HiPEAC INFO 54

COMPILATION ARCHITECTURE

MAY 2018

Computing Systems Week Gothenburg

The innovation issue: From research to market

Compilation for high-performance machine learning codes

FPGAs as a Service

contents

EU policy on digital innovation

Welcome to Gothenburg

HiPEAC voices: The art of innovation

18

HiPEAC Technology Transfer Award Winners

24

HiPEAC Vision 2018

25

Innovation Europe

This issue of the HiPEAC magazine is about innovation, something HiPEAC has been stimulating since the network was founded. In recent years, we've increased our efforts by creating technology transfer awards, by offering at least one entrepreneurial course at ACACES, our annual summer school, and by inviting successful entrepreneurs to speak at our events. Recently, we have seen an increase in startup creation by our members across Europe. Creating a successful product company in computing, scaling it up and selling it is not easy. Most companies have to wait until a key customer knocks on their door, and integrates a component into a successful product. Hence, they depend on somebody else's success. This can take a long time – in some cases, longer than an investor is ready to wait, meaning that some companies fail to survive.

However, this should not discourage us. The economy must continuously renew itself to compensate for the companies that shut down or move elsewhere. It is a normal and healthy process. In HiPEAC5, we will further support our members to launch startups or undertake technology transfer with an existing company, and we're delighted to have new partners ARTEMIS-IA and INNOVALIA on board to help us.

Our first step was to organize a Computing Systems Week dedicated to innovation in Gothenburg, where many of you will be reading this. We'll hear from startup founders about how they got their idea for the company, how the team was formed, where the initial investment came from, and how they are doing today. There will also be hands-on workshops on transforming research ideas into convincing business propositions.

To successfully launch a product or service, one should have a good idea, a lot of passion, a bit of luck and understand what it means to be an entrepreneur. That's why we've included an entrepreneurial track in this year's ACACES summer school (8-14 July) in collaboration with the TETRAMAX Innovation Action. This track will teach some of the basic skills every entrepreneur should have.

If creating a company is not an option, transferring technology to an existing company is equally valuable. The TETRAMAX project has several calls per year to help with technology transfer – find out more on p.9. Don't forget that if you're involved in a successful technology transfer in 2018, you can apply for a technology transfer award at the end of the year. Our aim is to turn the HiPEAC community into an innovation community.

Koen De Bosschere, HiPEAC coordinator

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. 779656.

Cover photo: Marcin Saj, marblemachine.org
Back cover photo: Magnus Själander
Design: www.magelaan.be
Editor: Madeleine Gray
Email: communication@hipeac.net

Innovation is the key to sustainable growth, and is our only choice if we want a thriving industry in Europe. Anne-Marie Sassen, Deputy Head of Unit for Digitising Industry at DG CONNECT, explains how the European Commission is supporting digital innovation.

Digital Innovation: a cornerstone of Eu

We Europeans have many strange obsessions. We like our welfare society, we want decent jobs and working conditions, we want to have long holidays and a good work-life balance, we want clean air and water, and we even want companies to pay taxes. So, how do you think that industry in Europe can compete with regions where salaries are low and all the other 'boundary conditions' for business are often very different?

"How can industry in Europe compete with regions where ´boundary conditions´ for business are very different?"

There are not many options available: in order to thrive and to create good and well-paid jobs, **European industry must be more innovative than its competitors**. Today, in most cases, innovation means digital innovation, and for this reason the digitization of industry is a cornerstone of the European policy. Spoiler alert: it is likely that this will be even more the case after 2020, if you consider that the European Commission, in a recent communication, recommended '**doubling the amounts currently invested in the digital economy to around EUR 70 billion over the period 2021-2027**'.

The name of the game is '**Digitising European Industry**'. It is an initiative aiming at making sure that **any business in Europe**, whatever its size, location or market sector, can have **easy access to the digital technology they need for their work**. Today this is not always true: while high-tech companies, that we in the HiPEAC community know well, are extremely advanced in the use of digital technology for every aspect of their business, the same cannot be said for the many small and medium enterprises (SMEs) that are the fabric of the European economy. Those SMEs are often very knowledgeable in their own field, but lack the resources and knowhow for introducing digital technology to reinvent their business; the same is often true for larger companies operating in more 'low-tech' sectors like agriculture or construction.

The European Commission wants to tackle this problem and to support the digital transformation of the economy. The most important weapon we have is a three-letter acronym: **DIH**, which stands for **Digital Innovation Hubs**. The DIH are the access doors of companies to the world of digitized industry, or 'industry 4.0'. They work together with **Competence Centres**, like universities or research organizations, to help SMEs in understanding, absorbing and deploying the digital technology they may need. Basically, they provide the necessary **link between technology and business requirements**.

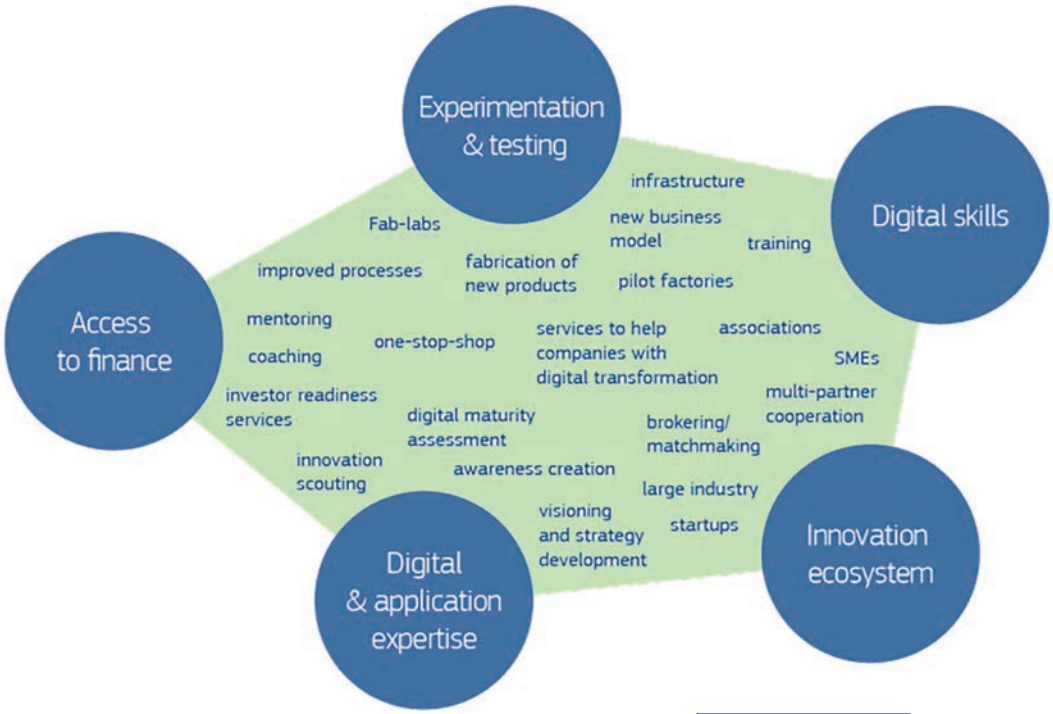
There are already many DIH across Europe, in most cases supported by national and regional initiatives for the digitization of industry like Plattform Industrie 4.0, Industrie du Future, Smart Industry, Piano Industria 4.0. The European Commission contributes to this effort by creating a network – so that, for example, an SME from Portugal can get access to a technology provider in Baden-Württem-

today and tomorrow, uropean policy

berg – and by managing working groups that bring together the relevant stakeholders to define together the way ahead.

The European Commission supports the network of DIH financially through Horizon 2020. In the workprogramme for the coming years you can find six calls (DT-ICT-01 to DT-ICT-06) dedicated to Digital Innovation Hubs in areas including photonics, robotics, big data, Smart Anything Everywhere and I4MS (ICT Innovation for Manufacturing SME).

It is easy to see that this is a big opportunity for the HiPEAC community: you represent the technology expertise that must necessarily be at the core of any Digital Innovation Hubs, together with the complementary business expertise. HiPEAC involvement is already a reality in several cases; one example is the project TETRAMAX, which is very active in the HiPEAC community and is funded within the framework of Smart Anything Everywhere. In the future, thanks to the European network of DIH, we can expect many more interesting opportunities to use digital technology to improve the production processes, the products and the business models of European industry. Any business sector will be affected, so stay tuned; it will be worth it.



FURTHER INFORMATION:

Catalogue of Digital Innovation Hubs
http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs-tool

Digital Innovation Hubs on the Europa website
https://ec.europa.eu/digital-single-market/en/digital-innovation-hubs

SAE
SmartAnythingEverywhere

I4MS

Computing the future in Manchester at HiPEAC18

With over 530 participants from 35 countries, the thirteenth HiPEAC conference demonstrated the breadth and depth of expertise within the HiPEAC community once again. Taking place in Manchester on 22-24 January 2018, it included sessions on everything from machine learning to secure and safe real-time systems, from exascale computing to the latest cyber-physical innovations.

As HiPEAC18 General Co-Chair John Goodacre (University of Manchester/Arm) noted, Manchester was a fitting location. 'Manchester has a long history of great inventions: the idea that a computer would run a program out of stored memory was done here, and the idea that it could run more than one application at the same time was demonstrated here. More recently, the SpiNNaker programme has been trying to simulate what a brain looks like with a million Arm processors.'

Keynote talks from Maria Girone (CERN openlab), Dileep Bhandarkar (Qualcomm Datacenter Technologies) and Daniel Belov (DeepMind) gave insights into computing challenges at CERN's Large Hadron Collider, emerging trends in data centres and the fascinating world of machine learning.

Artificial intelligence was also a major theme in many of the workshops, as Sandro D'Elia of the European Commission noted. 'This is a clear demonstration that this is a technology which is coming of age and which will change our lives in the future,' he commented. Sandro gave the HiPEAC community an overview of European research directions in digital technology, as well as highlighting trends in cyber-physical systems.

Complementing the conference programme – comprising the ACM TACO paper track, 27 workshops and eight tutorials in addition to the keynote talks – the exhibition once again highlighted university, project and industry-led research and innovations. Multinational companies including Arm, DeepMind, Atos and Samsung were joined by European small/medium enterprises (SMEs) in the industry exhibition, while many companies took advantage of the industrial session to pitch their work.

In addition to exchanging ideas and finding new clients, many companies came to the conference to attract candidates. In this, they had support from HiPEAC's recruitment services at the HiPEAC careers unit, featuring job offers from across Europe. For the first time, the conference also included a science, technology, engineering and mathematics (STEM) student day, where undergraduate students had the chance to learn about next-generation computing systems and find out what opportunities HiPEAC could offer them.

The HiPEAC team would like to thank our sponsors, who contributed a record amount to this year's event, and without whose support the event could not have been such a success.

Photos from the event can be viewed in the Google Photos album: bit.ly/HiPEAC18_photos

Visit the HiPEAC YouTube channel for videos from the event, including full-length videos of the keynote talks: bit.ly/HiPEAC18_videos

HiPEAC19 call for papers – deadline 1 June

The next HiPEAC conference will take place on 21-23 January 2019 in Valencia. Papers can be submitted to ACM TACO (Transactions on Architecture and Code Optimization) throughout the year. If you submit your paper by 1 June and it is accepted, you will be invited to present at HiPEAC19.

From 1 July 2018 to 1 July 2022, ACM TACO will be open access, meaning that all past and future papers, including papers accepted to this call, will be open access to all until July 2022 (at no cost to the authors).

Välkommen till Göteborg!

A young city which has nurtured a dynamic research and innovation ecosystem: Per Stenström, professor and head of division, Computer Engineering at Chalmers University of Technology, explains what makes Computing Systems Week host city Gothenburg so special.

Why is Gothenburg a hotspot for innovation?
With its maritime history and tradition as Sweden's biggest import/export centre, Gothenburg is the industrial centre of Sweden. The city has a real entrepreneurial energy, thanks in part to the automotive ecosystem which has been here for many decades. Several major international companies – including Volvo, manufacturing company SKF and Ericsson – have either their headquarters or a significant presence in the city. There has been significant investment in innovation, resulting in a strong innovation ecosystem with incubators. Here at Chalmers, our technology transfer office supports researchers to bring their ideas to practical use.

What are some of the most exciting projects at Chalmers at the moment?
Chalmers coordinates the Graphene Flagship project, a huge international collaboration with a €1 billion budget funded by the European Commission which will help exploit the potential of graphene – which is light, strong and flexible – to create new products.

Closer to the HiPEAC world of computing, the Knut and Alice Wallenberg Foundation – the largest private sponsor in Sweden – has funded research in autonomous vehicles in several Swedish universities, including Chalmers. The same foundation has invested one billion SEK in a quantum computing research initiative led by Chalmers, as well as funding a programme in artificial intelligence. As a small country, Sweden has traditionally invested a lot in high-tech research, which provides most of our opportunities for growth.

Photo © Per Pixel Petersson/Göteborg & Co

Name three things we probably didn't know about Gothenburg.
You might not know that Gothenburg is a relatively young city, by European standards; it was founded in 1621. The nascent city was formed by Dutch planners, who built the canals, and in fact the city was largely led by Dutch immigrants in its early years.

As for Chalmers, the name comes from its founder, William Chalmers, who was of Scottish descent. Strong trade links between Britain and Gothenburg important led to the Swedish city being referred to as 'little London'. William Chalmers was involved in building up the British East India Company, which traded with Asia, and made a fortune in the process. In his will, he left money for the foundation of a technical school for poor, talented youngsters. Chalmers only gained university status in 1900; so, like the city of Gothenburg, it is relatively young.

Where's the best place to grab a beer after a hard day at CSW?
There's a street called Avenyn – the Avenue – which really comes alive in May, when Swedes look forward to long summer days after the melancholic Swedish winter. With lots of outdoor bars, it's a great place to people watch.

Photo © Anders Wester/Göteborg & Co

Test your knowledge of Swedish

What is a 'fika'?
a) A troll.
b) A coffee break.
c) A swear word.

How do you say 'parallel computing rocks' in Swedish?
a) Jag är bra på korsstygn
b) Parallell parkering är min specialitet
c) Parallella datorer är super

Unleash your inner entrepreneur at ACACES 2018

HiPEAC has teamed up with TETRAMAX and EuroLab-4-HPC to offer you our best summer school yet. This year, ACACES, the summer school on Advanced Computer Architecture and Compilation for High-Performance and Embedded Systems, offers a TETRAMAX track focusing on innovation, with courses on intellectual property strategy, business prototyping and translating technology into commercial products and services. Meanwhile, EuroLab-4-HPS is organizing a special track focusing on high-performance computing (HPC), including courses on memory systems and graphics processing units (GPUs). All this in addition to a keynote by by Partha Ranganathan (Google) and courses on topics such as hardware architectures for deep neural networks and cybersecurity in relation to hardware.

Photo credit: Mahwish Arif

The summer school takes place in Fiuggi, Italy on 8-14 July 2018.

Further information: acaces.hipeac.net/2018

EuroLab-4-HPC continues building the HPC ecosystem

May saw the start of the EuroLab-4-HPC 2, the follow-up to the European Commission-funded Coordination and Support Action in the field of high-performance computing (HPC) ecosystem development.

'The overarching goal of the second edition of EuroLab-4-HPC is to strengthen academic excellence and innovation in HPC systems through a number of measures such as innovation events/programs, summer schools and research vision building activities.

The agenda of EuroLab-4-HPC is of key importance now that European Commission is investing more than a billion euros in HPC,' says EuroLab-4-HPC Coordinator Per Stenström.

In this new phase of the project, EuroLab-4-HPC will offer cross-site mobility grants to enable researchers to spend time at another research centre. The project is working on a long-term research roadmap for HPC, as well as helping define the HPC curriculum and providing support for technology transfer.

Further information: eurolab4hpc.eu

Get smarter with FED4SAE

Isabelle Dor, CEA-Leti

Through the Smart Anything Everywhere initiative, the European Commission is helping digitize European industry. FED4SAE is part of this strategy, targeting a large network of 'small' companies (startups, small/medium enterprises and midcaps), including both technology specialists and low-tech companies. These companies can request solutions for specific use cases, which FED4SAE's industrial and advanced platforms will provide.

FED4SAE also helps participants with innovation management, providing coaching and opportunities for networking to identify the best stakeholders – private investors, national/regional initiatives, business angels, potential customers, and so on. In summary, the initiative offers companies:

- technical expertise via **advanced platforms**
- product support via **industrial platforms**
- **innovation management**, focusing on business
- up to €60,000 in funding

Advanced platforms provide expertise, innovative technical solutions and/or testbeds which add value to the product. **Industrial platforms** give access to leading-edge technologies by market leaders in cyber-physical and embedded systems, which can bring the innovation to a state of maturity. **Innovation management** support helps get your innovation to market, via the FED4SAE consortium and the Smart Anything Everywhere ecosystem.

Experiments will be funded on applications such as smart cities, smart energy, smart manufacturing, smart mobility and smart health systems. Their solutions will comprise components such as sensors, data fusion/processing technology and actuators – the building blocks of cyber-physical systems. By definition, cyber-physical systems are feedback loop systems and therefore require a cybersecurity approach.

There are a further two calls: one from 29 May - 18 September 2018, and another to be launched in November 2018, which will last 12 weeks.

Further information: fed4sae.eu

FED4SAE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 761708

Innovation starts with TETRAMAX

Katrien Van Impe, Dissemination and Communication Officer, TETRAMAX

The first TETRAMAX call for bilateral technology transfer experiments (TTX) in customized low-energy computing closed at the end of February, attracting 36 applications from all over Europe with innovative ideas and budding international collaborations. Meanwhile, the first call for value-chain oriented and interdisciplinary TTX was launched at the end of February with a deadline at the end of May, and we look forward to the outcome of this multi-partner TTX call.

However, TETRAMAX offers more than funding. The project is investing in a novel tech brokerage portal: tetramax.eu/brokerage. The 23 TETRAMAX Competence Centres form the backbone of a European network with competence in all fields of customized low-energy computing for cyber-physical systems and the internet of things. Each Competence Centre is connected to Local Ecosystem Partners (LEPs) and Digital Innovation Hubs (DIHs).

Whether you are looking for an industry partner to experiment with your technology or you need a solution for a specific problem, the portal can help you find the right match for your innovative digitization projects. Our experienced Chief Technology Broker, Wei Jan Wang, is on hand to facilitate this matchmaking. Alternatively, you can contact your regional or national contact point to begin browsing potential partners. Visit tetramax.eu/brokerage for more information.

Earlier this year TETRAMAX organized workshops or presentations at events including the HiPEAC conference in January and the DATE 2018 conference in March. In July, we are holding a special TETRAMAX track at the ACACES 2018 summer school in Fiuggi, Italy. This will include courses on business prototyping, intellectual property strategy and technology transfer by Henrik Berglund, Marcus Holgersson (Chalmers University, Sweden) and Uday Phadke (Cartezia, UK). Register for this TETRAMAX track via the ACACES website: acaces.hipeac.net/2018

'Enabling ICT Innovations for European SMEs' session at DATE: Robin Schubert (BASELABS), Rainer Leupers (TETRAMAX), Isabelle Dor (FED4SAE), Bernd Janson (ZENIT), Juan Eusse (SILEXICA) and Luca Fanucci (Università di Pisa)

Further information: tetramax.eu

TETRAMAX has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 761349

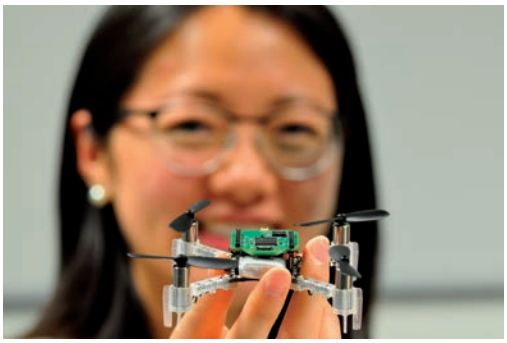
Bridging the gap with industry and innovation in HiPEAC5

The latest phase of HiPEAC will focus on reaching out to the industrial and innovation communities, the latter comprising those who help transform research results into products and services, such as digital innovation hubs, technology transfer officers and investors. We are delighted to welcome new partners ARTEMIS Industry Association and Innovalia to help us in this task.

Ad ten Berg, Office Director of ARTEMIS-IA says: 'ARTEMIS Industry Association is happy to be a partner in HiPEAC5, as HiPEAC is a widely recognized network in the domain of high-performance and embedded systems, which fits very well with the ARTEMIS focus on embedded intelligence. We strongly believe that the collaboration between HiPEAC and ARTEMIS will deliver a contribution to keep Europe at the forefront in the worldwide technology race to make products and systems more intelligent.'

'Innovalia's participation will bring new European industrial and innovation communities to HiPEAC, providing new areas of implementation and identifying new market needs, which will give the project higher impact. These new members will benefit from high-end digital technologies, expertise and the development of skills, which aligns perfectly with the Digitizing European Industry initiative and the intensive digitalization strategy in different sectors within Europe,' says Silvia de la Maza, Chief Innovation Officer, Innovalia.

If you're planning to prepare a project bid in electronic components and systems, why not try the ARTEMIS-IA ECS Collaboration Tool? Find your ideal balance of industrial and academic partners, then assess the best funding mechanism for your project. To find out more, visit ecscollaborationtool.eu.



PULP celebrates five years

Frank K. Gürkaynak, ETH Zürich

This year, ETH Zürich and the University of Bologna are celebrating five years of collaboration on the PULP (Parallel Ultra-Low-Power) project. To mark the occasion, they've announced three updates:

PULPissimo

An improved version of the project's RISC-V-based, open-source 32bit microcontroller system, which offers new features including an autonomous input/output subsystem, a new memory subsystem, support for hardware accelerators and a new interrupt controller.
github.com/pulp-platform/pulpissimo

Ariane

A Linux-ready, application-class 64bit RISC-V core supporting RV64, written completely in System Verilog.
github.com/pulp-platform/ariane

OPENPULP

The project's most advanced open-source release to date: an ultra-low power 'host' coupled with a powerful compute engine based on a tightly coupled cluster of eight cores. It features a low-latency memory interconnect enabling energy-efficient data sharing, an advanced direct memory access engine and an event unit for hardware-optimized synchronization and implementation of primitives typical of parallel programming models. In addition, it includes an energy-efficient shared instruction cache and support for shared-memory hardware accelerators.
github.com/pulp-platform/pulp

Keep up with the latest updates by visiting our website and following us on Twitter.
pulp-platform.org
@pulp_platform

Book: Hardware Security and Trust

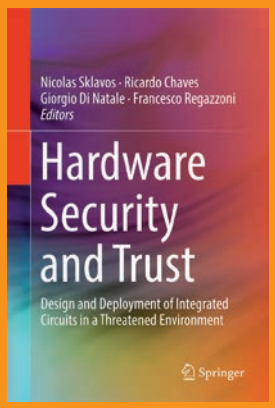
Design and Deployment of Integrated Circuits in a Threatened Environment

Nicolas Sklavos, Ricardo Chaves, Girogio Di Natale , Francesco Regazzoni (eds.)

This book provides a comprehensive introduction to hardware security, from specification to implementation. Applications discussed include embedded systems ranging from small RFID tags to satellites orbiting the earth. The authors describe a design and synthesis flow, which will transform a given circuit into a secure design incorporating counter-measures against fault attacks.

In order to address the conflict between testability and security, the authors describe innovative design-for-testability (DFT) computer-aided design (CAD) tools that support security challenges, engineered for compliance with existing, commercial tools. Secure protocols are discussed, which protect access to necessary test infrastructures and enable the design of secure access controllers.

Further information:
bit.ly/HW_security_trust



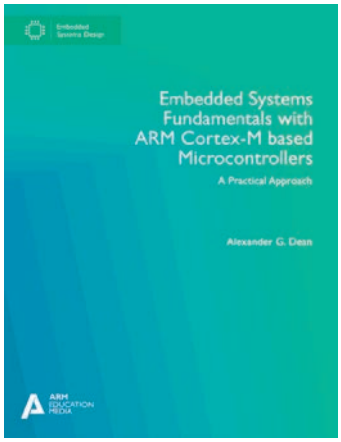
Arm Embedded Systems Textbook

Arm Education Media, a new publishing operation within Arm, has published its first textbook, for use in electrical engineering, electrical and computer engineering, and computer science departments.

Embedded Systems Fundamentals on Arm Cortex-M based Microcontrollers: A Practical Approach (ISBN 978-1-911531-03-6), by Dr Alexander G Dean (North Carolina State University), provides abstract examples that work on a real board, and introduces students to creating embedded systems using the Arm Cortex-M0+ CPU-based Kinetis KL25Z MCU. Topics covered include the CPU, interrupt system, peripherals, and programming. The concurrent operation of the CPU and peripherals is highlighted throughout as critical to creating cost-effective embedded systems. The book gives an early introduction to practical multitasking on the CPU, with the goals of improving responsiveness and software modularity while reducing CPU overhead. The interplay of interrupts, peripherals and schedulers is examined.

The textbook is available in print and digital formats, via Amazon and Bookshelf Online, respectively.

Further information:
armedumedia.com



Matthias Jung receives EDAA Outstanding Dissertation Award

Congratulations to HiPEAC member Matthias Jung, who received a European Design and Automation Association (EDAA) Outstanding Dissertation Award at DATE Conference 2018 for his PhD thesis 'System-Level Modeling, Analysis and Optimization of DRAM Memories and Controller Architectures'. The award was in the category of 'New directions in system-on-chip platforms co-design, novel emerging architectures and system-level management'. Matthias wrote his PhD thesis at the University of Kaiserslautern under the supervision of Prof. Norbert Wehn and Prof. Bruce Jacob (University of Maryland).

The dissertation will be published by Springer in the EDAA Outstanding Monographs series.

Further information: edaa.com

Michael O'Boyle awarded EPSRC Senior Research Fellowship

HiPEAC steering committee member Michael O'Boyle (University of Edinburgh) has been awarded a senior Research Fellowship by the UK's Engineering and Physical Sciences Research Council (ESPRC). In response to the end of Moore's Law and the resulting specialist, heterogeneous nature of computing hardware, Mike's fellowship aims to rethink how we connect software and hardware by introducing a more flexible language interface which can change from one processor to the next. This will allow existing software to use future hardware and allow hardware innovation to connect to new and emerging application areas such as robotics, augmented reality and deep learning.

The fellowship runs until 31 March 2023 and is in partnership with Arm, Codeplay, Inria, Microsoft, Northeastern University and the University of Texas at Austin.

Congratulations on behalf of the HiPEAC community!

Further information: bit.ly/Heterogeneous_Thinking

Dates for your diary

European HPC Summit Week
28 May – 1 June 2018, Ljubljana, Slovenia
exdci.eu/events/european-hpc-summit-week-2018

MECO 2018: 7th Mediterranean Conference on Embedded Computing
10-14 June 2018, Budvar, Montenegro
Featuring HiPEAC workshop and EUROMICRO/IEEE workshop on embedded and cyber-physical systems
embeddedcomputing.me/en

Euro-Par 2018: 24th International European Conference On Parallel and Distributed Computing
27-31 August 2018, Turin, Italy
europar2018.org

Euromicro DSD/SEAA: Euromicro Conferences on Digital System Design and Software Engineering and Advanced Applications
29-31 August 2018, Prague, Czech Republic
dsd-seaa2018.fit.cvut.cz

Embedded Systems Week
30 September – 5 October 2018, Turin, Italy
esweek.org

IoTSMS 2018: Fifth International Conference on Internet of Things: Systems, Management and Security
15-18 October 2018, Valencia, Spain
Submission date: 10 June 2018
emergingtechnet.org/IoTSMS2018/index.php

IEEE Nordic Circuits and Systems Conference 2018
30-31 October 2018, Tallinn, Estonia
Submission date: 24 August 2018
norcas.org

What are the secrets to transforming your research into innovations which have an impact on society? HiPEAC spoke to start-up founder Koen Bertels (TU Delft), innovation adviser Wei-Jan Wang (Chalmers University), technology transfer specialists Anna Escoda and Cristina Calatayud (Barcelona Supercomputing Center), and patent examiner Sylvain Lelait (European Patent Office) to find out.

The art of innovation

THE RESEARCHER TURNED START-UP FOUNDER

HiPEAC member Koen Bertels is head of the Quantum and Computer Engineering Department and head of the Quantum Computer Architecture Lab at TU Delft. He has launched two start-ups, Bluebee High Perfomance Genomics and Upsilon, as well as collaborating with the research centre QuTech. He is also involved in the TETRAMAX Innovation Action on technology transfer.

What are the challenges involved in transferring technology to the market?

First, as a scientist, you really need intellectual property of your results. So rather than publishing in a conference or journal, you need to ask yourself 'Do I need to protect this know-how, or this result?', because otherwise anyone could take your paper and do whatever they want with it.

The second key challenge is to find money, and that takes time. It's not like 'Oh, I have a great result, and the millions will just start flowing in as soon as I do something' – the 'doing something' is a big problem in itself. So another challenge is to define the product that you want to build. Many companies have a product in the initial phase which then evolves quite radically in the first one to two years because they have a completely wrong image of what the market actually expects.

> ***"We have to be conscious about where we want to stand in ten years, and realize that we have to start investing in quantum right now"***

The final challenge is that you have to be patient; you need to wait three, four, five, six years before you get any kind of meaningful result.

Your startup Bluebee was launched in 2011 and has raised €10 million in Series A funding. How did you go about starting the company?

Bluebee emerged out of my research at TU Delft on hardware accelerator technology combined with classical processors. The research came out of a European Commission-funded project; my colleagues at Delft and I thought the results were too good not to do anything with it, so we decided to start a company.

It took at least two and a half years before I pulled new people into the company. One is my colleague at Delft, Zaid Al-Ars, who works on genome sequencing. Genome sequencing is the future of medicine: doctors will be able to look at your genetic profile and extract the specific characteristics of whatever illness you may have, and we created a hardware accelerator technology to do that rapidly.

The second person I pulled in is a professional chief executive officer (CEO) – and that's also why you need money, because these people are quite expensive. Now, almost four years down the road, we've had substantial investment in the company. It's a bit too early to say that we're making a profit, but we did open an office last year in the US, and within the year we had three major contracts with US – not European, but US – companies.

This is an exciting time for your current main research interest, quantum computing. Why is it important to invest in quantum now?

Quantum technology is a futuristic technology in the sense that it will be at least 10 to 15 years before anything substantial is available on the market. QuTech, our research centre in Delft, is heavily sponsored by several US companies. Specifically, Intel sponsors our line of research. Now, does that mean there is absolutely no future for the European Union? No, but we would need substantial investment.

The key thing is to start now; it takes time before you create a team, before you have ideas, before you actually understand what quantum is all about and then you still have to think about what the product is. You may need people to do PhDs – so that's already four to five years – plus two to three years understanding your own technology, plus two to three years making the product. So you're easily talking about 10 years of investment for yourself, before you can do anything meaningful with that kind of technology.

Another reason to invest early is the reverse triangle, which means that the added value of services you can build on top of this technology will grow. And that's where the EU should be taking the initiative. We have the €1 billion Flagship on Quantum Technology, which is very good, but if you compare it to the budget that a company like Intel or even IBM has, it's basically peanuts. So we have to be very conscious about where we want to stand in ten years, and realize that we have to start investing ourselves, right now.

Video versions of this interview are available on the HiPEAC YouTube channel: bit.ly/HiPEAC_experts_YouTube

THE UNIVERSITY INNOVATION ADVISER

Wei Jan Wang is an innovation adviser at Chalmers University, helping the university's research results reach industry and society.

How do you promote innovation at Chalmers?

My job involves supporting the utilization and movement of research results from academia to industry and society, in order to create impact and improve lives. While no two projects are the same, the general process is similar. Working closely with researchers, we first work to understand the technology and generate hypotheses on target users, problems, and potential benefits of the solution. Open discussions with potential users are held to verify the hypotheses, and if a strong problem/solution fit is found, we proceed to verify the technology via prototyping or mockups. Along the way, there is support from other actors in the innovation system who provide business

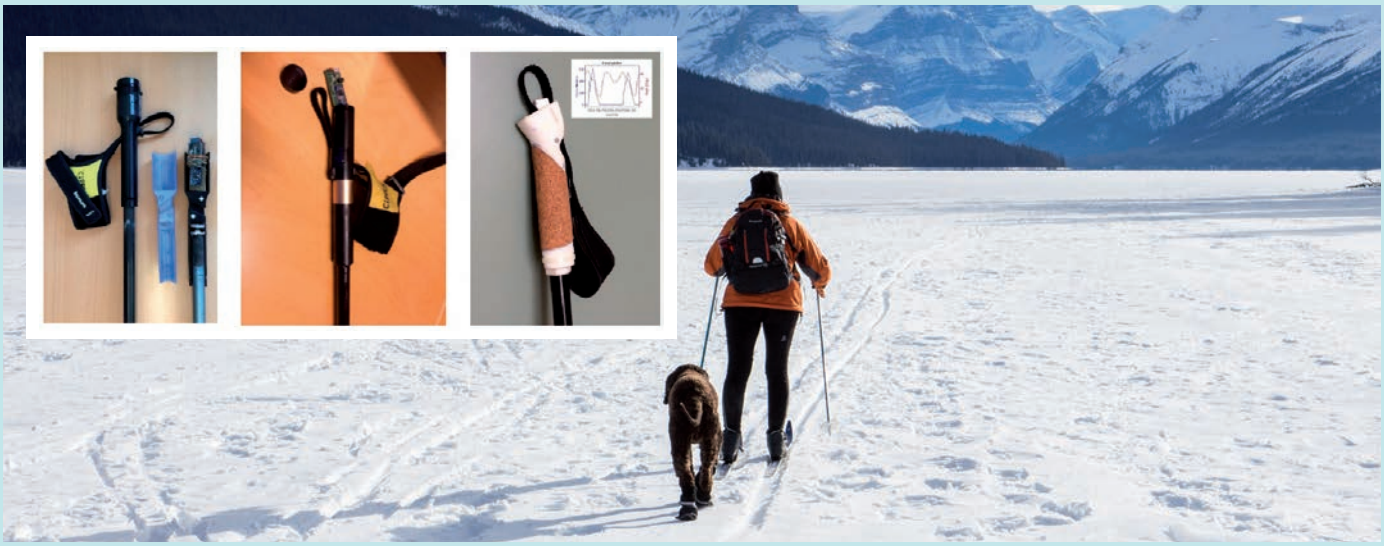


Photo by Jasper Guy on Unsplash

Ski poles get smarter thanks to a research project at Chalmers

incubation services, soft and risk capital financing, as well as a network of entrepreneurs and business advisers.

What successes have you had?

One example is a project involving power sensors for cross-country skiing. Currently, it is difficult for professional and amateur skiers to gather information on the power and angle of attack of the ski poles, which is important for technical analysis and improvement. The product – which is based on a master's degree research project – has a unique design and consists of wireless communications and micro-electrical mechanical system (MEMS) sensors. An understanding of the value chain, as well as of users and their expectations with regards to the measurements and graphical user interface, was important for success.

> *"TETRAMAX brings together leading academic expertise and the latest innovations in the internet of things and cyber-physical systems"*

You're involved in the TETRAMAX Innovation Action. How will this help stimulate innovation?

As the largest consortium for customized low energy computing in Europe, TETRAMAX provides a unique opportunity. This network brings together leading academic expertise and the latest innovations in the internet of things and cyber-physical systems, which is further supported by local support organizations including industry partners, small/medium enterprises (SMEs), startups and incubators. Innovation is facilitated by funding for technology transfer experiments, our online brokerage portal and service and extensive network, as well as continuous improvements through the sharing of best practices.

What makes Gothenburg an innovative place?

Gothenburg is a great city to live and work in, which attracts the talent and diversity that drives innovation and creativity. There is also a dedicated innovation system including incubators, science parks, research institutes, and state-funded organizations such as Business Region Gothenburg. These factors, coupled with the long history of research excellence, strong collaboration ties with industry, and an emphasis on sustainability, gender equality and innovation at the university, allow innovation to thrive.

innovationskontor.chalmers.se/en

THE HPC TECHNOLOGY TRANSFER OFFICE

Anna Escoda (pictured below right) and Cristina Calatayud (pictured below left) make up the technology transfer office for computer sciences at Barcelona Supercomputing Center (BSC).

What is the main aim of technology transfer?

Anna Escoda (AE): Technology transfer is about making sure research results reach industry, but also that they have an impact on society – although in order to reach society as a whole, they often need to go via industry.

High-performance computing is a crosscutting tool which provides different industry sectors with greater computing power in order to resolve complex problems. One example would be simulations of mechanics involving more than one variable, such as computational fluid dynamics. Another is genome analysis, which requires a high number of calculations to identify mutations, such as those related to cancers. This would require weeks of computation on conventional computers, as opposed to hours on a supercomputer.

What are the main services you offer?

Cristina Calatayud (CC): Our work focuses on both researchers and businesses. We work with the researchers to offer them

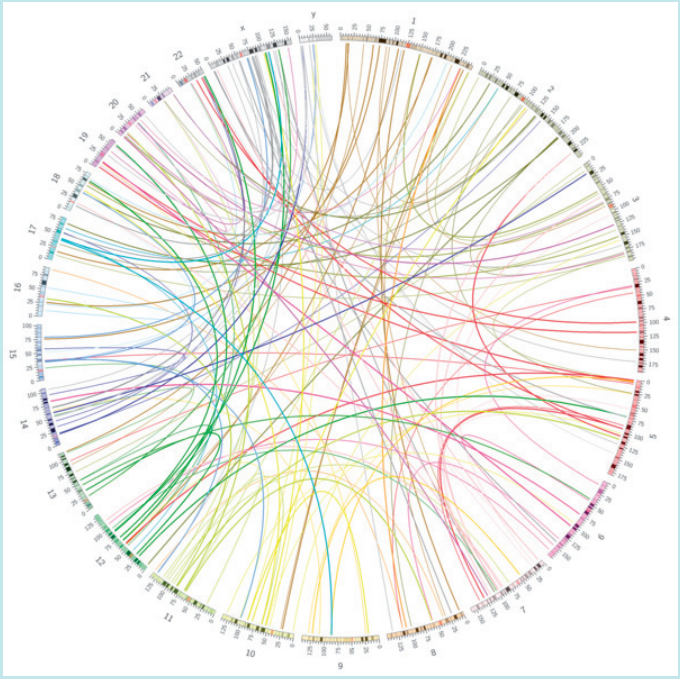
advice on intellectual property rights (IPR) management and the exploitation of their results. It's very important to establish a strategy based on what you want to achieve with a specific research result and the characteristics of the result itself. To better understand the context, we can commission market studies, patentability assessments and/or business plans, for example.

Normally, the first thing to decide is whether to go through the open or the protected route. If you want to work with the scientific community to advance the state of the art, an open route makes senses. However, if the result can give a company a competitive advantage, it's important to protect the result with the appropriate IPR and establish an exploitation agreement with the interested parties.

AE: Our work with businesses focuses on two main targets: large IT companies with whom we form long-term relationships, such as joint research centres, and small/medium enterprises (SMEs), which we try to involve in research, innovation and development (R+D+I) projects. We bring researchers and companies together, organizing in-house events and participating in brokerage events to find out more about industrial problems that our researchers could help solve.

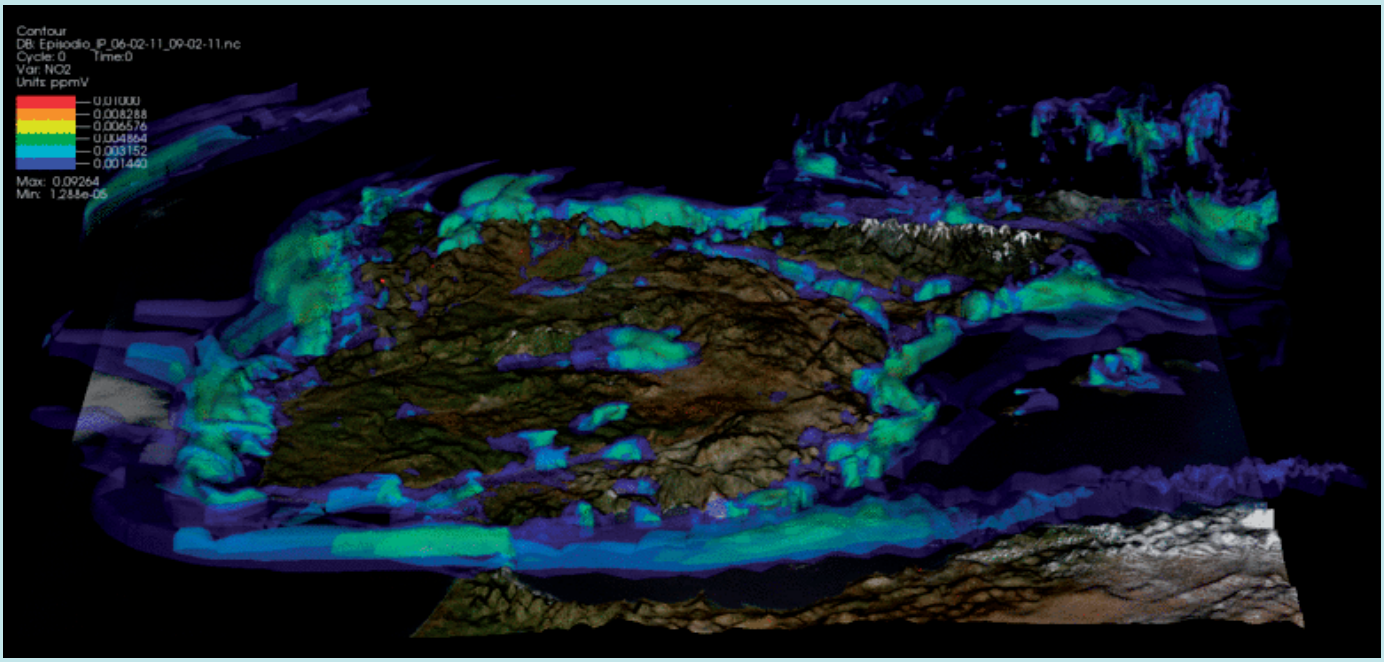
What we offer companies isn't so much infrastructure; they already have access to cloud services. Instead, we offer innovation and research services in the field of parallel computing.

CC: We aim to keep track of the technology being developed by the different departments at BSC, and help the different research groups identify the most promising results with regard to exploitation. We also help write the exploitation element of bids



BSC's services include air quality studies and computational genomics

for Horizon2020 and other types of projects. One was recently accepted for the European Commission's Future and Emerging Technologies (FET) Innovation Launchpad call. As part of this project, named QUAKE (Qbeast Utility Analysis to Market and Enterprise), we'll be helping prepare Qbeast, a scalable multi-dimensional indexing system that enables efficient multidimensional queries of data, for the market. Finally, we offer training for researchers – we've just launched a six-session training course called the Innovation Journey, which will teach researchers about IP protection and build their entrepreneurial skills.



What have been your greatest successes to date?

AE: Our work has focused on bringing a culture change towards an entrepreneurial way of thinking; thanks to this, an increasing number of researchers now consider an entrepreneurial career. Through initiatives such as the Innovation Journey training, we aim to help researchers understand the market and identify the problems their research could respond to. The markets we work with are still being created, which makes for a fast-paced environment.

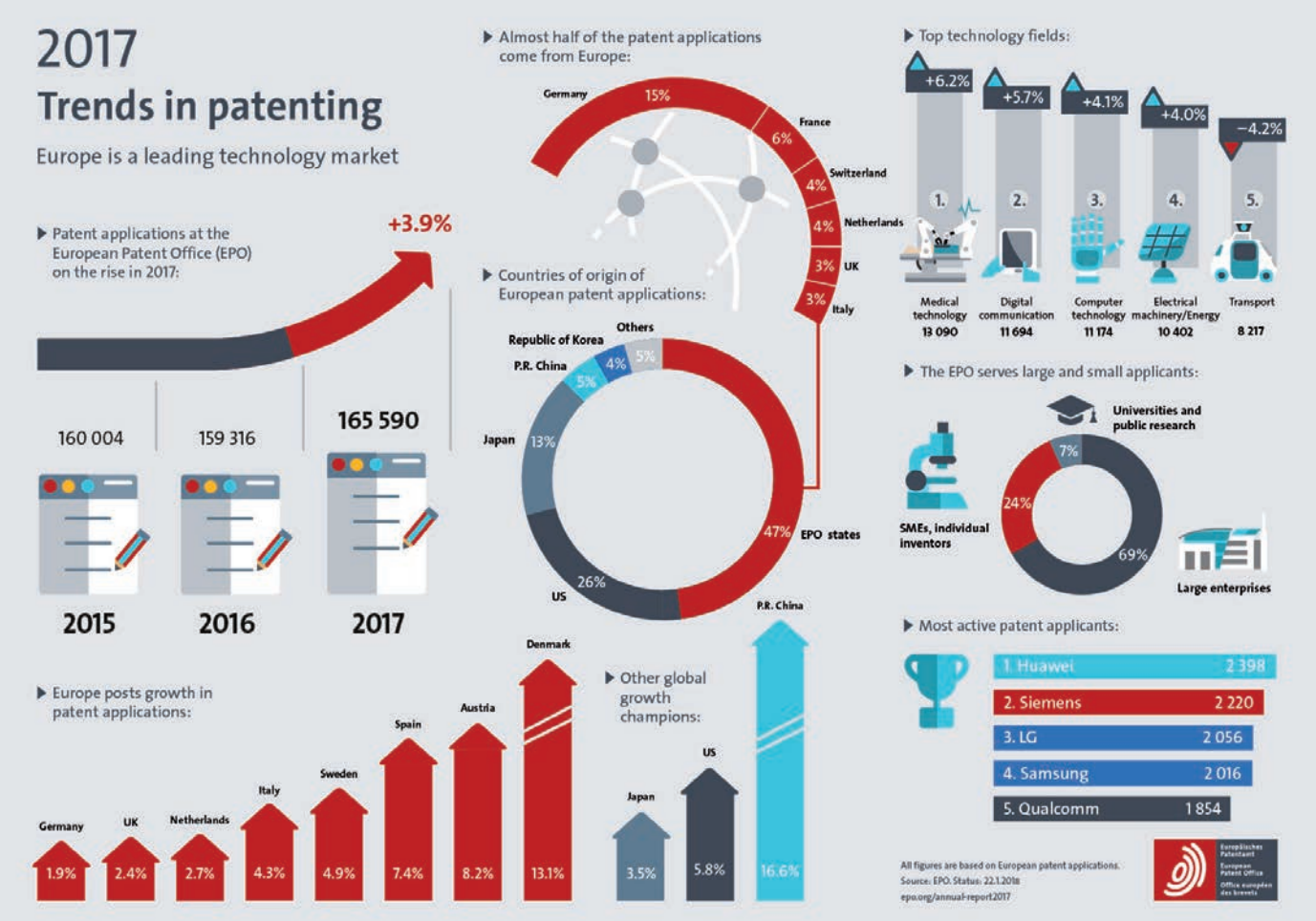
BSC has a rich portfolio of services to offer, spanning everything from air quality studies to genome analysis. We've helped researchers get more than 15 patents and we were instrumental in the launch of BSC's first spin-off, Nostrum Biodiscovery, with a second in the pipeline. Last year, we had 11 patents approved or pending approval, five joint research centres and 26 bilateral collaboration agreements with companies.

> *"Industrial PhDs can bridge the gap between HPC research and mature industries"*

What are the main challenges you have to face?

AE: Reaching SMEs and making it easy for them to get involved in the world of HPC is a major challenge. When we work with emerging technology companies, such as those working in biotech, smart cities or the internet of things, we share a common language. Barcelona has a digital R+D+I cluster comprising companies working in areas like big data, fintech and biotech, who find it easier to grasp the scientific value offered by BSC.

Reaching out to companies in mature industries is more difficult. Industrial PhDs are one way of bridging the gap, where the student's thesis provides a solution to improve an industrial process. If a good relationship with the student is established, it helps bring the company closer to the world of HPC.



Incubating innovation

Kemal A. Delic, Senior Technologist, Hewlett Packard

The history of humankind is marked by our ability to think creatively and deploy this thinking to improve our lives. By accelerating, augmenting and amplifying human capabilities, technology has had an impact on every aspect of human society.

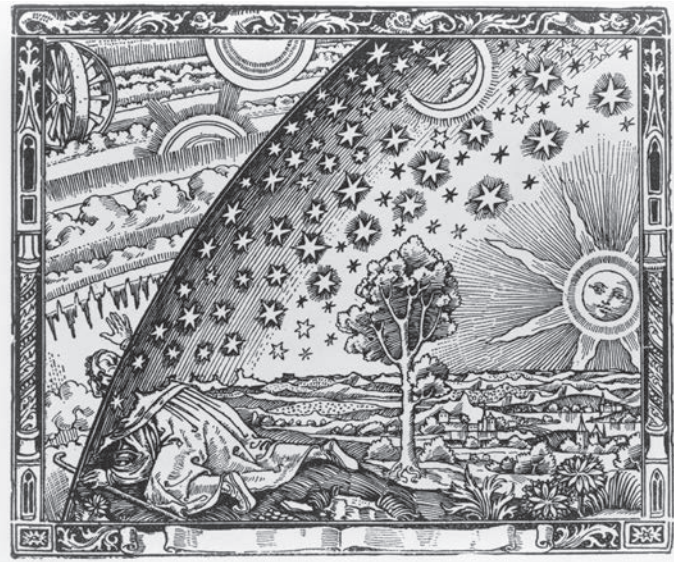
Key innovations which have shaped industrial progress include harnessing steam power in the eighteenth century; the invention of the telegraph in the nineteenth century; bringing electricity into everyday use in the nineteenth and twentieth centuries; and the introduction of computing, then the internet, in the twentieth and twenty-first centuries. As for scientific advances, the invention of the telescope (1586) and microscope (1590) spawned an entire new wave of discoveries, thanks to the light they shed on mysteries of the deep universe and the molecular world.

At the beginning of the twentieth century, the word 'computer' referred to a person who carried out numerical operations, mostly for accounting purposes. Today, it is an omnipresent device with multiple uses, from scientific instrument to business engine. In the future, artificial intelligence algorithms running on supercomputers performing 10 to the power of 18 calculations per second, fed by endless torrents of big data, will likely solve some of the biggest scientific questions of our day. This incredible amplification of human cognitive and sensory capabilities has been brought about mostly by incremental innovations.

Ideas to innovations

Think of how many ideas you have a week; if you're a creative thinker, you may have one a day. Of those, for the sake of argument, perhaps one in ten will be worth writing down; one in 100 might lead to a paper; one in 1,000 might be worth filing a patent on; and one in 10,000 might merit starting a company. Even if you do patent an idea, bear in mind that only one patent in 500 actually makes money for the inventor.

For companies, innovation gives a competitive advantage, or represents a technology basis for defensive or offensive purposes. Cultivating a spirit of innovation and establishing a process to select and scale ideas is therefore crucial. In such a process,



specialist workshops are held to trigger ideas which are then discussed further, creating clusters of interesting ideas. The ideas are then prioritized according to their level of inventiveness and then compared to ideas in either big technological libraries or patent repositories to establish the level of relevance to the company's business. A filtered list of refined ideas is then discussed, until they reach a level of maturity for patent filing. Patents are arranged into portfolios and carefully managed in a well-established intellectual property (IP) process.

Through this and other sources of innovation, thousands of ideas can be incubated before being tested in prototypes, trials or experiments, resulting in one or two products or services that are a major success.

Of course, the process will not always work in this way; it is often a long, winding road of testing, trials and persistence to achieve successful innovations. For example, the inventor of Ethernet took one year to develop his idea and convince corporate decision makers that it would create a new technology market – but it took another ten years to prove that he was right. However, a structured approach to innovation creation and IP management will significantly increase the chances of success.

FURTHER READING:
The Art of Innovation – Making Big Bets
http://bit.ly/Big_Bets_Innovation_KD

From real-time sound propagation simulation to migraine prediction technology, the 2017 HiPEAC Technology Transfer Award winners demonstrate how the HiPEAC community is delivering innovation with real impact. We spoke to some of the winners to find out more.

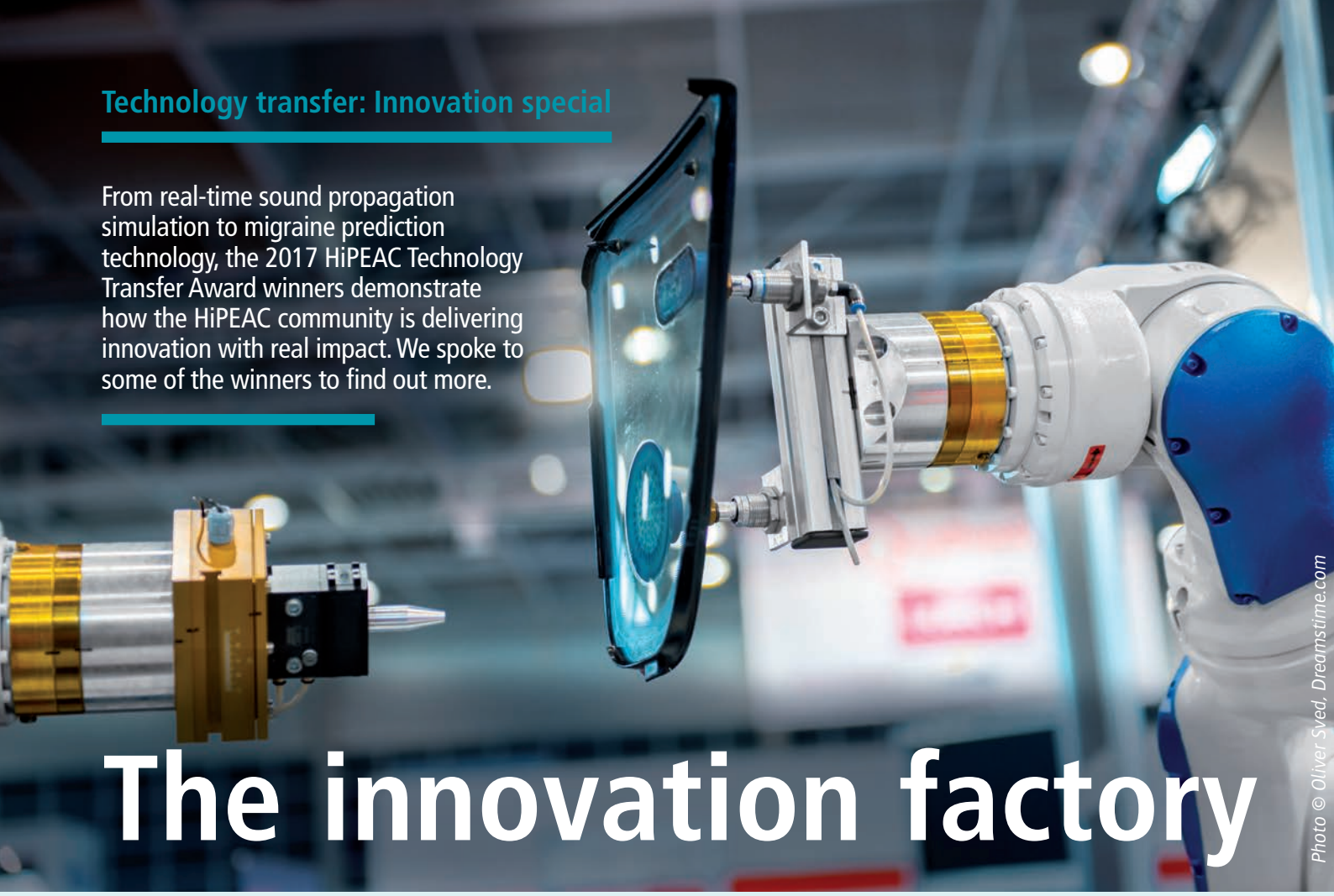


Photo © Oliver Sved, Dreamstime.com

The innovation factory

HiPEAC Technology Transfer Award winners 2017

VIRTUAL REALITY? SOUNDS LIKE THE REAL THING

RESEARCH CENTRE: AGH University of Science and Technology
COMPANY: Techmo

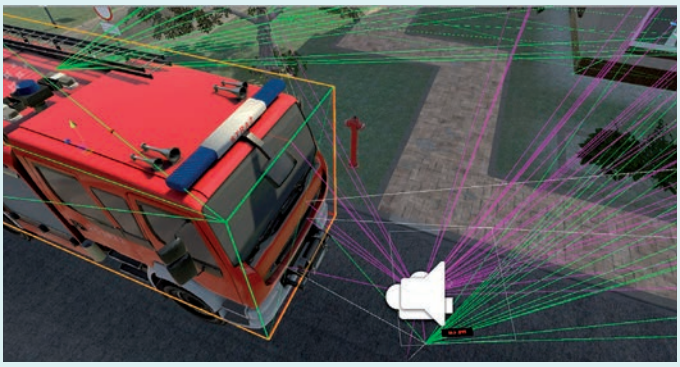
While interactive applications such as video games and virtual reality simulations offer incredible visual realism, the sound isn't always as impressive, something **Szymon Pałka** and his colleagues at AGH University of Science and Technology set out to change. 'There have been huge advances in graphics quality, but we felt this wasn't matched by audio quality,' he explains.

What started as a simple goal led to complex problems, however, requiring the development of advanced algorithms, according to Szymon. 'As a result, we created Sound ToolKit, which analyses the virtual environment (walls, obstacles, sound sources, etc.) to find out how the sound propagates in it. Next, our sound synthesis module uses the analysis data to render sound – which can be done in real time.'

Although the original intention was for SoundToolKit to be used in video games, 'more advanced hardware led to new use cases, where the software really shines,' says Szymon. 'The visual experience on virtual and augmented reality devices is so immersive that any discrepancy between what we think the real world should sound like and what we actually hear is even more noticeable. This is where our software comes in, helping content creators achieve a more coherent experience.'

The team built a working prototype at the university, but Szymon explains that it was only after the technology transfer to the Polish company Technmo (featured in *HiPEACinfo* 49) that they could transform the ideas into the full product: the SoundToolKit audio engine. 'The product has already been licensed for two productions, one of which is virtual reality training software for firefighters.'

Perhaps the toughest aspect of transferring the technology has been inculcating a culture change, according to Szymon. 'This technology requires a completely different approach to audio content creation to the most prevalent techniques used today. For this reason, we've continuously had to demonstrate how much improvement can be achieved by reorganizing the game creation process.'



SoundToolKit
Powered by TECHMO

As for the future, Szymon comments that development on performance and quality is a priority. 'The biggest challenge we're facing is to extend our solution so it will be more attractive for large game developer studies, who want greater control of the output audio – basically altering the simulation results as needed. We are also planning to create a system to augment artificial intelligence in interactive simulations by analysing which sounds in a particular situation reach the computer-controlled character.'

BRINGING HIGH-PERFORMANCE COMPUTING TO THE INDUSTRIAL IOT: OMPSS@FPGA

RESEARCH CENTRE: Barcelona Supercomputing Center (BSC) and Universitat Politècnica de Catalunya – Barcelona Tech (UPC)
COMPANY: Aingura IIoT

Aingura IIoT (formerly Plethora IIoT), part of the Etxe-Tar group of manufacturing companies, produces automotive parts. Their complex machines produce gigabytes of data per hour, which is collected and time-stamped with dedicated field-programmable gate array (FPGA) circuits. This data needs to be processed and filtered in real time, both to reduce it to a manageable size and to allow reactions on the fly.

However, this requires high computing performance which would normally be beyond the sensors' low-power Arm cores and Xilinx FPGAs, unless there was a way of exploiting these processors to the full. While code could be optimized for either the Arm cores or the FPGAs, synthesizing all the codes for both types of processor was a cumbersome, manual task.

Enter OmpSs@FPGA, an adapted version of BSC's flagship OmpSs technology. 'OmpSs is a simple way of allowing programs to be executed on several processors at the same time,' explains **Carlos Álvarez**, who accepted the HiPEAC Technology Transfer Award on behalf of a team of BSC-UPC researchers. 'It was originally designed for supercomputers, but now that even mobile and sensor computers have various processors, it is becoming increasingly relevant for embedded computing.'

This means that OmpSs can be used in the industrial internet of things (IoT) – or 'wearables for machines', as Carlos puts it. 'I was initially sceptical about the IoT. But I was immediately struck by the importance of the industrial IoT, from something as simple as a sensor detecting when a machine's temperature is rising too much and switching off the machine, to more intelligent applications using artificial intelligence (AI) and therefore needing more processors.'

BSC signed a collaboration agreement with Ikergune, the research and innovation wing of Etxe-Tar, to develop OmpSs for Aingura's industrial applications. Using OmpSs@FPGA allows the computation to be parallelized among all the free resources of the system from an initially sequential C program. This has allowed Aingura to introduce artificial intelligence algorithms into the sensors while keeping the cost of computational resources down. 'With a large number of transistors that you can regroup to execute computations on the fly, FPGAs offer the flexibility required for industrial applications at a lower cost,' says Carlos.

One of the main challenges faced when trying to transfer technology from the academic to industrial spheres, according to Carlos, is ensuring both that industry representatives are aware of the tools academic research can offer them and that academics are aware of real-life problems facing industry. Bringing academics into contact with industry is a good way to spot opportunities and to see how existing technology could be adapted to a different area.

'One of the use cases we developed during the EU-funded AXIOM (Agile, eXtensible, fast I/O Module for the cyber-physical era) project was facial recognition for smart home security applications. But the same software could easily be used for industrial IoT applications,' explains Carlos.

A LIGHTBULB MOMENT FOR SMART TRAFFIC MANAGEMENT

RESEARCH CENTRE: RWTH Aachen
COMPANY: ICE Gateway

When was the last time you were stuck in traffic, thinking about all the ways you'd rather be spending your time? 'Every driver is familiar with the frustration of time wasted in traffic jams,' says **Volker Lücken**, scientific staff member at RWTH Aachen. 'In addition, inefficient traffic flow leads to high emission levels, which has become an important political and societal topic in recent years – particularly in Germany.'

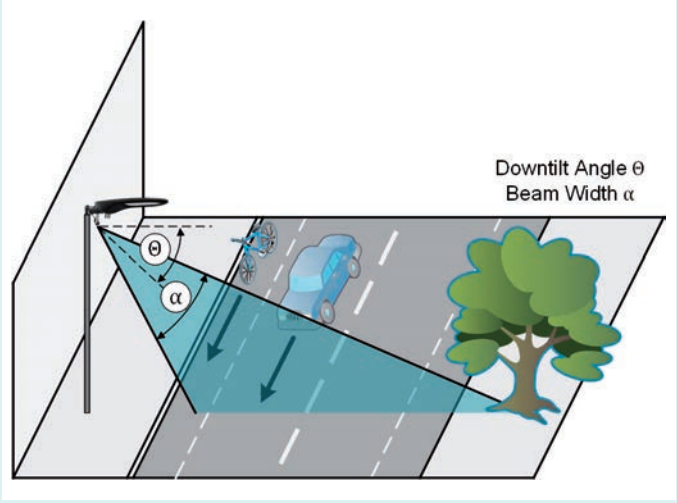
Traffic routing and congestion management can help. However, optimizing traffic flow requires real-time traffic monitoring, which can be both expensive and complex to integrate in the urban environment, Volker explains. Here, smart city approaches can provide innovation solutions. 'Our research group considers digital infrastructure as an essential building block towards connected, smart cities,' says Volker.

The group came up with a smart streetlamp as a response to the traffic management problem. 'Streetlamps cover all major urban areas, so if you add sensors, they can provide extensive coverage to monitor city traffic. They also have the added advantage of integrating seamlessly into the urban landscape,' he notes. 'With this in mind, we combined our research on sensor signal processing, statistical inference techniques and machine learning algorithms with our knowledge of embedded system concepts to develop a novel, ultrasonic sensing technology.'

The streetlamp includes processing, communication and actuation capabilities together with interfaces for external sensor components. A small, ultrasonic sensor allows multi-lane vehicle detection or parking space monitoring. 'This real-time data can be used by city authorities either to perform online vehicle routing or offline optimization of typical traffic patterns throughout the day,' says Volker.

Research-industry partnership

The initial technological concept was developed in strong cooperation with the group's industry partner, ICE Gateway – itself a spin-off of the university – which specializes in smart city lighting solutions. 'On the research side, we pursued this opportunity to intensify our investigations into statistical signal processing and object detection algorithms,' says Volker.



The smart streetlamp prototype

Once the initial prototypes were available, both partners were able to perform early evaluations and identify customer requirements. 'Our initial algorithmic approaches showed good results in simulations and lab evaluations. However, real-world scenarios threw up issues, meaning we had to go back to basics and adapt our concept based on this feedback. Facing these challenges at such an early stage meant we could overcome them in the following development iterations, which would not have been possible without the partnership with ICE Gateway,' Volker adds.

Once the technology had reached an initial level of maturity, the academic and industry partners successfully applied for publicly funded projects in the field of intelligent transportation and autonomous driving, allowing them to integrate their technology. Publishing papers and acquiring patents gave the research visibility and advanced the commercialization process. HiPEAC also played a part in supporting the development, as presentations at workshops gave the participants an external perspective on their concepts, algorithms and initial results.

So what does the future hold for this exciting new technology? 'The next step for the industry partner is large-scale deployment. On the research side, we will now focus on extracting more information about the traffic out of the signals and transferring the concepts to other applications of ultrasonic detection, for example in the security domain,' says Volker.

SAFE SYSTEMS FOR THE TRANSPORT OF THE FUTURE

RESEARCH CENTRE: Barcelona Supercomputing Center
COMPANY: Rapita Systems

With modern forms of transport being increasingly computerized, safe, reliable systems are paramount. Crucial here are critical real-time embedded systems like those that manage safety-related functions in cars (e.g. the auto-cruise system) and planes (e.g. the flight management system). 'Such systems require both functional and timing verification, with industry standards requiring guarantees that critical functions are performed within certain time limits,' explains **Francisco J. Cazorla**, leader of the Computer Architecture – Operating Systems (CAOS) group at Barcelona Supercomputing Center (BSC).

For single-core microcontrollers, advanced timing analysis tools, such Rapita's RapiTime, already provide all the evidence required for certification. However, using multicores to meet increasing performance requirements makes it more difficult to provide tight, reliable time limits for the execution of tasks, a fundamental requirement for certification. As Francisco observes, this is largely caused by the difficulty of predicting the worst-case interference of applications running on different cores when accessing shared hardware resources. 'In response, we've developed a multicore micro-benchmark technology known as MμBT, which has been successfully assessed with prototype research tools,' he says.

During the PROXIMA (Probabilistic real-time control of mixed-criticality multicore and manycore systems) project, which received funding from the European Commission, a partnership and framework agreement was created between BSC and Rapita under which BSC joined Rapita to provide consultancy services relating to multicore timing analysis. 'As part of the agreement, BSC and Rapita are integrating MμBT into Rapita's analysis tool RapiTime to offer industrial-quality worst-case execution time estimates for critical real-time embedded systems,' says Francisco. 'The integration of BSC's MμBT into Rapita's RapiTime is a step towards an industry-ready suite of verification tools for multicore processors.'

HiPEAC played a fundamental role in bringing the academic and industry partners together, says Francisco. 'Our relationship with Rapita, which is fundamental in bringing this technology into industrial domains, began as a result of HiPEAC; before that we had contact with researchers in the United States, but not so much in Europe.'

While steady progress has been made in increasing MμBT's technology readiness level (TRL) – thanks both to PROXIMA and to a European Regional Development Fund grant – Francisco notes that obtaining industry-ready tools is usually unrealistic within the scope of a three-year research project. 'Integration is complex due to the different types of multicore processor,' he comments. 'There is plenty more work to be done here, particularly as autonomous vehicles become more and more common.'

WEARABLE TAKES THE PAIN OUT OF MIGRAINES

BRAINGUARD

RESEARCH CENTRE: Complutense University of Madrid in collaboration with the Hospital de la Princesa
COMPANY: BrainGuard LLC

Anyone who has suffered from migraines will know how debilitating they can be, and they are surprisingly common. 'In different forms, migraines affect 10-15% of the world's population at least once a year,' says **José Ayala**, Associate Professor at the Complutense University of Madrid. 'They can last up to 72 hours, and no curative medication is yet available.' However, studies suggest that taking medication at the first sign of pain either eliminates or ameliorates pain in all cases. The problem, according to José, is that patients do not always recognize such signs.

In collaboration with neurologists and a neurobiologist from the Hospital de la Princesa (Madrid), José's team at the university's Department of Computer Architecture and Automation developed an algorithm capable of interpreting real-time changes in

Photo © Katarzyna Bialasiewicz | Dreamstime.com

physiological signs. 'This has been integrated into an unobtrusive wearable device which monitors various vital signs. Based on this, through machine learning, the algorithm can predict when a migraine will begin, and the device alerts the patient so that they can take pain relief in time,' says José.

The team quickly saw the commercial potential of the technology and took steps to protect the intellectual property through national patents and international extensions. Next, they created a business plan and began participating in acceleration and mentoring programmes. Thanks to this preparatory work, a start-up, BrainGuard LLC, was set up to manufacture and validate the technology. 'Our next step is to gain the investment necessary to manufacture the device commercially,' says José.

HiPEAC has a role to play in helping researchers develop entrepreneurial as well as research skills, and José is particularly appreciative of the opportunities for international collaboration the network offers; in 2011, he spent three months at the École polytechnique fédérale de Lausanne (EPFL) thanks to a HiPEAC Collaboration Grant. 'I've always considered it fundamental for researchers to spend time outside of their own country – not just when they're just starting out, but at regular intervals throughout their career,' he says.

'For my part, I've learned a lot and continue to learn from my international colleagues. Not just about their research topics, but how they manage research, how they transfer technology and the different paths which exist to help us internationalize our research and generate new research topics.'

2017 HIPEAC TECHNOLOGY TRANSFER AWARD WINNERS

Camilla Giunti (IngeniArs): IngeniArs S.r.l.: the result of technology transfer from research to business

Donatella Sciuto, on behalf of her PhD student **Rolando Brondolin**, (Politecnico di Milano): HyPPO - Hybrid Power-Capping and Performance-Aware Orchestration in the Cloud Computing Era

Francisco Cazorla (Barcelona Supercomputing Center): Enabling Real-Time Guarantees on Multicores with Rapita's Verification Suite and BSC's Micro-Benchmark Technology

Szymon Pałka (AGH University of Science and Technology): Realtime simulation of sound propagation in 3D environment

Aleksandar Ilic, (Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento): Cache-aware Roofline Model in Intel® Advisor

Gerd Ascheid, on behalf of his PhD student **Volker Lücken** (RWTH Aachen): A novel sidefire ultrasonic traffic sensing technology integrated into Smart City infrastructure

Carlos Álvarez (Universitat Politècnica de Catalunya - Barcelona Tech): OmpSs@FPGA for Industrial Internet of Things

Francisco Tirado, on behalf of his affiliate member **José Ayala** (Universidad Complutense de Madrid): BrainGuard, a brand-new start-up providing technology for migraine prediction

Martin Kersten (Centrum Wiskunde & Informatica): MonetDB, the column-store pioneers

Dimitris Gizopoulos (University of Athens): Framework for Full-System Hardware and Software Reliability Analysis at the Microarchitecture Level Detail

The next edition of the HiPEAC Vision, HiPEAC's biennial roadmap, will be launched at the 2019 HiPEAC conference in Valencia. Here, Editor-in-Chief Marc Duranton describes how the Vision draws on the HiPEAC community's expertise to predict the future, and gives us a sneak preview of key themes.

'Cyber-physical systems with AI and fog computing are an opportunity for Europe'

How do you go about creating the Vision?

It's a collective effort by the whole community. We sent a survey to HiPEAC members at the end of 2017 to get their input on what the key elements in our domain will be in 10 years, and we also get direct contributions by email. We've held consultation meetings, with more planned over the next few months; we've spoken to teachers and industry partners at the ACACES summer school in 2017 and will do so again in 2018; and we've exchanged views with organizations such as ARTEMIS-IA and ETP4HPC, the European Technology Platform for High-Performance Computing. And of course we actively monitor what's going on in scientific publications and the news. The editorial board then organizes all these inputs into a coherent and (hopefully) easy-to-read document.

Could you give us a sneak preview into some of the highlights?

It's very difficult to say, as the document is still taking shape. What I can say is that the key directions identified in previous editions are being confirmed, and their impact is increasing. Energy efficiency is still a major challenge, both for servers/high-performance computing (HPC) and for autonomous systems. As for societal issues, privacy and problems with trust in computing systems have been underlined by the recent events involving Facebook and Cambridge Analytica. Meanwhile, the arrival of the European Union's General Data Protection Regulation (GDPR) heralds the most important change in data regulation in 20 years.

Spectre and Meltdown exposed the vulnerabilities of microprocessors

What about security?

Security is still a major issue; we've seen more and more examples of malware and ransomware, which have even affected hospitals and shut down production lines, not to mention mining bitcoins using your computing power without your knowledge. Even our microprocessors have been shown to be vulnerable, thanks to the Meltdown, Spectre and BranchScope 'bugs'.

How can we make systems more trustworthy?

One root cause is the complexity of systems, both hardware and software, meaning that their specifications cannot encompass every aspect. Managing system complexity while increasing human productivity is an important direction. Another key aspect of trust is ensuring that systems do what they are supposed to do under all conditions.

In practice, coping with complexity will mean using more formal methods, analytical approaches, dynamic supervisors or artificial intelligence approaches.

Artificial intelligence is a buzzword you hear everywhere nowadays. Will it affect the HiPEAC community?

Definitely, although I don't like the term 'artificial intelligence', because these systems are far from being intelligent as most people understand the term. It would be more accurate to say that they mimic low-level cognitive processes. But yes, it will have an impact on our community.

First, we have to deliver the tools, hardware – accelerators – and software that improve the performance and energy efficiency of artificially intelligent applications. One example is the accelerators which have started to appear in mobile phones.

►

Artificial intelligence might also provide a solution for mastering system complexity; we should develop techniques inspired by these methods along with traditional approaches like operational research and formal methods to build better hardware and software.

This isn't a new theme, though – we've mentioned it in several editions of the Vision, for example the 2013 edition.

Can you summarize the main directions of the 2019 Vision? What does it mean for Europe?

As I mentioned, it's difficult to say as the document is still in flux, but it is clear that new applications and systems will leave cyberspace to interact increasingly with the real world. This is the era of cyber-physical systems, where computers will take the input from sensors and use actuators to control cars, factories and cities.

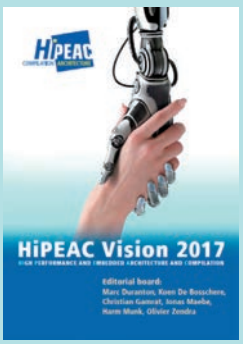
As natural data are harder to process – and the real world is less predictable – than numbers, systems will need to be more and more intelligent. They will achieve this thanks to artificial intelligence, reborn with a proposed new name of 'cognitive cybernetic and physical systems', or C²PS. However, to meet safety, privacy, efficiency and cost requirements, at least some of this intelligence will need to be near the sensors and actuators – that is, at the edge and not just located remotely in the cloud.

These new breeds of systems, combining cyber-physical systems, artificial intelligence (partly) located at the edge, and collaborating (sharing data, and possibly storage and computing resources, as in fog computing) are an opportunity for Europe to regain a strong position in computing, leveraging its knowhow in embedded systems, industry software and mathematics.

Is it too late to contribute to the 2019 HiPEAC Vision?

No – we are still holding meetings and discussions, so the main ideas might evolve from what we've discussed here. I would encourage anyone who wants to contribute their ideas to email us at vision@hipeac.net.

"I don´t like the term ´artificial intelligence´ – it´s more accurate to say these systems mimic low-level cognitive processes"



You can download the latest version of the HiPEAC Vision from the HiPEAC website hipeac.net/publications/vision

This edition features a software framework for better big data systems for the smart city domain, automatic optimization during software development, a mega project for model-driven engineering and an update on the Heterogeneity Alliance.

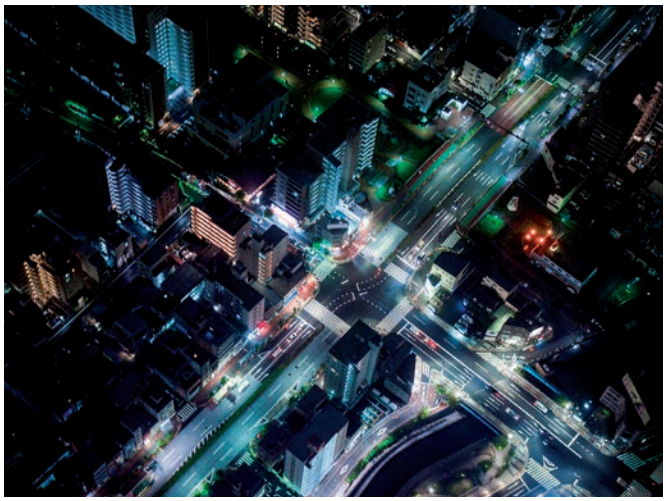
Innovation Europe

A CLASS ACT

Smart cities and autonomous vehicles are the near future. CLASS, coordinated by Barcelona Supercomputing Center (BSC), aims to develop a novel software architecture to help software developers to design and efficiently execute big-data analytics workloads with real-time constraints in distributed environments such as those deployed in smart cities.

To do so, CLASS aims to converge and evolve high-performance, low-power embedded and big data analytics computing technologies into a unified software architecture capable of efficiently coordinating and distributing computation resources along the compute continuum (from edge to cloud), while providing real-time guarantees.

The software architecture will be based on COMPSs, the software framework developed at BSC to design and execute high-performance applications in distributed cloud environments. COMPSs will be enhanced to support the distribution of computation among both cloud and edge computing resources, while providing real-time guarantees to the overall work-flow execution, as required by automotive systems.



This technology will be tested in a real urban area of one square kilometre in the city of Modena, using three prototype vehicles provided by Maserati and equipped with the necessary sensors and connectivity.

The project will pave the way towards better big-data systems for the smart city domain that will improve sustainability, services and safe mobility. In addition, the project will prepare the technological background for the advent of trustworthy connected vehicles.

'CLASS will develop a novel software framework for a new generation of highly distributed computing systems with data analytics and real-time requirements, capable of coordinating computing resources along the compute continuum. HiPEAC members are at the heart of this innovative project, whose technology will help bring about the smart cities of tomorrow,' says Eduardo Quiñones, CLASS project coordinator and BSC Computer Sciences researcher.

NAME: CLASS: Edge and Cloud Computation: a Highly Distributed Software for Big Data Analytics
START/END DATE: 01/01/2018 – 31/12/2020
KEY THEMES: Smart cities, connected vehicles, data analytics, big data
PARTNERS: **Spain**: Barcelona Supercomputing Center (BSC); **France**: Atos; **Italy**: Università degli studi di Modena e Reggio Emilia, Comune di Modena, Maserati; **Israel**: IBM Research – Haifa
BUDGET: €3.9 million
WEBSITE: class-project.eu

CLASS has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 780622

OPTIMIZE YOUR SOFTWARE AS YOU DEVELOP

SDK4ED

Software Development ToolKit for Energy Optimization and Technical Debt Elimination

Imagine if your software were automatically optimized for certain quality requirements, such as energy efficiency, dependability and performance, while avoiding technical debt, where choosing an easy solution leads to additional rework along the line. The European Commission-funded project SDK4ED – a Software Development ToolKit for Energy Optimization and Technical Debt Elimination – will provide tools to do exactly that.

The project is developing methods and tools to parse software artefacts, (source code, design models, test cases, etc.) and analyse them from the perspective of technical debt liability, considering the targeted hardware platform and the quality requirements provided. The tools will provide reports highlighting deficiencies, ranked by importance, taking the change log and probability of future maintenance into account. Tools monitoring energy consumption and identifying security vulnerabilities will also contribute to the optimization process.

SDK4ED will estimate the cost and the limitations associated with technical debt liabilities related to the energy-aware software development lifecycle. It will also establish a set of forecasting methods and best practices to assess options for repaying technical debt, appropriate timing and items to be improved under the energy consumption and dependability constraints imposed. Considering the trade-off between software quality and runtime constraints on energy consumption and security, the SDK4ED platform aims at deriving Pareto fronts that will assist developers during system optimization.

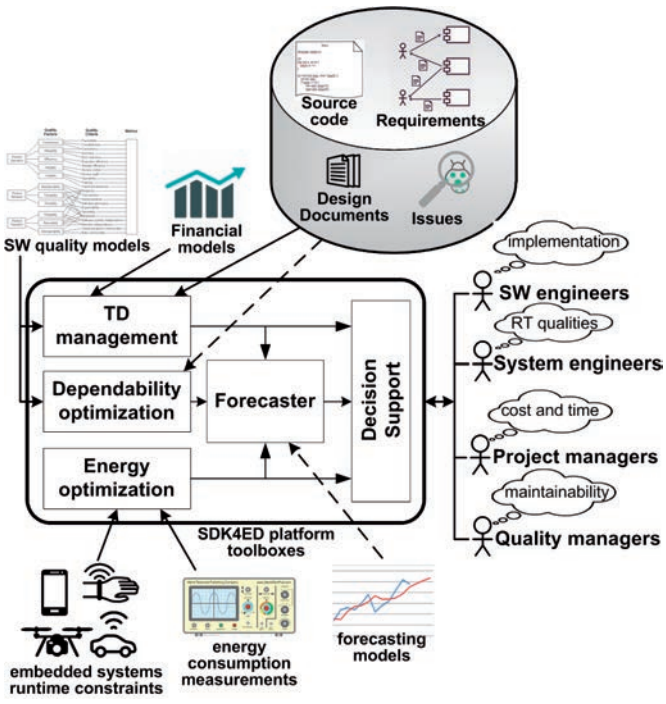
"SDK4ED will provide tools to automatically optimize software while avoiding technical debt"

Three representative use cases from the aerial vehicle, healthcare and automotive domains will be deployed in real operational environments to assess the impact of the project's innovations, and showcase the operations in general use context.

To find out more, contact Dionysis Kehagias (project manager). Email: diok@iti.gr

NAME: Software Development Toolkit for Energy Optimization and Technical Debt Elimination (SDK4ED)
START/END DATE: 1 Jan 2018 – 31 Dec 2020
KEYWORDS: Software engineering, embedded computing, energy efficiency
PARTNERS: **Greece**: Centre for Research and Technology Hellas (CERTH), University of Macedonia, Institute of Communications and Computer Systems (ICCS); **Netherlands**: Rijksuniversiteit Groningen, Neurasmus BV, TIOBE Software BV; **UK**: Imperial College of Science Technology and Medicine, Maxeler Technologies Limited; **Sweden**: CNET Svenska AB; **France**: Airbus DS; **Romania**: SC HoliSun SRL
BUDGET: €4.33M
WEBSITE: sdk4ed.eu

The SDK4ED project has received funding from the European Union's Horizon 2020 Programme under grant agreement no. 780572.



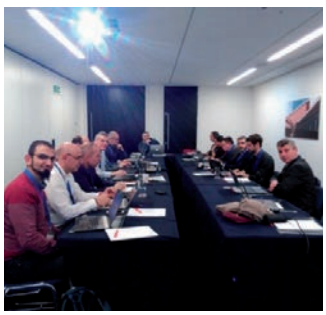
THE HETEROGENEITY ALLIANCE IN HANDS-ON MODE

Clara Pezuela, Atos, TANGO project coordinator

An update on the initiative

It's now been a year since the TANGO (Transparent heterogeneous hardware Architecture deployment for eNergy Gain in Operation) project launched the Heterogeneity Alliance. The Alliance is growing; currently, there are 18 members (nine research projects, two universities and seven industrial organizations). A first version of the reference architecture is now available, while a catalogue of tools is being populated with functional blocks for the different layers of the proposed architecture. The Alliance is also contributing to the next HiPEAC Vision document, as well as contributing to a book about the heterogeneity challenge in Europe and other activities.

Members organized a joint open workshop at the HiPEAC conference in Manchester and had the opportunity to share challenges and solutions in the area of heterogeneity; see 'Further reading', below, for presentations. The workshop highlighted different approaches to address the topic, which have been combined in the architecture document. A first board meeting of the Alliance members was also held at the event to discuss objectives, activities, governance and the future roadmap. Currently the Alliance is a HiPEAC stakeholder member, represented by the TANGO project coordinator, and participates actively in events and publications promoted by HiPEAC.



A reminder about the mission

Heterogeneous architectures have received considerable attention as an efficient approach to run applications and deliver services by combining different processor types in one system to improve absolute performance, minimize power consumption and/or lower costs. The impact of this heterogeneity on computing tasks is rapidly increasing, and consequently needs consideration.

The Alliance focuses on all phases of heterogeneous software, from the design phase to enhanced execution, parallel programming and optimized runtime, and considers a number of factors such as energy, performance, real-time, data locality and security. This will enable a new way of developing and executing next-generation applications.

As reported in previous issues of *HiPEACinfo*, the main aim of the Alliance is to create an open community (non-profit and non-legal) in which anyone interested in technological areas related to heterogeneity can participate. The community seeks to found a common, open and extendable set of technologies and tools around software development for heterogeneous hardware, which are attractive, easy to use and broader in scope and value, making them viable for mass adoption.

Benefits for members

The main benefits of the Heterogeneity Alliance for members are as follows:

- Take advantage of an additional channel to disseminate results (through the catalogue of tools).
- Include priorities and challenges in the reference architecture document.
- Collaborate with related research projects by conducting joint activities (workshops, books, publications, etc).
- Join forces with other projects and organizations in proposing roadmapping topics for the HiPEAC Vision and European Commission work programmes.

In exchange, the Alliance members commit to promoting the Alliance's goals and activities and to collaborating as far as possible in the joint activities.

If you want to influence the heterogeneity market, engage with potential competitors, partners and customers, or benefit from state-of-the-art catalogue tools, you should become a member of the Alliance. To join, visit our website: heterogeneityalliance.eu

FURTHER READING:

heterogeneityalliance.eu/alliance-members
bit.ly/HA_preliminary_reference_architecture
heterogeneityalliance.eu/catalogue
bit.ly/HA_HiPEAC18_workshop_presentations

TANGO is funded by the European Commission under the Horizon 2020 Framework Programme for Research and Innovation under grant agreement no. 687584.

MEGA MODELS

Agile development and validation for complex systems

To help Europe compete in an increasingly competitive electronics field, MegaM@Rt2 is developing toolsets to modernize the development and validation of complex systems. Here, project coordinator Gunnar Widforss (Mälardalen University) discusses the industry-ready model-driven engineering solutions being created by the project, which draws on the expertise of up to 100 participants.

How can model-driven engineering technologies benefit European industry?

Complex systems are now prevalent in several domains of cyber-physical systems, including the automotive and aerospace sectors, healthcare, industrial control and automation. Modern engineering practices are needed to ensure advances in productivity, quality and safety. At the same time, Europe's traditionally strong position in the embedded systems market (30%) is under threat as other geographies expand in this area.

ECSEL-JU, the European Commission's Joint Undertaking on Electronic Components and Systems for European Leadership, seeks to remedy this by investing in projects that strengthen industrial competitiveness, enable economic growth and improve sustainability, such as MegaM@Rt2. Model-driven engineering technologies can give Europe a competitive edge in software development by reducing development costs and time-to-market while maintaining high levels of quality and safety.

As cyber-physical real-time systems evolve, having the agility to react rapidly to hardware changes or usage scenarios is a critical success factor for businesses. Moreover, mastering systems that are ever more complex allows new use cases to emerge, based on optimization of larger problems or optimized solutions of existing problems.

How does model-driven engineering improve productivity?

This technology allows you to 'preview' new systems during the design phase, validate the design and make improvements, as opposed to testing later and finding that a lot of effort had been in vain.

With further development, model-driven engineering solutions would also allow you to monitor and manage a system at runtime, as well as integrating design and runtime engineering practices more efficiently. For example, it would allow comparison of the system at runtime with the original design to identify flaws, providing feedback to engineers as part of an agile software lifecycle. It would also provide a holistic, dynamic view of different aspects of the system while still on the design board, in order to validate and verify the design before implementation and deployment.

What is MegaM@Rt2 doing in this area?

The MegaM@Rt2 Framework comprises toolsets to help scale up model-driven technologies for real-life industrial projects. The toolsets are as follows:

- **Holistic Systems Engineering**, integrating existing industrial practices, verification and validation at system level.
- **Runtime Analysis**, conducted with monitoring, online testing and verification, as well as models@runtime techniques.
- **Model and Traceability Management**, which relates design models and runtime models, in addition to allowing mapping between models produced within the design phase or within the runtime phase.

Thanks to these toolsets, feedback from the runtime to design level will improve the design, while information collected during the design phase will improve the traceability and quality of the end product.

What use cases are you working on?

One important use case relates to safety-critical systems, such as flight control systems or door control systems for trains, which are expensive to develop due to the extensive verification and validation process required for certification. Due to this cost, once certified, engineers rarely change the design, and it is harder for them to reuse all or part of the designs for new solutions. Verification and validation at the design stage,

Photo credit: Peter-Paul Moschik on Unsplash

coupled with rapid prototyping and simulation using an agile methodology, would drastically reduce costs and promote the development of new systems.

Another use case involves analysing the system's execution logs, allowing engineers to understand the system's performance from execution logs and identify inconsistencies or problems with respect to the original design.

What are the challenges involved?

The first is organizational: MegaM@Rt2 has brought together over a hundred people from 27 organizations in six countries in order to develop its industry-ready framework. On the one hand are technology providers from academia and industry who provide components of differing technology readiness levels. On the other hand are large companies and small/medium enterprises developing cyber-physical systems for different sectors including avionics, rail, logistics, telecommunications, automotive and traffic management.

Technically, the main challenge is to bring together design and runtime engineering practices. While model-based technologies are popular at the design phase, their penetration at the runtime analysis phase is limited, so one challenge is to make toolsets at runtime 'speak' the common language of models. Another technical challenge is to make all toolsets interoperate to provide a seamless experience for engineers.

What are the project's main achievements so far?

Following analysis, we've identified different types of requirements for case studies to be supported by the MegaM@Rt2 framework: conceptual (use case) requirements, framework requirements and toolset (technology) requirements. These have been mapped to a preliminary high-level vision of the MegaM@Rt2 framework, and a first version of the framework architecture roadmap – which will prioritize, synchronize and guide the toolset development work – has been proposed.

What are the next steps?

The next important milestone is to carry out experiments where use case providers will test out the baseline technology and toolsets. These will help define the systems engineering scenarios to validate the framework, as well as providing baseline metrics that will need to be improved by subsequent versions of the framework.

This year, we are organizing a workshop on Model-Driven Engineering for Design-Runtime Interaction in Complex Systems (MDE@DeRun 2018), co-located with STAF 2018 (Software Technologies: Applications and Foundations) in Toulouse on 28 June. This will provide a great opportunity to discuss the kind of problems and solutions we are interested in with experts.

MegaM@Rt2 is funded by the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement no. 737494. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and Sweden, France, Spain, Italy, Finland, Czech Republic.

FPGA-accelerated cloud computing

Cathal McCabe, Xilinx University Program Manager for Europe, the Middle East and Africa, sets out the benefits of cloud computing with field-programmable gate arrays (FPGAs), which is now available as a service.

With rising computational demands and a focus on power efficiency, heterogeneous architectures are increasingly being adopted for next-generation cloud and high-performance computing applications.

For many years a technology for specialized applications, FPGAs entered the mainstream with the announcement from Amazon of F1 instances in the Amazon Web Services (AWS) EC2 cloud. The new F1 instances are compute nodes enabled with high-end FPGAs that can be programmed by the end user. The announcement heralded the arrival of FPGAs-as-a-Service (FaaS), a new cloud-computing model that re-defines the high-performance computing landscape. Developers now have universal access to leading-edge acceleration technology that can scale from occasional applications with limited resource and runtime requirements, to the most demanding jobs, requiring huge arrays of accelerators working in parallel, around the clock.

Representatives from Edico Genome, the Children's Hospital of Philadelphia and Amazon Web Services celebrate breaking the record for genome sequencing

Edico Genome was one of the first companies to take advantage of this technology, launching their genome sequencing on F1 – an ideal application for FPGA acceleration. Healthcare professionals around the world can now access genome sequencing on demand, without needing to acquire or maintain local hardware installations or to program these devices themselves.

In partnership with the Children's Hospital of Philadelphia, Edico shattered the Guinness World Record for genome sequencing by using 1,000 F1 instances to process 1,000 paediatric genomes in parallel in under two and a half hours. The previous record was 26 hours for a single genome in 2016. This was a dramatic demonstration of the power of scaling with AWS.

F1 configurations

The F1 instances are available in two configurations. The f1.2xlarge instance has a Xilinx All Programmable Virtex Ultrascale+ with 64 GB connected directly to the FPGA. The f1.16xlarge instance has eight Virtex Ultrascale+ devices each with 64 GB of DDR4 memory, giving 512 GB total memory available to the FPGAs. The eight devices can either be used as independent accelerators or can be connected to each other via a high-performance 400 Gbps bidirectional ring.

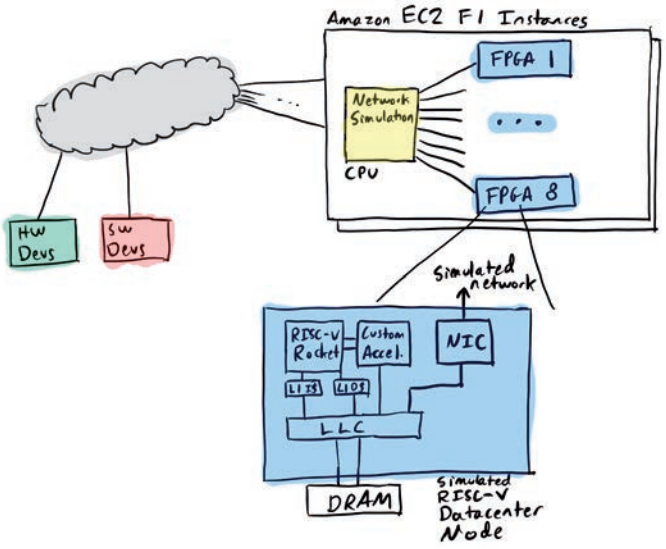
The Xilinx software to program these instances is also available in the AWS EC2 cloud. This includes the Xilinx SDAccel tools which software programmers can use to build designs for F1 instances using OpenCL. More experienced hardware developers have the option of building kernels using hardware description languages or high-level synthesis tools. Full debug and verification can be carried out on designs running on a remote instance. Applications running on an F1 instance can also be integrated with over 100 other AWS cloud services, providing developers with an ecosystem across the spectrum of high-performance computing.

FPGAs-as-a-Service offers a series of benefits both for industry and academia, as described below.

Academia

Educators can provide access to leading-edge customizable hardware in their classes, ranging from a small numbers of projects to classroom teaching where each student has access to their own F1 instance. Hardware and software updates are managed automatically by AWS.

The AWS platform is ideal for collaborative research and dissemination of results. A prototype or final system can be shared with anyone around the world instantly, and researchers can be sure that they are working on an identical infrastructure to their remote colleagues. All partners can access as many instances of a prototype for as much time as they require – no more waiting for boards to be delivered, time sharing of hardware, or travelling to the location with the only working system.



FireSim uses F1 as a HW/SW co-design environment and simulation platform (Image credit: Sagar Karandikar)

As an example, the FireSim project at UC Berkeley uses F1 as a hardware/software co-design environment and simulation platform for the open-source warehouse scale computing research based on the RocketChip RISC-V. Hardware teams in one or more locations can make modifications to a prototype that can be made available for software developers around the world to test and simulate.

Hardware developers can choose how much of the lower-level hardware detail to expose to the end user – if any. A top-level software API can be provided which treats the FPGA as a black box, with no specific domain knowledge of FPGAs required to take advantage of acceleration.

FireSim lowers the barrier to entry, making it easier for everyone to leverage the power of the FPGA-accelerated compute environment for RocketChip development research in this area.

Industry

A number of companies in Europe have already launched based on the FPGAs-as-a-service model.

Titan-IC, a spinout from Queen's University in Belfast, provides hardware-accelerated regular expression processing engines for security applications. The service can also be integrated with the full range of additional AWS services.

reconfigure.io provides tools to program FPGAs in the GO language which can then be deployed in the cloud. This simplifies the process of programming FPGAs by offering software developers a more familiar design language.

Xelera, a spinout from TU Darmstadt, develops novel datacentre processors on FPGAs that can operate much faster and with greater energy efficiency than traditional processor technology.

InAccel, a start-up from Greece, provides hardware accelerators that can be deployed to the Amazon AWS cloud. Applications can be integrated to widely used frameworks like Apache Spark. Machine learning, financial algorithms and computationally intensive algorithms including compression, encryption and hashing are available as accelerators.

XUP Jumpstart Program

The Xilinx University Program invites academics with software applications that would benefit from cloud acceleration to join the XUP AWS Jumpstart Program, which seeks to match hardware experts with software developers and domain experts who want to collaborate on high performance computing projects using F1 instances. For more information see xilinx.com/university or email Cathal.McCabe@xilinx.com.

XILINX
| UNIVERSITY PROGRAM

Digitalization allows for IT systems to be organized extremely efficiently. However, without a smart security structure, they also open the door to all kinds of unauthorized access. Against this background, a new approach, which the newly founded HENSOLDT Cyber is committed to, aims at harnessing the full potential of digitalization without giving infiltrators access.

Sensor specialist HENSOLDT offers protection in cyberspace

COMPANY: HENSOLDT Cyber GmbH
MAIN BUSINESS: Cybersecurity
LOCATION: Munich, Germany
WEBSITE: www.hensoldt-cyber.com

HENSOLDT, a sensor solutions provider active in defence electronics and flight safety, has joined with cyber specialists Secure Elements GmbH to found HENSOLDT Cyber GmBH. The new enterprise features the participation of HiPEAC17 keynote speaker Dr. Sandro Gaycken (cyber protection), HiPEAC steering committee member Prof. Dr. Rainer Leupers (chip design) and ACACES 2017 lecturer Prof. Dr. Gernot Heiser (micro kernel technology). It will develop security-hardened basic IT systems which are virtually impervious to hacker attacks and integrated hardware weaknesses.

'HENSOLDT's partner Secure Elements has a lot of experience in the cybersecurity world,' explains Marian Rachow, CEO of HENSOLDT Cyber. 'Together we will develop highly secure solutions which will protect IT systems against unauthorized access even at the operating levels of a computer. This approach avoids the shortcomings of conventional cyber protection, which only kicks in at higher application levels and can therefore be easily circumvented. To develop this we are still looking for more people to work with us in Munich.'

Cybersecurity can only be achieved systematically and on a highly secure operating system. Experience clearly shows that developing secure applications on top of insecure operating systems and untrustworthy hardware does not constitute a reliable security concept. If the lower levels – the hardware and operating system – are left open to attack, an attack is virtually guaranteed. Indeed, the lower levels that are particularly important, since they are part of the so-called TCB, the Trusted Computing Base.

An attack on the operating levels of a computer can manipulate or even sabotage every security measure at a higher level. Even the best encryption can be circumvented in this way, for example by means of an attack on the host system, and any hacker who wishes to do so can gain access to these systems via the low IT levels. The only way to avoid such problems is to use extremely secure, high-performance and robust operating systems, such as the seL4 microkernel. Rather than carrying on with 'business as usual', HENSOLDT Cyber will use these approaches to ensure cybersecurity from the hardware up.

Sandro Gaycken's keynote talk 'Cybersecurity: An unsolvable problem in the way of our IT-futures' is available to view on the HiPEAC YouTube channel bit.ly/Sandro_Gaycken_HiPEAC17

Sandro Gaycken, Rainer Leupers and Gernot Heiser are on the HENSOLDT Cyber GmBH team

Tensor Comprehensions: Just-in-time compilation for high-performance machine learning codes

Nicolas Vasilache (Facebook AI Research-FAIR), Oleksandr Zinenko (Inria and Département d'Informatique, École normale supérieure-DI ENS), Theodoros Theodoridis (ETH Zürich), Priya Goyal (FAIR), Zachary DeVito (FAIR), William S. Moses (Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory-MIT CSAIL), Sven Verdoolaege (FAIR), Andrew Adams (FAIR), Albert Cohen (FAIR, Inria and DI ENS)

Facebook AI Research (FAIR) has released Tensor Comprehensions. This C++ library and mathematical language helps bridge the gap between researchers, who communicate in terms of mathematical operations, and engineers focusing on the practical needs of running large-scale models on various hardware backends. The main differentiating feature of Tensor Comprehensions is that it represents a unique take on just-in-time compilation to produce the high-performance codes that the machine learning community needs, automatically and on-demand.

Order of magnitude productivity gains

The typical workflow for creating new high-performance machine learning (ML) layers can span days or weeks of engineering work through a two phase process:

1. A researcher writes a new layer at a numpy-level abstraction, chaining existing operations in a deep learning library like PyTorch, and tests it in small-scale experiments. The performance of the code implementing the validated idea needs to be accelerated by an order of magnitude to run large-scale experiments.
2. An engineer takes the layer and writes efficient code for graphics processing units (GPUs) and central processing units (CPUs). This engineer needs to:
   a. be a high-performance computing expert, of which only a limited supply of talent is available
   b. acquire context, map out a strategy, write and debug code
   c. undertake mundane tasks, such as verbose argument checking and adding boilerplate integration code, when moving the code to the backend

As a consequence, the deep learning community has grown to rely on high-performance libraries such as CuBLAS, MKL, and CuDNN to get high-performance code on GPUs and CPUs. Experimenting with ideas that deviate from the primitives provided in these libraries involves a level and magnitude of engineering that can be intimidating to researchers.

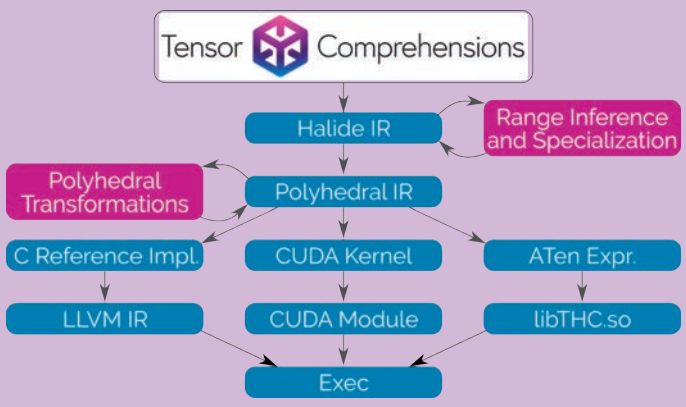
We anticipate great practical value in open-sourcing a package that shortens this process from days or weeks to minutes. With Tensor Comprehensions, our vision is for researchers to write their idea out in mathematical notation. This notation automatically gets compiled and tuned by our system, and the result is specialized code with good performance.

In this release, we provide:

- a mathematical notation to express a broad family of ML ideas in a simple syntax
- a C++ frontend for this mathematical notation based on Halide IR
- a polyhedral just-in-time (JIT) compiler based on Integer Set Library (ISL)
- a multi-threaded, multi-GPU autotuner based on evolutionary search

Related earlier work

A recent language that has become popular in the adjacent field of high-performance image processing is Halide. Halide uses similar high-level functional syntax to describe an image processing pipeline, and then, in a separate block of code, explicitly schedules it onto the hardware, specifying in detail how operations are tiled, vectorized, parallelized, and fused. This makes it a very productive language for people with architectural expertise, but it is difficult to use for most ML practitioners. Automatic scheduling of Halide is an active research area, but there is no good solution yet for ML code running on a GPU. ▶



Tensor Comprehensions uses the Halide compiler as a library. We build on Halide's intermediate representation (IR) and analysis tools, and pair it with polyhedral compilation techniques, so that you can write layers using similar high-level syntax but without the need to explicitly say how it is going to run. We also found ways to make our language even more concise, eliminating the need to specify loop bounds for reductions.

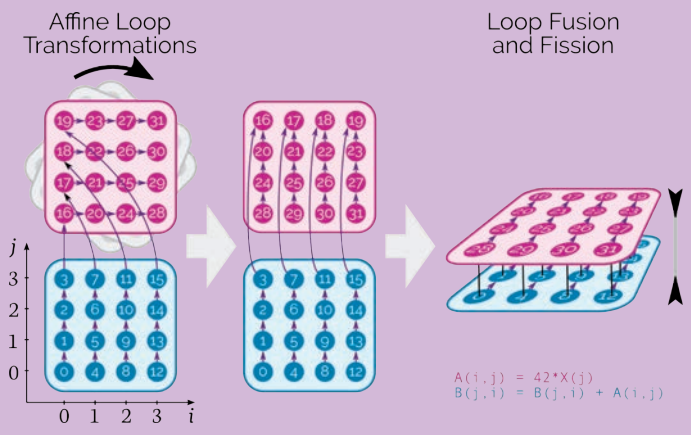
Automatically synthesizing (efficient) GPU kernels

Tensor Comprehensions use Halide and polyhedral compilation techniques to automatically synthesize CUDA kernels with delegated memory management and synchronization. This translation performs optimizations for general operator fusion, fast local memory, fast reductions and JIT specialization for specific sizes. Since we do not try to own or optimize memory management, our flow is easily and efficiently integrated into any ML framework and any language that allows calling C++ functions.

Contrary to classical compiler technology and library approaches, polyhedral compilation allows Tensor Comprehensions to schedule computations of individual tensor elements on demand for each new network.

At the CUDA level, it combines affine loop transformations, fusion/fission and automatic parallelization while ensuring data is correctly moved through the memory hierarchy. The numbers in the figure below show the order in which tensor elements were initially computed and arrows represent dependencies between them. In this example, the figure rotation corresponds to loop interchange which enables deep operator fusion.

To drive the search procedure, we also provide an integrated multi-threaded, multi-GPU autotuning library which uses evolutionary search to generate and evaluate thousands of implementation alternatives and select the best performing ones. You can call the tune function on your Tensor Comprehension and watch the performance improve, live. The best strategy is serialized via protocol buffers and reusable immediately or in offline scenarios.



On the performance side, while we still have many improvements in the works, Tensor Comprehensions can already match or beat the performance of current ML frameworks integrated with hand-tuned libraries, in favourable cases. This is mainly achieved by the ability to adapt code generation strategies to specific problem sizes. We are constantly conducting performance evaluations of the kernels produced automatically by Tensor Comprehensions. The early results demonstrate strong improvements on a variety of neural network models, compared to the default usage of vendor libraries such as CuDNN in the Caffe2 and PyTorch frameworks.

As we extend our contribution to more hardware backends, Tensor Comprehensions will complement fast libraries written by hardware manufacturers such as NVIDIA and Intel, and will be used in conjunction with libraries such as CUDNN, MKL or NNPack.

What to expect next

This release will allow researchers and programmers to write layers in a notation that is similar to the maths they use in their papers and communicate concisely the intent of their program. They will also be able to take that notation and translate it easily into a fast implementation in a matter of minutes rather than days. As the toolchain grows, we expect usability and performance to increase and benefit the whole community.

Tensor Comprehensions is integrated with the popular PyTorch and Caffe2 machine learning frameworks. We welcome feedback from other frameworks and teams. Email: tensorcomp@fb.com

Facebook has sponsored the HiPEAC conference in the past, and this work builds on earlier work by a long-term industry-academia collaboration called Polly Labs, supported by Arm, which won a HiPEAC Technology Transfer Award in 2015.

Access Tensor Comprehensions via github: github.com/facebookresearch/TensorComprehensions
The full version of this article originally appeared on the Facebook Research blog:
bit.ly/Facebook_Research_Tensor_Comp

Further reading:
Vasilache, et al. 'Tensor Comprehensions: Framework-Agnostic High-Performance Machine Learning Abstractions.' Ithaca, New York: Cornell University Library, 2018.
arxiv.org/abs/1802.04730

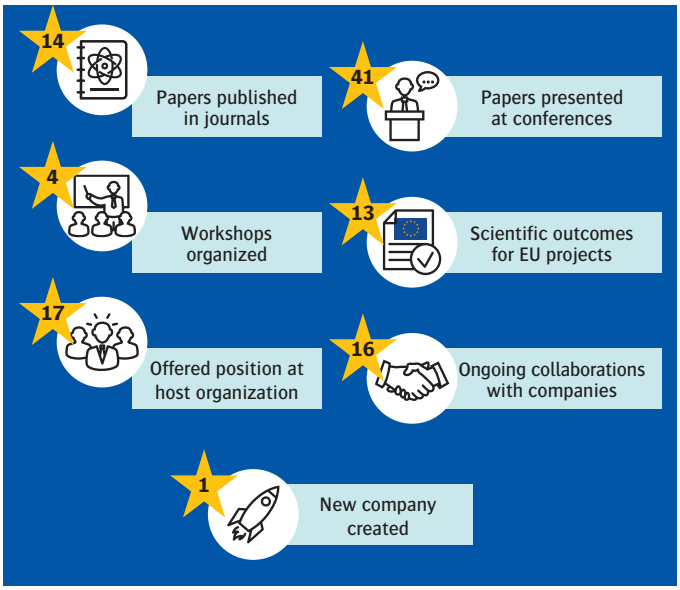
Over the years, HiPEAC has funded hundreds of mobility placements, helping researchers gain invaluable work experience. Xavier Salazar (Barcelona Supercomputing Center) sets out some of the opportunities on offer.

Boost your career with a mobility programme

Every year, many PhD students look for internship mobility opportunities, particularly over the summer break. Spending a few months at a different organization is fundamental for researchers' career development, especially when the mobility is between sectors (university/research centre - company). Students get the opportunity to widen their perspective on the work they usually do, perhaps having the opportunity to work on the practical side of their research or, conversely, getting a wider or a more in-depth view in a different environment.

HiPEAC mobility has become one of most successful instruments for structuring and connecting our community over the years. Since 2006, more than 300 PhD students have taken part in HiPEAC collaboration grants and internships. Today, some of the students who received a mobility grant while they were doing a PhD ten years ago are now successful senior researchers: research group leaders, project coordinators of major European Union project consortia, ERC Starting or Advanced Grant Holders. Others hold leading engineering positions at some of the most well-known companies in the field, such as Arm, Intel, etc.

To find out the impact of HiPEAC mobility programmes both on researchers' careers and on the organizations involved, we sent out a survey. This was answered by around 90 recipients of HiPEAC mobility grants. You can see a brief summary of the main findings below.



Missed out on this year's HiPEAC industrial internship mobility programme? Have a look at these other mobility programmes:

Digital Opportunity Traineeships: The Digital Opportunity traineeship initiative will provide cross-border traineeships for up to 6,000 students and recent graduates between 2018 and 2020. *Further information: bit.ly/Digital_Opportunity*

HPC Europa 3 – Transnational Access: Researchers visit high-performance computing (HPC) centres and/or scientific hosts who mentor them scientifically and technically to help them get the most out of HPC resources in their research. *Further information: hpc-europa.eu*

Eurolab-4-HPC cross-site mobility is a programme for short, interdisciplinary stays, targeted at PhD students, post-doctoral researchers and academics, working on multiple layers of the HPC system stack (architecture, runtime, tools, …). *Further information: eurolab4hpc.eu/collaboration*

PRACE summer of HPC offers summer placements at top high-performance computing (HPC) centres across Europe to late-stage undergraduates and early-stage postgraduate students. *Further information: summerofhpc.prace-ri.eu*

HiPEACjobs portal: Don't forget that internships and studentships are posted throughout the year on the HiPEACjobs portal. The table below shows trends on the HiPEAC Jobs portal for the last quarter.

Number of Jobs posted in the last quarter by institution	Jobs from relevant institutions, sponsors, contributors & projects
61 Jobs @ Research Centres	
84 Jobs @ Industry	
31 Jobs @ SMEs	
72 Jobs @ Universities	
47 Jobs @ H2020 Projects	
51 Internship Positions	

Interested in finding out more about HiPEAC mobility? Contact us at recruitment@hipeac.net
Further information: hipeac.net/jobs.

Arm Education Media: Addressing the engineering skills gap

Khaled Benkrid, Senior Director of Education and Research, Arm Ltd, explains how Arm's teaching materials help students and teachers keep up with the latest Arm technologies.

The skills gap

One of the most discussed topics in education policy circles these days is how to narrow the gap between what educational institutions are teaching, and the knowledge and skills required in today's job market. Often referred to as the 'skills gap', there is evidence that this gap is widening in engineering disciplines, with considerable socio-economic consequences.

As an academic now working in industry, it is clear to me that addressing this gap necessitates collaboration among four main stakeholders: government, education providers (including schools and universities), industry and learned societies. Industry stakeholders have a leading role to play, not only in facilitating educational institutions' access to the latest technologies, but also as content creators. The unprecedented pace of technological change in the last few decades makes it difficult, if not impossible, for educators to keep their teaching materials up-to-date, especially given the increasing pressure on educators to balance teaching with research. As the stakeholder responsible for the execution of technology roadmaps, industry is best positioned to produce content that matches this pace.

Arm Education Media

We at Arm have recently launched Arm Education Media, a subscription-based digital content hub offering interactive online courses and textbooks. This will allow academics, students, professional engineers and the wider training market to keep up with the latest technologies from the Arm ecosystem.

Arm Education Media's online courses combine theoretical and practical materials in the form of lecture slides and videos, interactive quizzes, and engaging lab videos demonstrating state-of-the-art software and hardware technologies. They are ideal for 'flipped classroom' pedagogy or individual self-study. The first eight online courses are on the following topics:

- Efficient and rapid embedded systems design
- The internet of things
- Digital signal processing
- Real-time operating systems
- Graphics and mobile gaming
- System-on-chip design

Our first textbook *Embedded Systems Fundamentals on Arm Cortex-M based Microcontrollers, A Practical Approach*, by Alex Dean (NCSU), was published in May 2017 and has been particularly well-received. More courses and textbooks will follow soon.

To ensure they maintain industry relevance, all of Arm Education Media's materials will operate a one-year maintenance cycle and a four-year major revamp cycle following major industry roadmaps. The materials have been designed in a modular and configurable manner to allow for the rapid creation of additional content to suit different learner needs.

Working in partnership

Arm Education Media is the culmination of several years of collaboration with educational institutions, industrial partners, students, recruiters and managers worldwide. It complements other initiatives and programs at Arm, including:

Arm Education Partnership (arm.com/markets/education)
Helps teachers and children learn with technology.

Arm University Program (arm.com/university)
Provides university academics with free teaching materials and technologies for computer engineering and related courses.

Arm Training (arm.com/support/training)
Provides professional engineers with training on core Arm technology topics

Whether you are an academic, researcher, student or professional engineer, visit the Arm Education Media website and explore our offerings. arm.com/resources/education

Career talk: Maximilian Odendahl, CEO and Co-Founder, Silexica

What was your main area of focus while you were a researcher?

I received my diploma in computer engineering from RWTH Aachen University and was formerly the Chief Engineer of the Chair for Software for Systems on Silicon, leading 15 research assistants. My work mostly focused on heterogeneous multi-processor system-on-chips, parallel programming models, software mapping exploration, scheduling and automatic data allocation.

How did you come to found Silexica?

All the research projects we did at the university had one thing in common: computing was changing faster than it had ever done in the past. Hardware was moving from single processors that were relatively easy to program, to several processors, to the thousands of processors needed for advanced technologies such as autonomous vehicles, 5G communications, robots and drones.

It was clear to me and the other co-founders that these super-computers would require new innovations to master them. For example, when you have an autonomous car computer making 320 trillion operations a second with low power consumption and zero room for error, there is no way these software algorithms can be tackled by manual source code inspection and software distribution by manual analysis in a spreadsheet. An automation tool is required – and that is why we launched our company Silexica and our solution SLX in 2014. As the demands of technology are increasing, so is the demand for our product.

What advice would you give other researchers thinking about starting a company ?

Talk to potential customers from day one and listen, listen, listen. And keep going, even if everyone tells you it doesn't make sense.

What makes you most proud about Silexica, or your career in general, so far?

The feeling of creating or building something and making it grow. Five of us started Silexica in a loft in Aachen with a plan, a vision and a pretty raw solution. Every single day since then the company and product has got bigger or stronger. We now have 50 incredibly smart and committed 'Silexicans' with offices in Japan and the USA. Another thing I find rewarding is having solutions to real and acute problems faced by the computing industry.

What makes Silexica a great place to work? Why should HiPEAC students apply for a job there?

The team spirit here is the best I have experienced. We have gathered so many outstanding software and hardware experts from around the world that they drive and empower each other on to achieve great things. Socially there is lots going on as well. We have monthly team events, healthy breakfasts, weekly team lunches, indoor football, fitness offers – but that still leaves plenty of time to work!

In this interview, we asked Esther Jiménez, dean of the School of Education at the Universitat Internacional de Catalunya and lecturer at IESE Business School, what makes a good leader and how we can attract more women to technology careers.

Career talk: Esther Jiménez, UIC Barcelona

What are the qualities necessary to be a good leader?

We tend to associate 'leadership' with power, or being the best at a given activity. However, in addition to knowledge and authority, a good leader needs integrity, empathy, emotional balance and self-control, qualities that inspire trust and produce the positive emotional atmosphere necessary to get the best out of people. A leader should be an example to others and should consider other people when making decisions.

How do diverse staff help make an organization more innovative?

A simple definition of innovation is 'something new which has a practical use'. Making an innovation – whether a product or a service – a reality requires both an idea and creative work carried out by a team of people. Conceptual perspectives on creativity emphasize the benefit of diverse points of view and knowledge bases to generate a greater variety of ideas. This makes a diverse group better placed to obtain results, and studies have shown this to be the case.

What are some of the barriers preventing women from reaching positions of leadership?

Women have been agents of change in the world of work, and have shown that they can contribute on the same terms as men.

The so-called glass ceiling, the barriers which companies impose against reconciling work, family and personal life, are what prevent women from being promoted. The determining factor is maternity: women don't tend to be discriminated against per se, but rather for the fact of being mothers. Our research confirms that:

- 69% of women are overstretched
- 20% have been unable to go back to work after having children
- 37% have not been promoted
- 45% were asked during the recruitment process if they were mothers or intended to be
- 53% have seen their career prospects dip as a result of having children
- 57% have had to give up jobs which were incompatible with having children

All this, together with stereotypes, mean that in practice women are unable to access leadership positions.

How can we attract more women to the technology field?

First of all, there need to be female role models. In addition, education has a fundamental role to play in awakening girls' interest in science, technology, engineering and mathematics (STEM) subjects. Teachers need to make use of appropriate content, materials and methodologies to create a learning environment that inspires girls to study areas related to this field.

What can we do to keep a good work-life balance?

Work, family and social life are fundamental, inherent and complementary dimensions of human beings. However, nowadays work seems to dominate everything, either because there's too much or too little of it. This distortion has consequences for people's health and quality of life. Research is a highly absorbing activity, and I would advise researchers not to lost sight of the need to have a balanced, well-rounded life. Make sure you schedule in activities in all three areas of life. Personal leadership comes first, and is the most important.

This issue's featured thesis considers how to improve performance using tailored compiling passes. HiPEAC's network includes almost 900 PhD students, producing an impressive body of research between them.

Three-minute thesis

NAME: Ricardo J. F. Nobre
RESEARCH CENTER: University of Porto, Faculty of Engineering / INESC TEC
SUPERVISOR: Prof. Dr. João M. P. Cardoso
THESIS TITLE: Efficient Target and Application Specific Selection and Ordering of Compiler Passes

Featured research: Free pass to better code

Code optimization is a vital part of current production-grade compilers. Optimization is achieved by the execution of a sequence of passes, each performing well-delimited actions over a compiler-specific internal representation of the program being compiled. The impact of applying a given compiler pass depends on which passes were executed before, which passes are going to be executed after, the features of the program being compiled, the output code generator, and the particular characteristics of the target architecture.

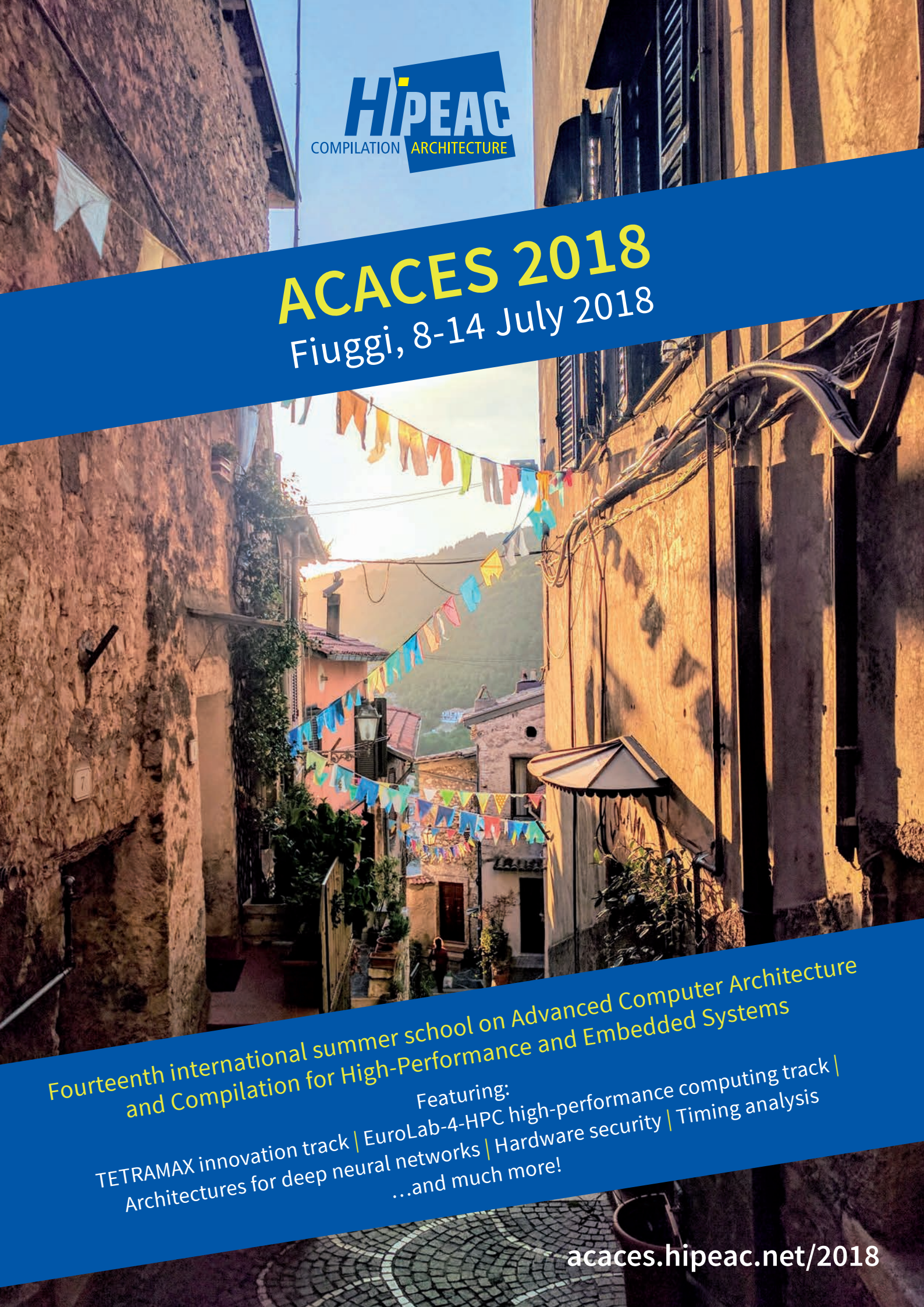
Programmers usually rely on generic compiler optimization flags, such as GCC –O3, which typically represent the execution of fixed sequences of analysis and transformation passes. This means that it is often possible to achieve better compiled code by using sequences of passes tailored to the specific input source code and target architecture. Compiler sequence specialization can result in considerable execution performance, energy consumption, and/or memory footprint enhancements in comparison with the standard optimization flags. Domains such as embedded systems or high-performance computing (HPC) tend to prioritize metrics related to energy efficiency, which typically receives less attention from compiler developers, making these domains benefit further from using specialized sequences.

SPeCS: The Special-Purpose Computing Systems, languages and tools group at the University of Porto

However, generating suitable compiler sequences can be challenging, especially with the growing number of compiler passes that are becoming available in modern compilers. In order to leverage the potential available thanks to compiler sequence specialization, it is necessary not only to select which compiler passes to execute, but also to select their order of execution. As a result, the exploration space quickly becomes too large, meaning that efficient design space exploration (DSE) algorithms and/or heuristics are required to prune the exploration space, as otherwise exhaustive exploration would be unfeasible.

We propose a number of DSE schemes, which we implemented and validated targeting processors typical of embedded systems, smartphones and supercomputers, as well as graphics processing units (GPUs). Our DSE schemes rely on prior knowledge about compiler sequences and on iterative and non-iterative approaches to propose compiler sequences. Using our approaches, we were able to efficiently and consistently achieve, with a reasonable overhead, significant performance improvements in comparison to any standard optimization flag available in current compilers such as GCC and LLVM. Finally, we developed a modular DSE system for compiler sequence specialization. This is used to make the implementation and validation of new DSE schemes easier, which, in addition, has the potential to reduce the entry barrier of new researchers to the field.

Ricardo's supervisor João M. P. Cardoso, Associate Professor in the Department of Informatics Engineering (DEI), commented: 'The main contributions of Ricardo Nobre's PhD thesis include a framework to explore flag selection and phase ordering in the context of representative compilers (e.g., a CoSy-based compiler, LLVM and GCC), and a technique that uses accumulated knowledge about phase ordering to find suitable phase orders for new functions/applications. His thesis provides results of many experiments and shows the impact of phase ordering on energy consumption and execution time for different microprocessors and GPUs.'



HiPEAC
COMPILATION ARCHITECTURE

ACACES 2018

Fiuggi, 8-14 July 2018

Fourteenth international summer school on Advanced Computer Architecture and Compilation for High-Performance and Embedded Systems

Featuring:

TETRAMAX innovation track | EuroLab-4-HPC high-performance computing track | Architectures for deep neural networks | Hardware security | Timing analysis

…and much more!

acaces.hipeac.net/2018