Recore Systems concludes that: 'during his internship at Recore Systems, Mariano contributed with his research to an advantageous testbench environment for real-time data analytics devices. Likewise, we showed Mariano how his research is already applicable in real practice.'

The HiPEAC network has over 800 PhD students who defend, on average, at least two theses per week. Being an affiliated PhD student of a HiPEAC member gives access to a vibrant and dynamic research community spanning academia, large industry and SMEs. It also provides the opportunity to apply for internships and collaboration grants and to attend networking events and the annual ACACES Summer School.

Three-minute thesis

NAME: Ali Azarian
RESEARCH CENTRE: Faculty of Engineering, University of Porto
ADVISOR: Prof. João M. P. Cardoso
THESIS: Task-level Pipelining in Configurable Multicore Architectures

Multicore processors and parallel platforms are providing the opportunity to increase the performance of applications. Reconfigurable computing devices such as Field Programmable Gate Arrays (FPGAs) are also providing the opportunity for computing and storage resources to meet the specific needs of an application. Nowadays, other accelerators such as GPUs also have potential for high performance for many applications that use a large number of cores which run in parallel to accelerate the execution of applications. However, GPUs require the use of specific programming tools and techniques to achieve high performance. Also, the advantages of using GPUs depend on the parallelism potential of the applications. Thus, GPUs are not suitable as accelerators for some applications.

To efficiently exploit the advantages of multicore architectures, parallel programming and parallelization techniques to speed up the processing of an application are becoming more and more important. A variety of applications in the domain of image, video and signal processing (e.g. Wavelet Transform, Fast DCT, FIR-Edge, etc.) consists of linear sequential stages which can be dependent on or independent of each other. In most of these sequential programs, the output of a stage is the input of the next stage. One possibility to improve performance is to provide pipelining schemes to allow the processing of the next input before the subsequent stages have completed their process. In pipeline parallelism, the stages can operate simultaneously and process different data. A parallel execution in pipeline parallelism is obtained by partitioning the data into a stream of data elements that flow through the pipeline stages one after another.

In the domain of pipeline parallelism, task-level pipelining is also an important technique to speed up processing of an application, especially when dealing with applications consisting of producer/consumer (P/C) tasks in multicore-based systems. In these applications, producer tasks output data to be processed by the consumer tasks. Using task-level pipelining, a consumer computing stage (e.g. consisting of a loop or a set of nested loops and also identified herein as task) may start execution, before the end of the producer computing stage, based on data availability. Performance gains can be achieved as the consumer can process data as soon as it becomes available. Task-level pipelining may provide additional speedups over the ones achieved when exploring other forms of parallelism. In the presence of multicore-based systems, task-level pipelining can be achieved by mapping each task to a distinct core and by synchronizing their execution according to data availability. It can accelerate the overall execution of applications by partially overlapping the execution of data-dependent tasks.

This thesis proposes techniques for pipelining tasks able to deal with in-order and out-of-order communication patterns between P/C pairs. We address fine- and coarse-grained data synchronization techniques to achieve pipelining execution in FPGA-based multicore architectures and the use of inter-stage buffers to communicate data and to synchronize the cores associated with the P/C tasks. Furthermore, we propose techniques to reduce the off-chip memory accesses in the context of fine-grained data synchronization schemes. The experimental results reveal noticeable performance improvements of the multicore-based task-level pipelining approach for a number of benchmarks over a single core without using task-level pipelining.

Dates for your diary

23rd IEEE International Symposium on High Performance Computer Architecture (HPCA 2017)
22nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP 2017)
15th International Symposium on Code Generation and Optimization (CGO 2017)
4-8 February 2017, Austin, Texas, USA
http://hpca2017.org
http://conf.researchr.org/home/PPoPP-2017
http://cgo.org/cgo2017

26th International Conference on Compiler Construction (CC 2017)
5-6 February 2017, Austin, Texas, USA
http://conf.researchr.org/home/CC-2017

25th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP 2017)
6-8 March 2017, St Petersburg, Russia
http://pdp2017.org

Embedded World Conference 2017
Including a Special Session on High Performance Embedded Architectures led by HiPEAC
14-16 March 2017, Nuremburg, Germany
http://www.embedded-world.eu/home.html

Design, Automation and Test in Europe (DATE17)
27-31 March 2017, Lausanne, Switzerland
https://www.date-conference.com

30th International Conference on Architecture of Computing Systems (ARCS 2017)
3-6 April 2017, Vienna, Austria
http://arcs2017.itec.kit.edu

European HPC Summit Week 2017
15-19 May 2017, Barcelona, Spain
https://exdci.eu/events/european-hpc-summit-week-2017

10th International Symposium on High-Level Parallel Programming and Applications (HLPP 2017)
10-11 July 2017, Valladolid, Spain
https://hlpp2017.infor.uva.es

27th International Conference on Field-Programmable Logic and Applications (FPL 2017)
4-8 September 2017, Ghent, Belgium
www.fpl2017.org