

HIPEAC

info 56

JANUARY 2019

HiPEAC
conference
2019
Valencia

The road ahead for computing systems

Monica Lam on keeping the web open

Alberto Sangiovanni Vincentelli on building tech businesses

Koen Bertels on quantum computing

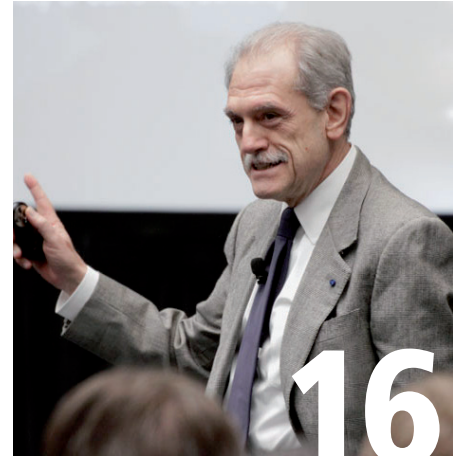
Tech talk 2030



Benvinguts a València



Monica Lam on open-source voice assistants



Starting and scaling a successful tech business

<p>3 Welcome <i>Koen De Bosschere</i></p> <p>4 Policy corner The future of technology – looking into the crystal ball <i>Sandro D’Elia</i></p> <p>6 News</p> <p>14 HiPEAC voices ‘We are witnessing the creation of closed, proprietary linguistic webs’ <i>Monica Lam</i></p> <p>16 HiPEAC voices ‘Do not think that SME status is the final game’ <i>Alberto Sangiovanni Vincentelli</i></p> <p>18 Technology 2030 Computing for the future? The way forward for computing systems <i>Marc Duranton, Madeleine Gray and Marcin Ostasz</i></p> <p>23 Technology 2030 Tech talk 2030</p> <p>24 Future compute special ‘A quantum computer is, theoretically, the most powerful computational device we can build’ <i>Koen Bertels</i></p> <p>26 Future compute special Quantum Collective Knowledge <i>Anton Lokhmotov</i></p> <p>27 Future compute special Beyond digital: Brain-picking for new computing models <i>Wim JC Melis</i></p> <p>29 Peac performance Machine-learning IoT solution for noise classification <i>Sabri Pllana</i></p>	<p>30 SME snapshot UltraSoC: Smarter systems thanks to self-aware chips <i>Rupert Baines</i></p> <p>33 Innovation Europe M2DC: The future of modular microserver technology <i>João Pita Costa, Ariel Oleksiak, Micha vor dem Berge and Mario Porrman</i></p> <p>34 Innovation Europe TULIPP: High-performance image processing for embedded computers <i>Philippe Millet, Diana Göhringer, Michael Grinberg, Igor Tchouchenkov, Magnus Jahre, Magnus Peterson, Ben Rodriguez, Flemming Christensen and Fabien Marty</i></p> <p>35 Innovation Europe Software for the big data era with E2Data <i>Juan Fumero</i></p> <p>36 Innovation Europe A RECIPE for HPC success <i>William Fornaciari</i></p> <p>37 Innovation Europe Solving heterogeneous challenges with the Heterogeneity Alliance <i>Clara Pezuela</i></p> <p>38 Innovation Europe ANTAREX: Technologies for energy-efficient HPC systems towards exascale <i>Cristina Silvano and João MP Cardoso</i></p> <p>40 HiPEAC futures More than a website: The value of HiPEAC Jobs Career talk: Vincent Hindriksen, Stream HPC Three-minute thesis: Custom accelerators made simple</p>
--	--



Computing for the future



Koen Bertels on quantum



UltraSoC's self-aware chips



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



hipecac.net



@hipecac



hipecac.net/linkedin



HiPEAC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 779656.

Cover photo: David Becker on Unsplash

Back cover photo: Papagnoc on Pixabay

Design: www.magelaan.be

Editor: Madeleine Gray

Email: communication@hipecac.net

First of all, I would like to wish you a healthy and prosperous 2019, personally as well as professionally. As we're celebrating the new year with a brand new HiPEAC Vision, setting out key issues for computing systems over the next ten years, this issue of *HiPEACinfo* is dedicated to futuristic technologies and the experts creating them.

I will remember 2018 as the year of extreme weather conditions. The summer was unusually dry; the browning of the landscape in parts of Europe could even be observed from space. Later in the year, we saw major wildfires like the one in Paradise, United States with 291 people either dead or missing at the time of writing, and 14,000 homes completely destroyed.

Climate experts warn us that we will experience more extreme weather conditions as the global temperature rises. A substantial reduction in agricultural productivity might lead to food shortages, increased prices, social unrest, famine, migration and possibly war. I strongly believe that climate change is the mother of all societal challenges this century. If we fail to stop it in the next 50 years, we might face a catastrophic collapse of society as we know it by the end of this century.

Recently, the European Commission therefore announced its ambitious plan for a climate-neutral Europe by 2050. It is clear that decarbonizing Europe in the next 30 years will not be possible without a serious change in lifestyle. Resistance is futile – the longer we wait, the higher the price we will pay. Fortunately, this ambitious plan also creates enormous opportunities for our community.

All products and processes that can be dematerialized by digitizing them require less energy to produce and to transport. The service industry is less energy-intensive than the manufacturing sector and computing is already 100% electric. We have all the key enabling technologies to meet virtually and to create virtual reality experiences that might become a substitute for some forms of tourism. Computing is also the key to more efficient private and public transportation, to smart homes, smart cities, etc. In other words, without advanced computing solutions, there won't be a climate-neutral Europe by 2050. So I strongly believe that the mission of our community should be to work on excellent climate-neutral alternatives for fossil fuel based processes.

Koen De Bosschere, HiPEAC coordinator



How should Europe approach the future of digital technology? Answering this will be tricky, but at least we know that some good policy tools will be available. In this article, DG CONNECT's Sandro D'Elia introduces Horizon Europe and Digital Europe and explains how these programmes aim at making Europe a better place to live and work.

The future of technology —

If we compare the daily routine of someone today with someone living 50 years ago, we see that technology has greatly improved the average quality of life. Technology, in particular digital technology, is no longer exclusively in labs; on the contrary, it has a direct impact on our daily lives. This is a clear global trend: technology is becoming more and more important for the economy and for society. It forms part of almost every moment of our personal and social lives, and its value in economic terms is enormous.

Indeed, today technology is a strategic element of human and social development. The difference between those who have direct access to technology and those who depend on others is growing at all levels: between citizens in the same city or in the same region, between countries, between macroregions at the global level. The world is splitting between digital leaders and digital colonies, and this is not a positive development.

Moreover, technology creates wealth and jobs. Of course, there are specific cases where this is not true: in some jobs, computers and robots will soon replace humans, but if you look at the overall picture and check the numbers, you will discover that companies (and countries!) which invest more in technology are those where more jobs are created.

In this context, we cannot do business as usual if we want to keep Europe in the very small club of global technological leaders. Supporting research is not enough; we do need to develop cutting-edge technologies, but we need also to bring those technologies out of the labs and deploy them widely across Europe, to drive the digital transformation of society.

This is why the European Commission has proposed something radically new for the future. In the period 2021-2027 we propose to invest €9.2 billion in the new 'Digital Europe Programme', which will complement the future 'Horizon Europe' programme, the follow-up to Horizon 2020. Digital Europe will focus on capacity building and deployment, with the support of co-investment from Member States, and will have a specific emphasis on high-performance computing, artificial intelligence, cybersecurity and advanced digital skills.

"I dream of computers capable of helping me even if I have not read the manual, and which I will be able to instruct precisely without having to become an expert in programming languages, IDE, debuggers, profilers and test automation"



looking into the crystal ball

This is a very important change: for the first time, the European Union (EU) will have a financial tool to invest specifically in digital technologies, and to make sure that those technologies are widely used across the whole of society and in all economic sectors. With Digital Europe, we will be able to support strategic investments that require EU-level cooperation, like exascale computing centres or large test facilities, which are beyond the capacity of individual Member States.

Horizon Europe and Digital Europe will be complementary in their support of digital tech: Horizon Europe will mainly fund research and technological development including piloting,

proof-of-concept, testing, innovation and pre-commercial deployment, while Digital Europe will focus on large-scale digital capacity and infrastructure building, with the objective of wide deployment across Europe of critical existing or tested innovative digital solutions.

In conclusion, with the Horizon Europe and Digital Europe programmes, the European Union will have the instruments needed to support the development and deployment of digital technologies in the coming years.

We don't yet know what these technologies will become, but I would like to place a few bets. I believe that in 2025 'artificial intelli-

gence' will be significantly different from the neural networks we see today, and that mainstream computer components will consume only a fraction of the huge amount of energy they need now. However, what I would really like to see is a different way of interacting with computing devices, in which we humans do not have to adapt our way of thinking and working to the constraints of technology as is the case today. I dream of computers capable of helping me even if I have not read the manual, and which I will be able to instruct precisely without having to become an expert in programming languages, integrated development environment (IDE), debuggers, profilers and test automation.

Will this ever be possible? I am not sure, but I would be happy to see this as a result of European research in the coming years. With Horizon Europe and Digital Europe, there will be enough financial resources for this and many other exciting developments.

MORE INFORMATION:

An interview with Sandro on the Horizon Europe and Digital Europe programmes is available on the HiPEAC YouTube channel: bit.ly/HorizonEU-DigitalEU

**Investing in the future:
Digital Europe Programme**

€1.3 billion for Deployment, use of digital capacities & interoperability

- Provide more **interoperable** public services
- Support digital & related **technologies** for industry & SMEs
- Support and follow latest **tech developments**
- Bring digital technologies to **public sector** and areas of public interest
- Offer **tests** and **pilots** to public administrations
- Build up and strengthen the network of **Digital Innovation Hubs**

#EUBudget #DigitalEurope

European Commission



Photo credit: Xavier Martorell

A Computing Systems Week with a view

With over 200 participants, HiPEAC's autumn Computing Systems Week, hosted by the Foundation for Research and Technology – Hellas (FORTH), was a huge success. Keynote talks by Dushyanth Narayanan (Microsoft Research), Franck Capello (Argonne National Laboratory) and Bastian Koller (University of Stuttgart / HLRS High Performance Computing Center) discussed cloud computing, lossy compression and high-performance computing (HPC) respectively.

For anyone interested in the latest industry developments, the HiPEAC Industry Partner Programme provided an indispensable introduction to some of the most innovative companies in computing. Meanwhile, students got the chance to prove themselves and win a Neural Compute Stick, kindly provided by HiPEAC member Intel Movidius, in the HiPEAC Student Challenge. Continuing the student track, the Inspiring Futures session featured compelling talks helping early career researchers decide on the right career path for them.

The next Computing Systems Week takes place in Edinburgh on 16-18 April, with a theme of HPC and innovation. We look forward to seeing you there.

Videos from Computing Systems Week Heraklion, including full talks and brief interviews, are available to view on the HiPEAC YouTube channel.

youtube.com/c/HiPEAC



The main theme of data centres, storage and networking was explored in a variety of sessions, including dedicated Eurolab4HPC sessions on open source, accelerators and storage. Meanwhile, talks on the latest developments in European strategy addressed the forthcoming ICT-01-2019 call for proposals, the European Processor Initiative and cyber-physical systems recommendations from Platforms4CPS.



Olivier Zendra joins HiPEAC steering committee, replacing Albert Cohen

After several years on the HiPEAC steering committee, Albert Cohen is now standing down, as he has left HiPEAC partner INRIA to join Google. The HiPEAC steering committee has benefitted enormously from Albert's expertise and industry knowledge over the years. We would like to thank him for his contribution and look forward to working

with him as a HiPEAC industry member in his new role.

INRIA will now be represented on the steering committee by Olivier Zendra, an active HiPEAC member who is on the HiPEAC Vision editorial board. We look forward to working more closely with Olivier over the next few years.



Olivier Zendra (left) will replace Albert Cohen (right) on the HiPEAC steering committee

Benvinguts a València! welcome to Valencia



A rich culture, delicious food and sunshine are just some of the things that make Valencia, this year's HiPEAC conference city, special. HiPEAC19 General Chair José Duato explains.



What makes Valencia such a good place to host HiPEAC19?

A popular tourist destination, Valencia offers the visitor a wide variety of things to do. These include visiting historic buildings and unique futuristic architecture, enjoying our rich and varied gastronomy, heading to the beach or visiting museums, all with mild temperatures during January. Valencia enjoys an average of over 300 sunny days per year, but please don't blame me if it ends up raining during the HiPEAC conference.

In 2018, the Valencia Conference Centre, the venue for HiPEAC19, won the Apex Award from the International Association of Convention Centres, reflecting the highest client rating, so participants are in good hands.

From a more academic point of view, two universities have their main campus in Valencia. They complement each other very well and are among the top ten universities in Spain. Currently, more than 80,000 students are enrolled in those universities. The quality of the universities, the hospitality of the people, the charm of the city, and the Mediterranean climate combine to make Valencia a very attractive city both for international projects and student exchange programmes.

What are some of the most exciting projects taking place at UPV at the moment?

The most challenging project we recently started consists of developing a network of highly specialized, directly interconnected hardware accelerators for deep learning. The main difficulty is not designing the hardware architecture but finding an efficient way to map sparsely interconnected neural networks into the regular, densely interconnected arrays of processing units provided by the hardware.

Another very challenging project we are starting, in collaboration with Delft University of Technology, is the design of efficient communication support for moving quantum states around in large 2D qubit lattices with only neighbour-to-neighbour communication. This is necessary to enable quantum computers to scale to large sizes.

In addition to these projects, we are continuing to develop our more traditional research lines, most of them in collaboration with industry. In particular, we have finally found a definitive solution to the network congestion problem, and are developing specific mechanisms to implement it in lossless ethernet, also patenting some of the ideas and standardizing the mechanisms through IEEE. We have also developed an infrastructure that enables us to design and

deploy complex manycore hardware designs on a network of field-programmable gate arrays (FPGAs). In addition, our rCUDA software technology to virtualize and enable remote access to graphics processing unit (GPU) accelerators in a cluster is very mature and now also incorporates support for deep learning libraries.

What should HiPEAC19 participants try or do during their stay in Valencia?

I would definitely recommend visiting some of the historical monuments in the old city, and trying our varied gastronomy. The buildings in the City of Arts and Sciences are a wonderful example of futuristic architecture; indeed, some of the scenes of the movie *Tomorrowland* were recorded there. Meanwhile, if you're interested in sea life, the Oceanographic Park contains more than 45,000 animals from 500 different marine species.

In the old city my favourite building is the Lonja de la Seda, the silk trade centre from the 15th century, but there are many other monuments within walking distance. Finally, if you'd like to try something other than the well-known paella, I would recommend the 'arròs banda' or the 'arròs del senyoret', preferably at one of the many restaurants at Malvarrosa beach or in La Marina, followed by a walk next to the sea.

SpiNNaker reaches target of one million Arm cores

Steve Furber, University of Manchester

The SpiNNaker (Spiking Neural Network Architecture) project at the University of Manchester, UK, held an event on 2 November to celebrate achieving the target originally set in 2009 to deliver a platform for brain science incorporating over a million Arm processor cores.

The agenda included talks by Collette Fagan (University of Manchester Vice-President of Research), Anna Angus-Smyth (Head of ICT at the UK Engineering and Physical Sciences Research Council), Steve Furber (Head of the SpiNNaker project), Sacha von Albada (Jülich Forschungszentrum, Human Brain Project partner) and Michael Denham (Chief Executive, Mindtrace Ltd), covering the history of the project and the use of the machine in brain modelling and commercial development. There were also tours to see the million-core machine, and suitably celebratory refreshments, including specially commissioned SpiNNaker badges and balloons, and cupcakes carrying a range of logos! Many past members of the SpiNNaker team joined us, adding to the sense of reunion and celebration.

A half-million core SpiNNaker has been offering a free service under the auspices of the Flagship Human Brain Project for two-and-a-half years to external users across Europe and around the globe, and over 100 SpiNNaker boards have been delivered to user groups from North America to New Zealand, so the technology is widely available. It offers both the world's largest (in terms of the number of spiking neurons that can be modelled in biological real time) and most flexible (through the use of software for neuron and synapse models) neuromorphic platform available today. The expansion to the full million cores fulfils a long-term ambition and opens up the possibility of running larger models, such as a full mouse brain, as well as running many smaller models at the same time.

apt.cs.manchester.ac.uk/projects/SpiNNaker



Left: the million-core SpiNNaker machine.



Right: SpiNNaker team members are asked to stand at the event

“SpiNNaker offers the world’s largest and most flexible neuromorphic platform”

TANGO white paper: Simplifying heterogeneous hardware architectures

Clara Pezuela, Atos Spain



In heterogeneous systems, achieving optimization in performance and power consumption is challenging for developers, as software needs to be correctly fitted to the capabilities of the underlying hardware. The Horizon 2020 TANGO project, which concluded on 31 December 2018, recently published a white paper highlighting the outcomes of the project.

TANGO aims to simplify the way developers approach the development of next-generation applications based on heterogeneous hardware architectures, configurations and software systems including heterogeneous clusters, chips and programmable logic devices.

TANGO impacts both the information technology (IT) industry and market in Europe, by designing more flexible software abstractions and improved system architectures to fully exploit the benefits of these heterogeneous platforms, while addressing energy optimization at the same time.

The primary outcome is the TANGO toolbox, a set of modular, open and interoperable tools to design, model, develop and execute optimized software in heterogeneous hardware, through the optimization of various dimensions of software design and operations such as energy efficiency, performance and dependability on target architectures.

TANGO can be used in several scenarios, such as to facilitate the programming, automatic deployment and energy optimization of high-performance computing (HPC) or embedded applications in heterogeneous devices, for example.

TANGO has launched the Heterogeneity Alliance, a joint initiative with other projects and organizations to create a community of interest and business around heterogeneity challenges, nurturing strong research collaboration, integration and effective promotion of results.

Turn to p.37 for the latest news on the Heterogeneity Alliance

bit.ly/TANGO_white_paper

Last chance to apply for FED4SAE funding and expertise

Isabelle Dor, CEA-Leti



The FED4SAE Innovation Action, part of the European Commission's Smart Anything Everywhere initiative, has launched its third and final call for smart application projects. The deadline for submitting applications is 5 March 2019, and 15 projects will be selected for funding.

Now in its third year, FED4SAE brings companies together with both major cyber-physical system (CPS) platform providers and competence centres. In addition to up to €58,000 of funding, the Innovation Action provides access to leading-edge industrial platforms, advanced technologies and testbeds.

Additional support through expertise, knowhow, coaching, design support and technology transfer from FED4SAE partners enables rapid development, as well as lowering the entry costs for designing and optimizing. This leads to new business and innovation opportunities, which in turn may lead to quantifiable increases in the participating companies' market share, productivity and industrial capacities.

The kind of experiments which will be funded will address any smart application domain, involving both advanced and industrial platforms supported by FED4SAE. The experiments will be pan-European, allowing the awarded company to collaborate with cross-border advanced platforms and industrial partners.

FED4SAE is already supporting 16 projects with SMEs coming from all over Europe (Serbia, Spain, UK, Italy, Cyprus, France, Hungary, etc.) and is looking forward to granting 15 additional experiments. You just need to provide the idea and we will help you get it to market – don't miss this opportunity!

Further information: fed4sae.eu



TETRAMAX: Full speed ahead towards innovation

Annual Industrial Advisory Board meeting held in Aachen

Rainer Leupers, TETRAMAX Project Coordinator



TETRAMAX, the one-stop shop for digitizing European small / medium enterprises (SMEs), reflected on its first year of operation during a meeting with its Industrial Advisory Board (IAB) on 13 September at RWTH Aachen University. Among the 45 attendees were high-calibre, director-level IAB members as well as representatives from all 23 consortium partners. In addition, all Technology Transfer Experiments (TTX) currently funded by TETRAMAX via open calls introduced their novel customized low-energy computing technologies and discussed their business plans.

The IAB meeting provided an invaluable external bird's-eye view on achievements so far and future strategies. Likewise, it marked a major team-building milestone and was also instrumental in preparing carefully for the successful formal EC review in October.

Interested in leveraging TETRAMAX for your tech transfer endeavours? Visit our website to learn more about our TTX open call schedule (short proposals, up to €100,000 in funding), the EU-wide technology brokerage network, and other unique offerings.

"An impressively led and managed European competence network that drives true technology brokerage sustainability. It breaks the code on how to effectively and efficiently link EU funding, university research and industry success. Its current focus on computing will pave the way to the next exponential productivity jump to support Industry 5.0 via the heart of the European social and economic fabric: SMEs and midcaps."

Dr. Oscar Gallego Castilla, GE Digital Europe Director, General Electric

"TETRAMAX is an impressive example of what can be achieved if European partners push for one common goal: transfer academic excellence into new businesses. Building upon years of experience in similar programmes by applying constant learnings and novel ideas provides an extremely vivid framework for success."

Dr. Matthias Weiss, Managing Director, CommSolid GmbH

Further information: tetramax.eu

dividiti join the MLPerf initiative

Anton Lokhmotov, dividiti



At dividiti, we've been crusading for community-driven benchmarking for over a decade, so we're pleased to see that kindred ideas are beginning to capture the imagination of a broad community interested in benchmarking artificial intelligence (AI) / machine learning (ML) systems (models / software / hardware) via the MLPerf initiative. Community involvement means using at least three good things: representative workloads, rigorous and fair methodology, and state-of-the-art workflow automation.

First, benchmarking workloads included into MLPerf should be representative of real workloads and thus enable meaningful comparison between systems. For example, much of early deep learning benchmarking involved AlexNet, the winner of the ImageNet 2012 classification challenge. AlexNet includes 11x11, 5x5 and 3x3 convolution layers. Modern networks, however, tend to use 3x3 and 1x1 convolution layers. Therefore, even if one knows how fast a system can run AlexNet, one cannot easily predict how fast the same system will run another network. MLPerf should steer clear of such non-representative workloads.

Second, benchmarking methodology should be refined over time to avoid intended or unintended bias. For example, traditional benchmarking suites often collapse multiple scores into one 'summary' score. It may make sense when you deal with one metric (typically, the execution time) and similar workloads. When you consider multiple metrics (such as the execution time and accuracy necessary for benchmarking inference) and diverse workloads (for example, object detection and speech recognition), however, it becomes extremely hard, if not impossible, to summarize sensibly.

“In the first ACM ReQuEST competition, we collected five portable and customizable Collective Knowledge workflows for image classification inference”

Third, we hope that over time the AI / ML benchmarking community will transition to using automated reusable workflows. No matter how detailed a description of an experiment is, if the experiment involves any manual steps, it is susceptible to human errors and omissions. Given the complexity of AI / ML workloads and systems (comprising models, tunable hyperparameters, datasets, frameworks, libraries, platforms, etc.), workflow automation is a must to achieve reproducibility and thus credibility. In addition, automation helps both researchers and practitioners alike by reusing best practice, reducing effort and improving quality.

In the first ACM ReQuEST competition, we collected five such portable and customizable Collective Knowledge workflows for image classification inference covering diverse deep learning models and systems ranging from IoT devices to servers. The community can now build upon these workflows to play with new systems, models, datasets, etc., while avoiding many pitfalls awaiting the unwary (such as thermal throttling on mobile platforms).

We have already contributed to MLPerf one of these image classification workflows, as well as workflows for object detection and speech recognition. These are among the very first MLPerf inference workflows, both for edge and cloud.

We are excited about the benefits that community involvement brings to benchmarking and optimization of emerging workloads from AI / ML to quantum computing, and encourage you to join a growing Collective Knowledge community!

FURTHER READING:

MLPerf

mlperf.org

Collective Knowledge framework

cknowledge.org

Collective Knowledge community

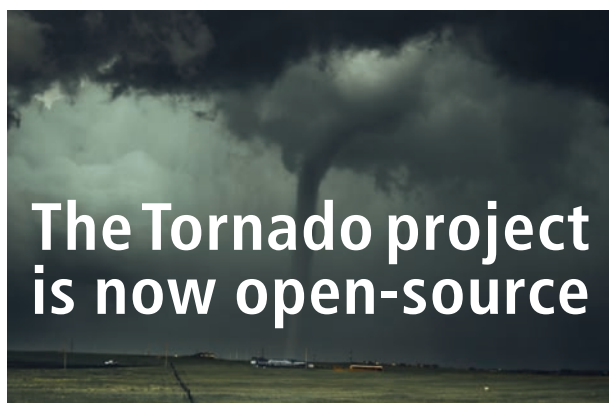
cknowledge.org/contacts

First ACM ReQuEST Workshop/Tournament on Reproducible Software/Hardware Co-design of Pareto-Efficient Deep Learning

cknowledge.org/request

LinkedIn article by Grigori Fursin on results, reusable workflows and ACM proceedings from the ReQuEST workshop

bit.ly/LinkedIn_1st_ReQuEST_workshop



The Tornado project is now open-source

Start accelerating your Java programs on GPUs and FPGAs for free

Juan Fumero, University of Manchester



Tornado, a practical heterogeneous programming framework for Java, has been released as an open-source project in GitHub. Tornado enables automatic just-in-time (JIT) compilation and acceleration of Java programs on any OpenCL

compatible device, such as multicore central processing units (CPUs), graphics processing units (GPUs) and field-programmable gate arrays (FPGAs).

Tornado provides a minimal parallel task-based application programming interface that programmers can use to express computation to be offloaded and accelerated on GPUs and FPGAs. In addition, Tornado includes a runtime that automatically manages and optimizes memory transfers between heterogeneous devices and the main CPU. The Tornado JIT compiler extends the new Oracle Graal JIT compiler with new information that guides the compilation and optimization process for heterogeneous architectures, allowing to reuse and easily extend typical Java virtual machine optimizations for new devices.

Currently, Tornado is included as a component in a European project called E2Data. This project aims to provide end-to-end solutions for accelerating big-data workloads that fully exploits distributed memory systems and heterogeneous hardware. Tornado supports the E2Data project by providing a compiler framework and a runtime system in which compile and execute, big data programs implemented in Java on heterogeneous hardware.

Read more about E2Data on p.35

Tornado project

github.com/beehive-lab/Tornado

Graal JIT compiler

github.com/oracle/graal

When a HERO comes along

HERO: PULP's Open Heterogeneous Research Platform



Alessandro Capotondi (University of Bologna), Primin Vogel and Andreas Kurth (both ETH Zürich)

The Parallel Ultra Low Power Platform (PULP) recently released HERO, its Open Heterogeneous Research Platform. HERO combines a PULP-based open-source parallel manycore accelerator implemented on FPGA with a hard Arm Cortex A multicore host processor running full-stack Linux. It is the first heterogeneous system architecture to mix a powerful Arm multicore host with a highly parallel and scalable manycore accelerator based on RISC-V cores.

HERO offers a complete hardware and software platform which advances the state of the art of transparent accelerator programming using the OpenMP v4.5 accelerator model. The programmer writes a single application source file for the host and uses OpenMP directives for parallelization and accelerator offloading. Lower-level details such as differing instruction set architectures (ISAs) as well as shared virtual memory (SVM) between host and accelerator are handled by our heterogeneous toolchain based on GCC 7, runtime libraries, kernel driver and our open-source hardware intellectual property (IP).

As such, HERO greatly simplifies heterogeneous systems programming and forms a complete basis for future system-level research and industrial design comprising both the hardware and software side of heterogeneous systems.

To get started, we suggest you try the HERO software development kit (SDK) repository, as detailed below.

HERO SDK repository

github.com/pulp-platform/hero-sdk

PULP hardware repository

github.com/pulp-platform/bigpulp

HERO documentation

pulp-platform.org/hero/doc

[@pulp_platform](https://twitter.com/pulp_platform)

Call for submissions: Year-round Championship Value Prediction

Arthur Perais, Qualcomm



Following the success of the first Championship Value Prediction workshop, held at the ACM / IEEE International Symposium on Computer Architecture (ISCA) 2018, the Championship Value Prediction (CVP) organizing committee is pleased to announce that CVP is now a year-round competition.

We believe this new format is beneficial: while there is no deadline, you still have the opportunity to attend and present at a workshop. Indeed, the organizing committee pledges to organize a CVP workshop at top-tier conferences regularly and to invite contestants to present their work, depending on the number of new submissions.

In addition, CVP provides a high-quality framework including industry-generated instruction traces for computer science and electrical and computer engineering (ECE) students looking for an interesting project with potential real-world impact.

The goal for this competition is to compare different value prediction algorithms in a common framework. Predictors will be evaluated for all instructions producing a general-purpose register. Predictors must be implemented within a fixed storage budget as specified in the competition rules. The simple and transparent evaluation process enables dissemination of results and techniques to the larger computer architecture community and allows independent verification of results.

All source code, write-ups and performance results will be made publicly available through the leaderboard.

We look forward to receiving new submissions, and we encourage you to spread the word about CVP's new format and mission to colleagues and / or students who could potentially be interested.

CVP leaderboard [↗ microarch.org/cvp1/contestants.html](http://microarch.org/cvp1/contestants.html)

New CVP website [↗ microarch.org/cvp1](http://microarch.org/cvp1)

CVP-1 website [↗ microarch.org/cvp1/cvp1](http://microarch.org/cvp1/cvp1)



Journal: Advances in Cyber-Physical Systems



Professor Anatoliy Melnyk,
Lviv Polytechnic National
University

The editorial board invites HiPEAC members to submit manuscripts to the journal Advances in Cyber-Physical Systems (ACPS), which offers an open-access option and fast processing times. Peer-reviewed and open access, the journal is published by the Lviv Polytechnic National University, Ukraine.

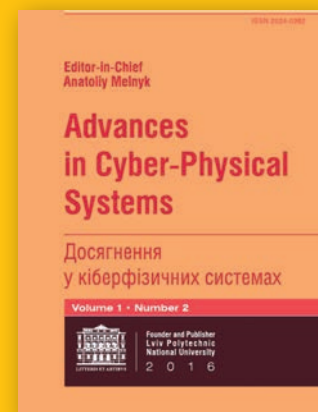
ACPS publishes timely, novel and high-quality recent results that advance the state of the art and practice in cyber-physical systems (CPS). These include theoretical foundations, system modelling, architecture and design for embedded, high-performance, reconfigurable and self-configurable systems, benchmarking and more. The full list of topic is available on the ACPS website.

ACPS accepts original articles in areas of its scope, including technical contributions, short communications and novel surveys. Proposals for special issues in cutting-edge and newly developing areas of cyber-physical systems are encouraged, and should be discussed with us.

The editorial board is guided in its work by international ethics of scientific publications and maintains a policy of open access. Articles in the journal are published free of charge. The journal is published in English twice a year.

Advances in Cyber-Physical Systems

[↗ science.lpnu.ua/acps](http://science.lpnu.ua/acps)

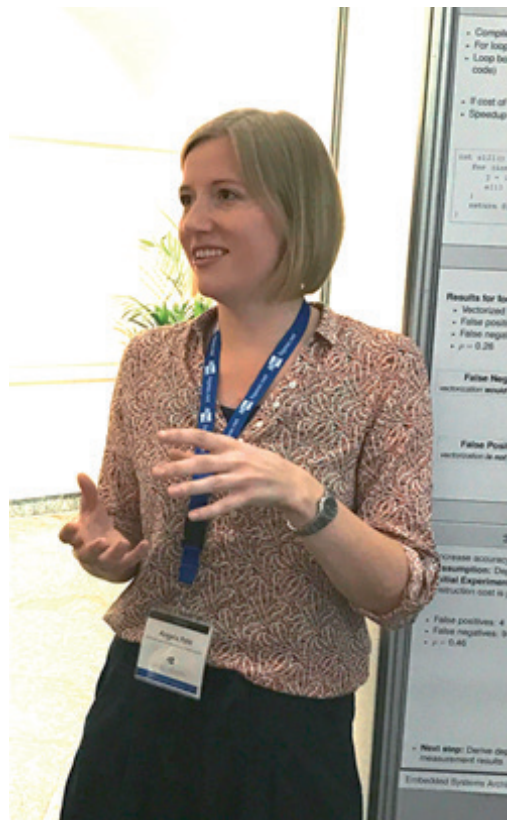


Angela Pohl wins Best Presentation Award at SCOPES 2018

HiPEAC student Angela Pohl received the Best Presentation Award at the 21st International Workshop on Software and Compilers for Embedded Systems (SCOPES18). The PhD student from TU Berlin presented the full paper 'Control Flow Vectorization for ARM NEON', which was co-authored by Nicolás Morini, Biagio Cosenza, and Ben Juurlink. In this work, the authors discuss the capabilities of compilers' auto-vectorization passes and present strategies to overcome the lack of masked instructions, which are critical to vectorize loops with control flow. The work was selected for the award through votes by attendees at the conference.

The 21st edition of SCOPES was held in St. Goar, Germany, and showcased more than 20 presentations from the field of embedded systems. More information about this ACM sponsored event can be found on the website below.

scopesconf.org/scopes-18



Angela Pohl at the ACACES 2017 poster session

Dates for your diary

ISC High Performance 2019

16-20 June 2019, Frankfurt, Germany

February 2019: deadlines for tutorials, project posters, the PhD Foun, workshops, birds-of-a-feather sessions and research posters.

isc-hpc.com/isc-2019

HPCA2019: International Symposium on High-Performance Computer Architecture

16-20 February 2019, Washington DC, USA

HiPEAC Paper Award conference

hpca2019.seas.gwu.edu

ISVLSI (IEEE Symposium on VLSI)

15-17 July 2018, Miami, Florida, USA

Paper deadline: 17 February 2019

isvlsi.org

DATE19: Design, Automation and Test in Europe

19-23 March 2019, Florence, Italy

date-conference.com

Don't miss HiPEAC activities: in addition to our usual booth and jobs wall, this year we will be organizing an 'Inspiring Futures!' careers session in the Exhibition Theatre.

EMIT2019: Emerging Technologies Conference

9-11 April 2019, Huddersfield, UK

emit.tech



Computing Systems Week Edinburgh

16-18 April 2019, Edinburgh, UK

hipeac.net/csw/2019/edinburgh

FCCM 2019: 27th IEEE International Symposium on Field-Programmable Custom Computing Machines

28-30 April 2019, San Diego, California, USA

HiPEAC Paper Award conference

fccm.org/2019

EuroHPC Summit Week 2019

13-17 May 2019, Poznań, Poland

exdci.eu/events/eurohpc-summit-week-2019



Virtual assistants are becoming increasingly popular, but with a few companies dominating the market, consumer choice is limited and the open web is under threat. We caught up with HiPEAC19 keynote speaker Monica Lam, Professor of Computer Science at Stanford University, to find out why it's so important to establish open-source virtual assistant technology, accessed using natural language.

'We are witnessing the creation of closed, proprietary linguistic webs'

How did you get into computer science in the first place?

Growing up, I didn't come across any subject or profession that appealed to me. That all changed when I started college and took my first computer science class. This was 1977, and I had never seen a computer before, since there were no personal computers then. I fell in love with computer science right away. I found it fascinating how programming lets us translate one's thoughts into powerful systems so easily.

Why is it problematic for just a few companies to control the digital services we rely on?

Simply put, monopolies stifle innovation. Computer science has always benefitted from having a lot of startups exploring new ideas, and the competition makes every company work harder to provide better services. There are many examples of how monopolies lead to ossification in the field. Microsoft was ruled a monopoly and has been found to abuse its power to stifle competition. The first smartphones were Windows phones, but they lost out to more innovative products like the iPhone and Android, which were made possible by open-source derivatives

of Unix and Linux. In general, we want to make technology available to everybody to enable many innovative products.

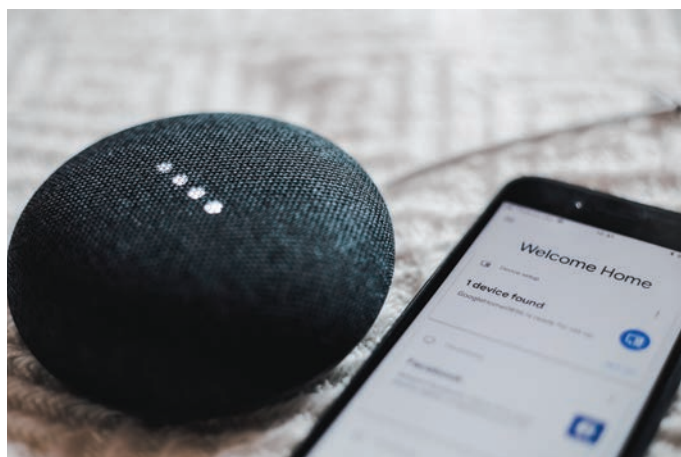
Do we have to sacrifice privacy in return for the convenience of seamless digital services?

We don't – the key is competition. Consider email as an example. Email is an open protocol: many different companies can provide an email service. There are free services that read user emails and serve ads; there are also paid services that don't. In fact, you can even run your own email server at home.

How do virtual assistants represent a paradigm shift in terms of the way the web works?

In the past, we had to learn the computer's interface; now we get to interact with the computer using the human interface – natural language. Natural language will be the primary interface in the era of ubiquitous and mobile computing.

Virtual assistants let us access the web and internet of things (IoT) like a browser does. But there is a huge difference between the



Virtual assistants are becoming increasingly popular Photo credits: Bence Boros and Andres Urena on Unsplash

two. On a browser, we can access any public webpage. In contrast, developers need to enter their services onto the proprietary platforms owned and controlled by the virtual assistant before we can access these services, which are referred to as skills or actions. In a sense, we are witnessing the creation of closed, proprietary linguistic webs.

We absolutely need to keep the web open – we need an open repository of skills that can be accessed by anyone in natural language.

What is Almond? Why are open source and interoperability so crucial to this project?

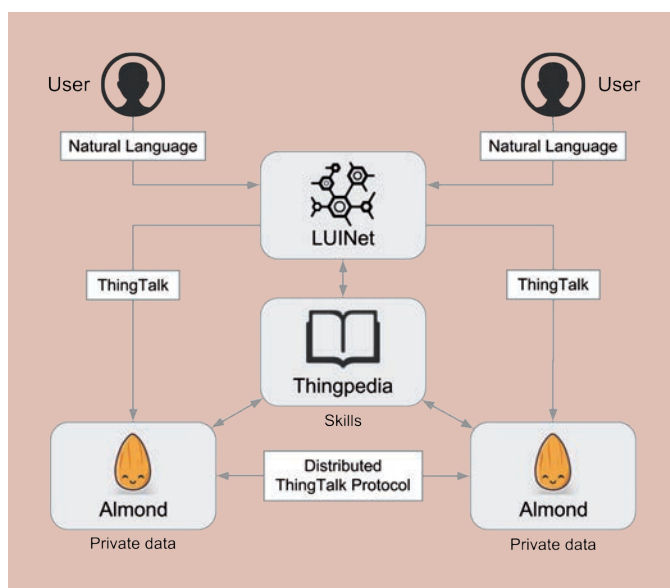
Developing virtual assistants is not only expensive; natural language research is also still an open problem. Amazon has more than 5,000 employees working on Alexa, and the product is still far from perfect. There are only a few other companies that can attempt to build a virtual assistant on their own. On the other hand, natural language interfaces are necessary for all customer-facing software in the future. Natural language interfaces are not a technology that a business can outsource to companies like Amazon and Google. A business needs to keep its own data and its own customer relationships, and it cannot rely on companies that are or could become its competitors.

The only alternative to monopolistic platforms is open-source technology. We need to pool our research and development efforts together to create competitive virtual assistant technology.

Almond is a project we started in 2015 to build the foundation for open source virtual assistant development. Our virtual assistant lets users compose new skills in natural language and gives them control over who, what, when, where, and how their data are to be shared. Most importantly, we have developed tools that make it easy to extend the functionalities of Almond. We are making it open, and are inviting academia and industry to contribute to it and use it in their research.



Photo © Kadettmann | Dreamstime.com



In addition, it has a federated architecture, meaning that individuals can choose different Almond service providers, and these assistants can interoperate to support sharing. The combination of open-source and interoperability is key to providing a choice of services and privacy to those who want it.

I am very thankful to all the students who made Almond possible: Giovanni Campagna, Silei Xu, Mehrad Moradshahi, Michael Fischer, Rakesh Ramesh, Richard Yang. We also got a lot of advice from Richard Socher, Chief Scientist at Salesforce.

Do you think there is enough will and enough resources to make open-source projects like Almond work?

Yes. This technology is critical to every business, and has already received a lot of interest from many different institutions. We have also seen successes with large-scale collaboration; for example, the crowdsourced Wikipedia is bigger than any proprietary database. Another example is software-defined networks (SDN): through an open-source collaboration of many companies, a new approach to designing networking equipment is being adopted across the industry globally.

“The combination of open source and interoperability is key to providing a choice of services and privacy”

'Do not think that SME status is the final game'

Photo credit: © Carlos Cacetano | Dreamstime.com

The founder of two leading companies in electronic design automation, HiPEAC19 keynote speaker Alberto Sangiovanni Vincentelli – the Edgar L. and Harold H. Buttner Chair of Electrical Engineering and Computer Sciences at the University of California at Berkeley – has first-hand experience of what makes a technology business successful. We spoke to Alberto to find out how he got into computer engineering, where he would start a business now and what the future holds for chip design.



“The most interesting direction, in my opinion, is the development of sensors and actuators – the ‘more than Moore’ direction”

How did you first get into computer engineering? What then attracted you to your areas of focus, like cyber-physical systems and design automation?

I actually got into computer engineering in a roundabout way. I started as a theory person working on systems and numerical analysis. When I moved from Europe to California in 1975, I was lured into more practical side of engineering: integrated circuits. In 1983, the algorithms I developed with colleagues such as Richard Newton and Bob Brayton caught the attention of industrial concerns such as Intel, Motorola, ATT and others, so we decided to start two companies: Cadence and Synopsys. These two are now the leaders in electronic design automation and are public companies traded on NASDAQ.

In 1988, I decided to move towards cyber-physical systems with a foray into the automotive domain when I started collaborating first with Magneti Marelli, and then Mercedes, BMW and General Motors. Recently I have been involved in the aerospace domain with Collins Aerospace and in construction with Lendlease.

Where do you think chip design is going to go over the next ten years, given the end of Dennard scaling and Moore's law?

The most interesting direction in my opinion, given my interest in cyber-physical systems, is the development of sensors and actuators (the ‘more than Moore’ direction) whereby the integration of different technologies and physical domains is taking place. The increasing cost and complexity of moving to new nodes is making life difficult for most of the semiconductor companies, hence the Global Foundries announcement that it won't be pursuing 7nm technology. I personally believe that the strong movement towards consolidation is going to substantially reduce the number of viable semiconductor companies, even if governments are trying to oppose some of the mergers and acquisitions in this domain on the ground of strategic national interests.



The Silicon Valley ecosystem is world famous, while Singapore is a gateway to East Asia. Photo credits: Jakub Gorajek and Fancy Crave on Unsplash

Would you rather start up a business in Europe, the United States or somewhere else? On the flip side, where would you rather study?

Today, I would still start up a business in the United States, although I have started businesses in Europe. The Silicon Valley ecosystem is still the best in terms of offering opportunities to entrepreneurs, even though the cost of locating in Silicon Valley and the scarcity of human resources is making this area of the world less appealing. Of course, China is the big land of opportunities even if it is quite difficult to start a business there considering regulations and the weakness of business protection legislature. Singapore is another country where the government is very business friendly and it can be used as a gateway to East Asia.



*Milan, where Alberto studied as an undergraduate
Photo credit: Igor Saveliev on Pixabay*

On the education side, I am very grateful for my Italian high-school education and for my years at the Politecnico di Milano. This education gave me a broad cultural base that has served me exceptionally well both for my academic and industrial career. However, I am sure that combining this background with a PhD in United States would have given me insights and competences that I had to build on my own.

What advice would you give someone thinking about setting up a technology business based on their research?

Be sure that your business is ‘defensible’ – i.e. based on strong scientific foundations to protect your ideas beyond the protection which can be provided by patents. Intercept the future needs of your chosen customer base to build a large business over time. Do not think that small/ medium enterprise (SME) status is the final game. Think carefully on how to scale up your business over the years.

What technology would you most like to become a reality by 2030?

I must say that most of the time technology develops into business opportunities in unexpected ways. Figuring out what will become a reality in five years is already a challenge. In more than 10 years it is impossible, at least for me...

“Think carefully on how to scale up your business over the years”

Every two years, we publish the HiPEAC Vision, which highlights ongoing evolutions in the computing systems domain and outlines recommendations to guide future developments. Published in January 2019, the seventh edition is a thorough exploration of the field, from business trends and societal issues to the technology options being investigated and the position of Europe in the field.

Computing for the future?

The way forward for computing systems

HIPEAC VISION 2019 IN A NUTSHELL

The overarching message of the HiPEAC Vision 2019 is that computing is at a crossroads. Digital technology has made dizzying progress over the last 50 years, and has changed our lives immeasurably. Remember when you actually had to go to a shop to buy your music or take out a film? Or when you had to physically go to a bank to move money from one account to another? Or when written communications took days to arrive, or the only way to find your way somewhere was to look it up on a map?

Today, the possibilities of an interconnected, heterogeneous and intelligent world are only just starting to make themselves known. Improved techniques in artificial intelligence are opening up new ways of dealing with the vast amounts of data being produced, and are even paving the way to a whole new way of interacting with computers.

Computing is a continuum from huge high-performance computing (HPC) installations and vast data centres providing cloud computing to the tiniest devices at the edge of the network. Voice assistants and self-driving cars are just some of the most well-known applications of new artificial intelligence technologies drawing upon this spectrum, with undreamed-of ones yet to emerge. Computing systems are becoming increasingly part of the fabric of the physical world, from robot surgeons to delivery drones; conversely, physical objects are getting their own digital twins.

However, as the HiPEAC community is well aware, hardware has hit a serious roadblock. Transistor scaling and performance scaling are no longer in sync, which has led to a series of work-arounds in order to deliver smaller and smaller components while



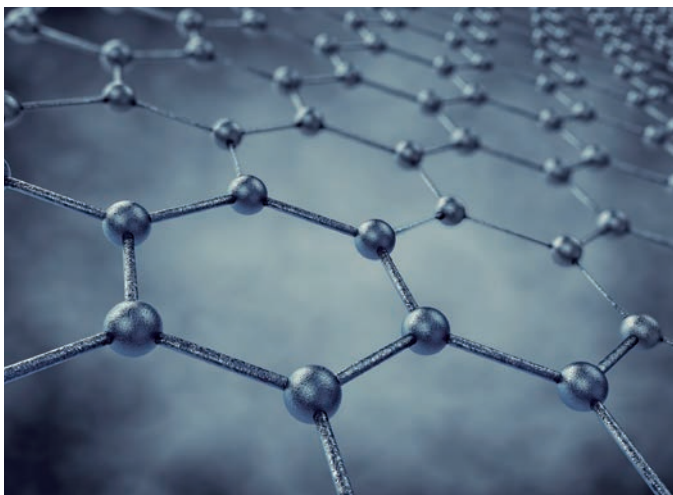
Photo © Klevo | Dreamstime.com



still increasing performance. Meanwhile, the ever-increasing complexity of computing systems and changing computing needs are creating serious problems for software, complicated by quality and legacy issues.

So where should we go from here? With no clear frontrunner in approaches at the moment, we should investigate a variety of options. Where technologies are concerned, these include everything from stacking chips in 3D to using new materials such as graphene or even synthetic DNA. Research should continue into novel architectures which respond to the needs of the data age, such as in-memory computing. The end of semiconductor scaling

Photo © Nobeastoffierce | Dreamstime.com



Future computing devices may be based on graphene or synthetic DNA



Photo © Thossaphol Somsri | Dreamstime.com

represents an opportunity for Europe to take the lead in radical new computing paradigms, such as quantum or neuromorphic.

One thing is for sure: whatever path we take, the solutions will need to be accepted by the general public. They will need to be safe, so that they do not cause us to be harmed, and secure, so that hackers can't break in. If they make a decision that concerns us, we should be able to find out why. The novel systems will be more long-lived than ever before, and the devices embedded in them should be designed to last, including their software. Last but not least, they will need to use less energy and fewer resources in general.

HiPEAC Vision 2019 recommendations

Accelerate, accelerate, specialize and automate

The only way to continue performance scaling in the short to medium term is to specialize hardware for important application domains, such as artificial intelligence and processing near memory. This specialization will require significant investment, and will only be economically viable if it is automated. Open-source hardware could boost innovative solutions. To facilitate the integration of accelerators into a system and to manage the increasing complexity, new intelligent tools and frameworks will be needed for both hardware and software.

Develop alternative architectures

Not only hardware technology but also architectures need to be revisited in light of the end of exponential scaling in computing power, and to improve energy and efficiency. Alternatives to the von Neumann architecture should be investigated, responding to the needs of modern computing, especially the need to process vast amounts of data. Taking inspiration from the example of neural networks, Europe should revisit innovative concepts from the past, which may have been made viable by new technology and production techniques. The new computing models can be applied to specific application areas for efficiency benefits. ▶

Build computers you can trust

With computers forming part of every aspect of our lives, any solutions developed must lead to trustable computing systems. They need to be secure – with watertight protections against malicious attacks – and safe, not harming people when they interact with their environment. This is particularly important for connected and cyber-physical systems. They also need to be reliable despite being increasingly complex, and here artificial intelligence could help by writing software and developing systems.

As ICT systems increasingly make decisions based on machine learning, the algorithms and the decisions they provide should be explainable enough to build trust.

Get looking for CMOS alternatives

The end of complementary-metal-oxide semiconductor – or CMOS – scaling means that all bets are on as to what technology will look like in 2030. With no one technology emerging as a clear front-runner, Europe should continue investing in research and help get results to market so that it will be at the heart of new technology developments. These technologies are unlikely to supplant CMOS, but instead will complement it.

Post-CMOS technologies might provide good solutions for the innovative sensor / actuator / interface technologies, which will play a crucial role in cyber-physical systems and wireless sensor networks.

Treat the computing infrastructure as a continuum, from the edge to the cloud

From microcontrollers with sensors and actuators, to concentrators, to micro-servers, to cloud and high-performance computing, computing is on a continuum, and self-contained systems are now themselves components of a large system. Interoperability is key; systems need to collaborate to give the best service to users. There is a need for dynamic devices, which can adapt intelligently.

Europe should encourage collaboration between different communities – such as software versus hardware, cloud and high-performance computing (HPC) versus the edge, to help break down silos – thereby making better use of resources, reducing energy consumption or latency as needed.

Shift value towards the edge

Europe needs to play to its strengths. That means building on strong industries like automotive, aerospace and trains, and electronic components and systems for embedded computing. Providing intelligence at the edge should be a major priority, aiming for a wide range of cognitive cyber-physical devices, and not necessarily always chasing the most advanced CMOS technology.

Investing in mature technologies (above 10nm) doesn't mean giving up on ambition. Interposer and chiplet technology will bring down costs and allow different technologies to slot together, such as analogue, power converters, memories, digital and photonics.

Lead on the use of collective data

Europe should develop the ethical use of state-owned, collective or domain data. This will allow the continent to develop its strength in AI-based solutions based on large amounts of data, without relying on the big business-to-consumer (B2C) technology companies. Solutions to ensure the privacy and security of data should be developed and enforced.

Become a leader in energy-efficient, sustainable electronics

Europe should become a leader in the design of sustainable electronics, the recycling of computing devices and modularity, prolonging the life of information and communication technology (ICT) systems. Innovative approaches should be developed to increase the longevity of electronic systems, through certification and virtualization, modularity, specific supervision, etc.

Conversely, computing should also be used to find solutions to the sustainability crisis facing the planet and bring the ecological footprint of Europe to within the continent's biocapacity.

Invest in the future workforce

The impact of computing on employment cannot be underestimated. Disappearing medium-skilled jobs will increase income inequality and may lead to social unrest. Europe should continue to invest in training programmes for workers at risk of losing their jobs, and try to reintegrate them in the job market at the highest level.

On the other hand, automation in AI may deliver the productivity growth the continent needs. In addition, Europe must invest in digital skills to maintain its innovation potential and remain competitive. Areas to be prioritized include machine learning, security, blockchain, architecture, system design and tools.

Develop a robust digital ethics framework

Computing has become such a powerful commodity that we should start thinking about whether everything that can be done should, in fact, be done. It is time to invest in digital ethics as a discipline to guide us to the future and to make sure that all computing professionals receive basic training in it. Digital ethics should also support policy makers to make decisions.

The HiPEAC Vision 2019 editorial board is as follows: Marc Duranton (editor-in-chief), Koen De Bosschere, Bart Coppens, Christian Gamrat, Madeleine Gray, Harm Munk, Emre Ozer, Tullio Vardanega and Olivier Zendra.

THE ROADMAPPING ACTIVITIES OF ETP4HPC



Marcin Ostasz, ETP4HPC / Barcelona Supercomputing Center (BSC)

ETP4HPC is the European High-Performance Computing (HPC) Technology Platform, an industry-led think tank whose main objective is the maintenance of a

European HPC Technology Roadmap. Based on this roadmap, the European Commission (EC) defines the priorities of its HPC technology and related research programmes: every project proposal submitted within those programmes should address the objectives identified by ETP4HPC.



ETP4HPC issued its first Strategic Research Agenda in 2013 and since then the SRA has been updated continuously, with the latest version published in 2017. Over the course of Horizon 2020, at any given time, there has been a valid SRA in place in order to allow the EC to build its programme and let projects define their scopes. The SRA is defined based on a model of HPC technology

development, which includes elements of the HPC system stack such architecture, system software, programming models, etc. To date, the ETP4HPC SRA has contributed to the implementation of almost 30 HPC technology projects with a total budget of €200 million. These projects have produced over 120 intellectual property items – tangible results that can be used to develop the European HPC ecosystem.

The next SRA will become part of the mechanisms of EuroHPC, an initiative of the EC and Member States to deliver HPC systems based on European technology. While the previous SRAs were developed by ETP4HPC taking into account the input of other stakeholders, the next issue will have to reflect the integration of various technologies: HPC, big data, artificial intelligence (AI) and the internet of things (IoT). This SRA will be written jointly with the Big Data Value Association (BDVA), the European big data community, whose representatives will be involved in the technological governance of EuroHPC.

“The next SRA will reflect the integration of high-performance computing, big data, artificial intelligence and the internet of things”

A lot of work in preparation of this SRA has already been completed. ETP4HPC and BDVA teams have been working together to identify commonalities in their technological approaches. Use cases (for example, the autonomous vehicle) have been identified to delineate the areas where the two technologies interact. Furthermore, ETP4HPC, BDVA and HiPEAC are working on their respective Visions for the next EC work programme; the next ETP4HPC SRA will stem from this HPC document. The completion of the next SRA is scheduled for the end of 2019 and the process of its writing will begin in January 2019.

The ETP4HPC and BDVA teams produced together a first white paper (see ‘Further reading’, below). This document is intended to feed the discussion on current strengths and differences between the (software and hardware) stacks of big-data computing and HPC, and how the current strengths of one stack may address a shortcoming/need in the other stack.

For more information on ETP4HPC’s work and collaborations, attend the joint workshop with HiPEAC at the HiPEAC conference in Valencia, on Tuesday, 22 January

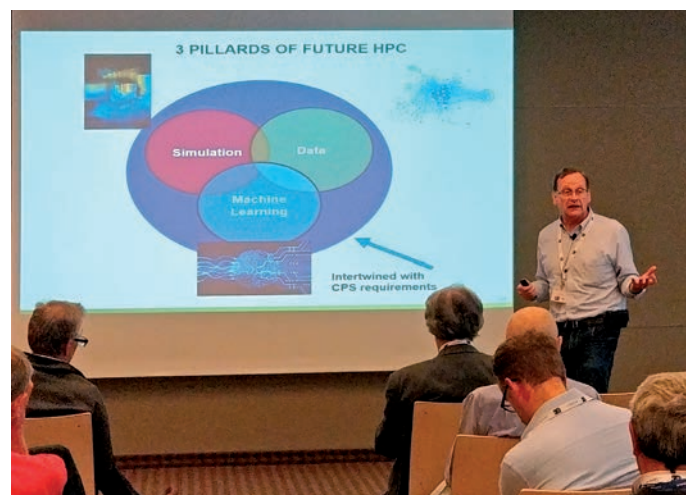
bit.ly/HiPEAC19_Vision_workshop

FURTHER READING:

ETP4HPC Strategic Research Agenda

etp4hpc.eu/sra

White paper: “The Technology Stacks of High Performance Computing and Big Data Computing: What they can learn from each other”
etp4hpc.eu/bigdata



ETP4HPC and HiPEAC have organized joint roadmapping workshops

MAPPING THE WAY AHEAD IN ELECTRONIC COMPONENTS AND SYSTEMS



March 2018 saw the publication of the first ever joint Electronics Components and Systems Strategic Research Agenda (ECS-SRA) between the European industry associations ARTEMIS-IA, AENEAS and EPoSS. This common framework for research, development and innovation (R+D+I) identifies key research priorities, technology challenges and application areas to drive Europe's transformation into a digital society while delivering societal and economic value.

The ECS-SRA incorporates the work of over 200 experts and feedback from the ECS community. A living document, the ECS-SRA is to be fine-tuned annually and subject to a major review every three years. In November 2018, the draft 2019 edition was released and feedback was provided during EFECs2018, the European Forum on Electronic Components and Systems.

Download the draft ECS-SRA 2019 here: bit.ly/ECS-SRA_2019_draft

PLATFORMS4CPS IDENTIFIES PRIORITIES FOR CYBER-PHYSICAL SYSTEMS

In 2018, Platforms4CPS released its final publication, *Platforms4CPS: Key Outcomes and Recommendations*, outlining key outcomes from the European Union (EU)-funded project and setting out recommendations for future research. The document aims to help Europe build on its strengths in the cyber-physical systems (CPS) domain by supporting the right technology areas, facilitating technology transfer, addressing societal concerns about new technologies and ensuring trustable systems.

The key needs identified for the future may be summarized as follows:

- Increase digital capacity and capability through Digital Innovation Hubs.
- Enhance multi-disciplinarity and cross-fertilization.
- Foster collaboration, European coordination and defragmentation across Europe.
- Support large-scale demonstrators in key areas such as autonomous driving.

- Tackle the confused landscape of business support for small and medium enterprises (SMEs).
- Explore CPS-enabled business models and business services, and facilitate access to these by SMEs.
- Help SMEs in allaying fears that are significant barriers to adoption, such as risks around cybersecurity.
- Encourage the development of common standards to connect different technologies.
- Establish a 'science of design for CPS'.
- Address the skills shortage and encourage systematic engagement between education and industry.
- Revitalize EU engineering education to provide T-shape (broad and deep) education.
- Ensure that European citizens can rely on European-supplied, trusted systems.

Download *Platforms4CPS: Key Outcomes and Recommendations* from the project website: platforms4cps.eu/resources

SURVEYING THE FUTURE OF HIGH-PERFORMANCE COMPUTING

The EU-funded coordination and support action Eurolab4HPC aims to strengthen excellence and innovation in academic research in high-performance computing (HPC). One of the project's main activities is its roadmapping document, the *Eurolab4HPC Long-Term Vision on High-Performance Computing*.



The first edition was released in August 2017; it analyses examples of 'application pull' – such as high-performance data analytics (HPDA) – and examples of 'technology push' in the short-, medium- and long-term, from 3D stacking to nanotubes. These are used to highlight key trends in high-performance computing, such as the convergence of HPC and HPDA, the impact of new non-volatile memories, programmability and green information and communication technology (ICT).

A new edition of the *Eurolab4HPC Long-Term Vision* is due for release in January 2020.

Further information is available on the Eurolab4HPC website:

eurolab4hpc.eu/vision

Tech talk 2030

What will technology look like in 2030? We asked experts from the HiPEAC community about the technological developments they would like to see over the next 10 or so years.



“ Everybody has a **personally trainable virtual assistant**, for work and personal tasks, that they can fully trust. No strings attached.
Monica Lam, Stanford University ”



“ **‘Brain-inspired AI!’**
Steve Furber, University of Manchester ”

“

A fundamental **breakthrough in battery technology**. Current batteries have an energy density which is 50 times lower than petrol. If they could store 5-10 times more energy, we would only have to charge our devices twice a week, electric cars could run for over 1,000km after being charged only once and houses could store electric energy for a week.
HiPEAC Coordinator Koen De Bosschere, Ghent University ”



“



By 2030, I hope to see **quantum computers** based on chips developed here at Barcelona Supercomputing Center being used all over the world.
Mateo Valero, Barcelona Supercomputing Center ”



“ Technology improvements in **accessible accelerator computing**.
As accelerator technology becomes increasingly pervasive as a solution to low-power, high-FLOP rate computing it is essential that the barrier to uptake is minimized, particularly for languages such as Python which is increasingly being used by science communities.
Toni Collis, Appentra Solutions / Women in HPC ”

“

By 2030 I hope that we have advanced **quantum technology** to a point where **every country** has the know-how and people to work on it.
Koen Bertels, Delft University of Technology / QuTech ”



“

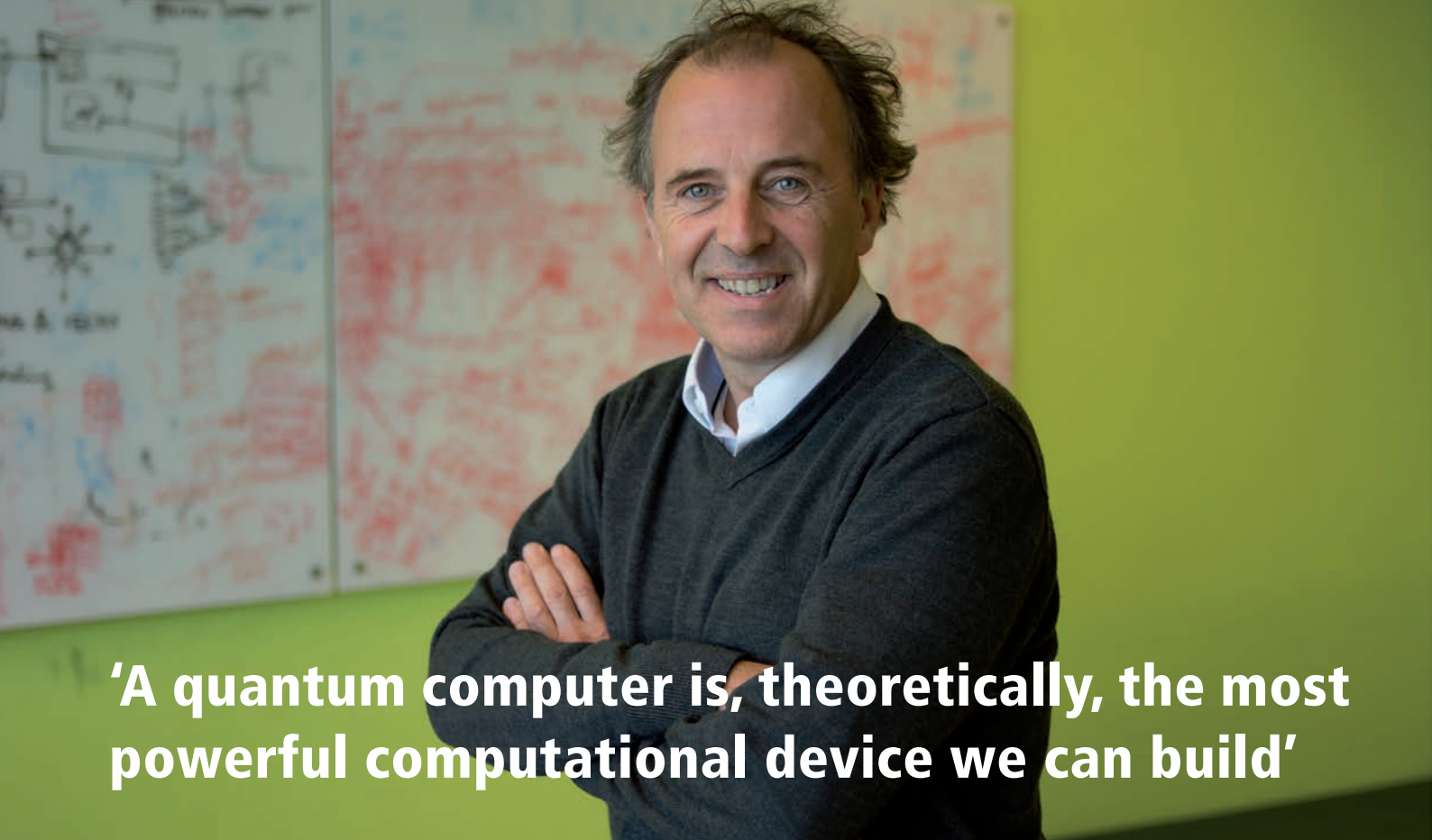
To deal with ever-growing complexity, the design and optimization process will rely more and more on **automated, portable and customizable workflows** with standardized ‘plug&play’ components. **Sharing and reusing** such workflows and components will become essential to accelerate innovation and technology transfer, while dramatically reducing costs and time-to-market.
Anton Lokhmotov, dividiti ”



“

One or more approaches – analogue or hybrid circuits, photonics, etc. – leading to **commercial neuromorphic chips** by 2030. Looking further into the future, perhaps by 2050 we could have large enough quantum accelerators and associated quantum algorithms to solve challenging problems for which we do not have enough computing power today.
José Duato, Universitat Politècnica de València ”





'A quantum computer is, theoretically, the most powerful computational device we can build'

Quantum computing has been heralded as one of the major new paradigms in computing; yet the weird and wonderful world of entanglement and superposition is a radical departure from contemporary computing systems. We spoke to HiPEAC19 keynote speaker Koen Bertels (Delft University of Technology / QuTech) to find out about his journey to quantum and how he sees this field developing in future.

How did you become a computer engineer?

Many years ago, a good friend of mine, Stamatis Vassiliadis, was a computing engineering professor in Delft. I was a younger associate professor in Wallonia, the French speaking part of Belgium. We did a couple of studies together and he invited me to become an assistant professor in Delft. That's how I got into the computer engineering domain.

Sadly, Stamatis died too soon and, through circumstances, I started assuming his responsibilities. I worked for 15 years on field-programmable gate array (FPGA) accelerators but got very annoyed by the fact that nobody seemed interested in that kind of technology.

Around five years ago, I moved over to quantum computing. After I'd been working on it for a year, what happened is that Intel decided to buy Altera, a FPGA company from the United States, for a very large amount, which then meant that the computing community began being interested in FPGAs. But at that time I lost the connection with that community, as I started working in quantum computing, which hardly anybody from the computer engineering world was interested in.

The switch to quantum computing entailed a lot of very hard studying the first years, happily assisted by Carmina Almudéver. We're now the core team of the Quantum Computer Architecture Lab and we divide the research between each other.

What are some of the highlights of your career so far?

With the support of colleagues in the computer engineering lab, I've graduated around 30 PhD students on FPGA and other topics. We've tried to foster a very international community in

the department and continued working with many of them even after students graduate.

It is also important to be aware that a core objective of the Quantum Computer Architecture Lab, known as QCA, is to develop international collaboration in the quantum computing project. The goal is really to make sure that the technology is available worldwide. For example, we try to establish a collaboration with graduate students once they've returned to their country of origin and help them set up a quantum research centre. This process is in the bootstrapping phase, so we do not yet have concrete results yet, but we are doing it right now with Pakistan and hopefully soon with some European universities. The idea is that TUDelft alumni find around two top students a year and then Pakistan can send them to Delft for a PhD in quantum computing. Once they graduate, we send them back to Pakistan and with our alumni they can work on developing a meaningful quantum accelerator.

In that context, we are currently working on the development of a quantum genome sequencing accelerator that aims to provide a tool that computes the DNA profile of every single individual on earth. So this is an open invitation to everyone to come up with other interesting ideas that are important for humankind and on which we can start collaborating.

Will quantum become the computing paradigm of the future? Should Europe invest heavily in quantum in preference to other technological directions?

From a purely performance point of view, a quantum computer is, theoretically, the most powerful computational device we can build. That is mainly due to the superposition and entanglement features. We should not forget that the field is only around 30 years old and a lot of things still have to be discovered and developed. Another thing to remember is that quantum phenomena are not digital but analogue, so that creates additional scientific challenges. The control, however, is – at least for now – a digital computer that controls all aspects of the quantum devices. That means we are currently building a quantum accelerator, similar to the FPGA or others that now are included in almost any computer.

Europe should realize that now is the right moment to start the development at all levels of what, at least in Delft, we think the layers are of any quantum device that we can build. We also need to develop the generation of scientists that will work on these topics. Topics range from quantum algorithms, quantum programming languages and compilers, runtime support, microarchitecture, cryogenic complementary metal-oxide-semiconductor (cryoCMOS) and, ultimately, the quantum chip.

So I do think that quantum is the computing paradigm of the future. We can reach compute performances which are substantially

higher than those we can achieve with any supercomputer currently available. It is not even known how much faster those computers will be, but some estimates suggest they could be more than 1,000 times faster than existing supercomputers.

What needs to happen before quantum becomes mainstream?

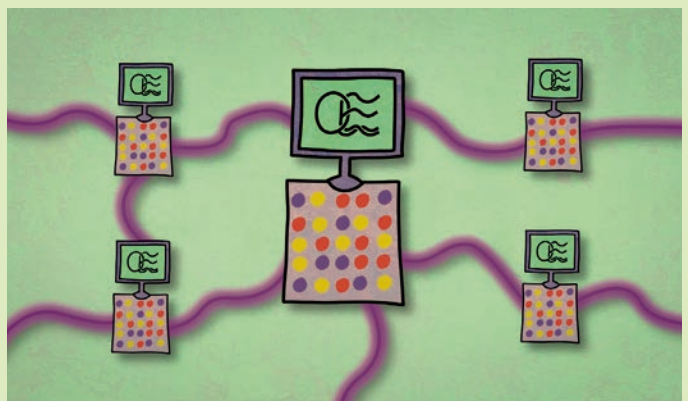
It is extremely important that there is an explicit or implicit decision about the underlying quantum technology that will be predominantly used to build such a machine. Currently, there are five main technologies in competition with each other:

- nitrogen-vacancy (N-V) centre, one of the imperfections in the structure of diamond, useful for its photoluminescence
- the Majorana fermion, a theoretical particle predicted by Ettore Majorana which would form the basis for topological quantum computers
- ion traps, a combination of electric or magnetic fields used to capture charged particles
- semiconductor qubits, using existing CMOS technology to implement quantum information processing
- superconductor qubits, an implementation of a quantum computer in superconducting electrical circuits

It's possible that multiple technologies will survive and have different properties. Delft is working on N-V centres to build a quantum internet. The superconducting qubit is the most dominant technology globally and our team is working very intensively with QuTech's Leo DiCarlo on making good progress here.

What is also very important is the stability of the qubit. In CMOS we are used to having around 10^{-15} error rates. In qubits, 10^{-3} error rates are considered a 'good' result at the moment, so there is a substantially higher error rate than CMOS. That implies that a lot of errors occur at any point in time and make the computation very unpredictable.

Being analogue, any quantum device also needs to compute the same algorithm multiple times and the outcomes of each run are assembled such that a statistical overview can be shown, allowing the human user to decide what the exact answer is of the quantum computer.



QUANTUM COLLECTIVE KNOWLEDGE



As Roberto Viola, Director General of the European Commission's DG CONNECT recently noted, the history of quantum physics and mechanics is well-rooted in Europe. However, when it comes to developing a quantum computer, the continent faces stiff competition both from vertical companies and government-funded initiatives elsewhere in the world.

'As most of us know, quantum computers have the potential to solve certain problems dramatically faster than conventional computers, with applications in areas such as machine learning, drug discovery, materials, optimization, finance and cryptography,' explains Anton Lokhmotov (a co-founder of dividiti and a principal contributor to Collective Knowledge). 'That's one good reason why we need to start preparing now.'

But there's another, more personal reason why researchers should get involved in quantum, says Anton. 'Quantum computing is in its very early days. Imagine you are living in the 1950s, when many things we take for granted today simply did not exist: hardware is unreliable, programming languages are machine-specific, compilers are non-optimizing, and so on. With so many challenges, many opportunities and discoveries also lie ahead. If you are entering this field today, chances are you may become one of its heroes tomorrow. Fancy being quantum's Maurice Wilkes? Niklaus Wirth? Edsger Dijkstra?

A quantum problem shared

To advance the field, knowledge sharing is crucial, Anton notes. 'First, it is still not a given that we will ever achieve so called "quantum supremacy" where we convincingly demonstrate that a quantum computer can outperform a conventional one; for example, simulate precisely the behaviour of a complex chemical compound. With only a few thousand people working in this field worldwide, we need to maximize every opportunity to achieve significant breakthroughs.'

Another reason for collaborating, says Anton, is that 'practical and scalable quantum computing will be the result of a multidisciplinary collaboration: between physicists devising ways to implement robust qubits; computer engineers creating actual devices and optimizing their performance, reliability and costs; algorithm designers mapping problems onto a mix of quantum and classical resources; and so on'.



This is where the community-driven initiative Quantum Collective Knowledge comes in. 'Designed to pinpoint the state-of-the-art and forecast future developments in this field, Quantum Collective Knowledge (QCK) builds upon Collective Knowledge (CK), our universal open framework for reproducible and collaborative research and development (R+D),' explains Anton.

QCK uses unified workflows with reusable components which can run on classical and quantum platforms, can be extended by the community, and are connected to a public dashboard, according to Anton. 'We believe that QCK will be instrumental in unlocking the power of quantum computing for everyone.'

The initiative is seeking realistic, useful and diverse workloads and has run two QCK hackathons so far, with more planned for 2019. 'Based on these highly successful events, we are launching an Open VQE Challenge, where participants will be able to crowdsource optimal solutions to a quantum chemistry problem on a quantum simulator or real quantum device from the comfort of their homes (or labs),' adds Anton.

The team is also discussing how to scale up their activities (to cover more algorithms, applications, frameworks, platforms and so on) and to share research artifacts and workflows (such as code and data) in a common portable, interoperable and customizable format. 'Our ultimate goal is to accelerate all aspects of R+D, education and technology transfer to build scalable quantum technologies,' Anton concludes.

FURTHER READING:

Quantum Collective Knowledge

cknowledge.org/quantum

Collective Knowledge

cknowledge.org

Reusable quantum components

github.com/ctuning/ck-quantum

CK dashboard

cknowledge.org/dashboard



BEYOND DIGITAL: BRAIN-PICKING FOR NEW COMPUTING MODELS

Image credit: © Sergey Khakimullin | Dreamstime.com



There may still be a long way to go before we get to a fully digitalized society, but there are compelling reasons to ‘start looking beyond digital and develop the technologies that will follow’, argues Dr. Wim JC Melis, senior lecturer at the University of Greenwich.

‘If you look at the history of our current technology, you’ll notice that digital was developed as a solution for long-distance communication,’ says Wim. ‘This led to the development of gates for logic and mathematical functionality, and that turned into our current, mainly von Neumann architectures – a great achievement which will continue to be useful for quite a bit into the future.’ The digital abstraction has also manifested itself in a number of technological implementations, notes Wim, ‘from electro-mechanical switches to transistors, with new developments looking at optical and quantum implementations’.

That said, the end of Moore’s Law and a changing application landscape are showing up some of the limitations of digital. ‘Hence, one needs to wonder whether our current platforms deliver optimal efficiency and performance for all current and future applications,’ says Wim, ‘which leads to the conclusion that large datasets are not what these architectures were designed for. As for artificial intelligence (AI), the abilities and accuracy achievable by digital have been limited in comparison to humans.’

So where do we look for alternatives? ‘Computer scientists have taken inspiration from biology, with developments such as neural network models and networks modelling the electrical aspects of the human brain,’ says Wim. ‘However, it’s important to keep in mind that there we have limited knowledge of how real neurons

“How can an analogue system such as the brain be so tolerant to noise and component failure?”

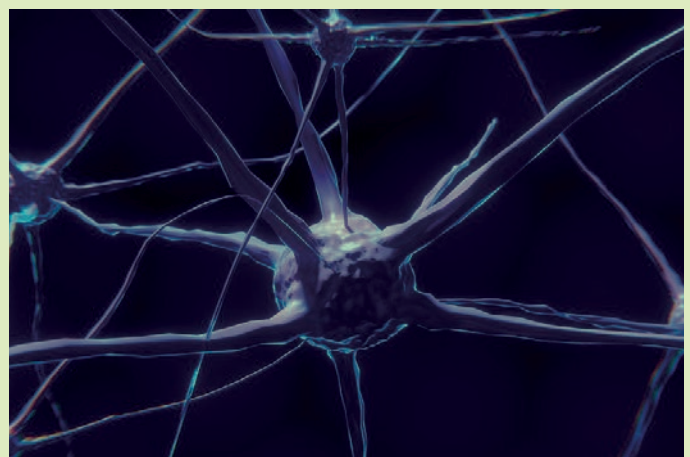
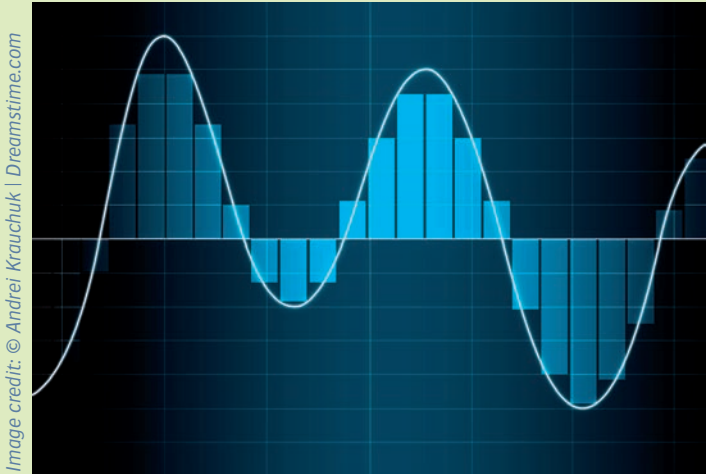


Photo © Colin Behrens on pixabay



operate, and also that their operation is not purely electrical, with chemicals, for example, also playing a part. Clearly, this limited knowledge influences the creation of their artificial counterparts.’

What we can be fairly sure about, says Wim, is that our brains do not operate in a digital manner. ‘The brain seems to manage a large continuous data stream from our sensors and operate very reliably, while showing flexibility and resilience.’ The way forward, he suggests, is to ‘understand the brain from a functional level and use that to create a new abstraction, rather than await a full understanding of the lowest levels’. ‘If one does look at the brain from this functional perspective, then theories from hierarchical temporal memory (HTM) to the simplicity and power (SP) theory of intelligence lead to the same conclusion, namely that similarity matching and/or information association are key to overall operation.’

One of the main reasons why digital systems replaced analogue ones, notes Wim, is how they deal with noise. How, then, can an analogue system such as the brain be so tolerant to noise and component failure? The answer lies in two concepts, he says. ‘First, it is a non-deterministic system that “just” needs to identify similarity and rank the results. Second, there is a massive amount of feedback within the system that combines with feed forward and corrects information where necessary. So as a system it would learn the patterns that need to be recognized, and then recognize those as they appear.’

“The brain seems to manage a large continuous data stream from our sensors and operate very reliably”

Wim points out that the brain mostly likely learns these patterns from sensor information, which is kept in analogue format, as analogue has a much higher data density than digital. ‘This higher data density aids the processing, storage and communication, which is exactly where our digital systems are struggling, since they have the worst possible data density.’

As a new abstraction, says Wim, creating a similarity model is achievable and can be implemented on technologies including those already being used or investigated for digital systems, such as semiconductor, optical and quantum. ‘However, the implementation platform does not affect the data density, and taking into account that any computing platform has a tight relationship between data representation, the way that information is processed and the implementation on the platform, selecting the correct data representation is key to creating a new system.’

There is also a direct link between the data representation / format and the applications for which a platform is particularly good, he points out. ‘For example, digital platforms are natural winners for logic / control / mathematical / deterministic operations, while anything else requires substantial algorithm development to “map back” towards this framework, resulting in inefficiencies and other compromises.’

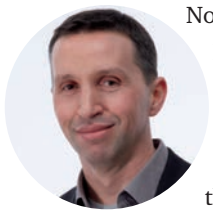
So when will these new platforms become available? ‘Designing a basic “cell” that does similarity matching in a non-mathematical way has already been achieved,’ says Wim, ‘but these cells need to be brought together into an actual architecture that can deal with larger datasets, learn information and so on.’ This will require input from the machine learning community in addition to the architectural work, he notes. ‘Doing all this well, could create a platform that delivers human-like intelligence, but how this links with consciousness and the subconscious is just one of the topics which is still unclear.’

One thing that is clear, says Wim, and that is that ‘using massive feedback makes the system fully reversible, which brings significant energy savings’. However, he points out that it will take time before such an architecture becomes commercially available, and when it eventually does it may become a co-processor to current processors, rather than a standalone platform. ‘Nevertheless, we need to start developing these alternatives as soon as possible, and that will require significant capital and human investment.’

Noise pollution represents a genuine health hazard. Here, Sabri Pllana (Linnaeus University) explains how his team have developed an effective machine learning solution to help combat this problem.

Sound effects

A machine learning-driven IoT solution for noise classification



Noise is any undesired environmental sound.

Recent studies have found that exposure to noise pollution may increase the risk for health issues, such as heart attack, obesity, impaired sleep or depression. Following the Environmental Noise Directive (END) 2002/49/EC, every European Union (EU) member state must assess environmental noise and develop noise maps every five years.

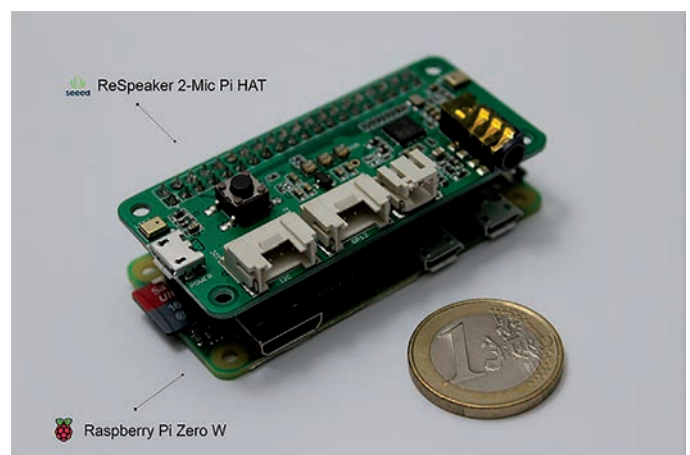
As sources of noise, including traffic, construction sites, music and sport events, may change over time, there is a need for continuous monitoring of noise. Health-damaging noise often occurs for only few minutes or hours, and it is not enough to measure the noise level every five years. Furthermore, sounds at the same decibel (dB) level may be perceived either as annoying noise or as a pleasant music. Therefore, it is necessary to go beyond the state-of-the-art approaches that measure only the dB level and actually identify the type of the noise. The internet of things (IoT) is a promising technology for improving many domains, such as e-health, and it may be also used to address the issue of noise pollution.

We have developed an approach for noise classification in smart cities using machine learning on a low-power and inexpensive

IoT unit that comprises a Raspberry Pi Zero W and a ReSpeaker 2-Mic Pi HAT (Figure 1). The Raspberry Pi Zero W is the size of a credit card and costs only about €10; it may be used as an affordable computer to learn programming or to build smart devices, such as our solution for noise classification. We use ReSpeaker 2-Mic Pi HAT, a dual-mic array expansion board for Raspberry Pi, for sound sensing. Mel-frequency cepstral coefficients (MFCC) are extracted as audio features and applied to two classifiers: support vector machine (SVM) and k-nearest neighbours (KNN). We have implemented all algorithms used for our solution of the noise classification problem in Python.

To investigate the performance of the system, we conducted experiments with eight different classes of environmental sounds: quietness, silence, car horn, children playing, gunshot, jackhammer, siren, and street music. For the training of the system we selected noise-relevant environmental sound clips from popular sound datasets, such as UrbanSound8K and Sound Events. Our dataset contains more than 3,000 sound excerpts with lengths of up to four seconds. We observed in our experiments with various environmental sounds (such as car horns, jackhammer, or street music) that our solution provides high noise classification accuracy in the range of 85%-100%. Future work will investigate the usefulness of our solution for a large number of Raspberry Pi devices.

“Exposure to noise pollution may increase the risk for health issues, such as heart attack, obesity, impaired sleep or depression”



The proposed noise classification hardware platform consists of a Raspberry Pi Zero W and a ReSpeaker 2-Mic Pi Hat

Smarter systems thanks to self-aware chips

Shining a light on hidden inefficiencies, security risks and energy consumption, UltraSoC's embedded analytics solutions enable self-aware chips and provide a powerful answer to the complexity of modern computing systems. Here, UltraSoC Chief Executive Officer Rupert Baines explains how the company's innovative technology helps cut development costs and improve performance by monitoring 'from within'.

“Having an intelligent analytics infrastructure embedded into the hardware of an SoC gives you visibility of the real-world behaviour of the entire system”

What are 'self-aware' or 'self-monitoring' chips?

Modern electronic products are complex and powerful. A smartphone today delivers the processing power of a supercomputer from the 1980s. The average internet search will draw on computers around the world – often around 2,000 'nodes' are involved in answering your request to find your nearest Starbucks.

The heart of these systems is the system-on-chip (SoC), and the makers of these chips are constantly under pressure to deliver extra performance and speed while of course maintaining reliability, security and other differentiating performance criteria. One approach to solving these problems is to put monitoring and analytics hardware into the chip itself. Such a monitoring infrastructure can be completely independent of the 'main system' itself – it monitors and observes – but it can also be used to make the system itself 'aware' of interesting events and behaviours. With these capabilities embedded into the SoC, the system can be self-aware: 'environmentally aware' (what is happening in the rest of the SoC and wider system); and also able to raise alerts, so that the system itself can take action. UltraSoC provides intellectual property (IP) to embed these self-analytical capabilities into an SoC.

What are the benefits of self-aware chips?

Having an intelligent analytics infrastructure embedded into the hardware of an SoC gives you intimate visibility of the real-world behaviour of the entire system. The ultimate benefits therefore can include improved resistance against malicious intrusions; enhanced product safety (identifying anomalous and potentially dangerous behaviour); reduced system power consumption; and overall better performance. Just as importantly, it's possible to fine-tune the final products, even when they are in the field.

A great example is optimizing the performance of high-performance computing (HPC) platforms. Optimizing the performance of these systems translates directly into dollars. In e-commerce, every microsecond of user latency reduces ad clicks and makes it less likely that the user will complete a transaction. The size of HPC systems means that scaling and performance optimization

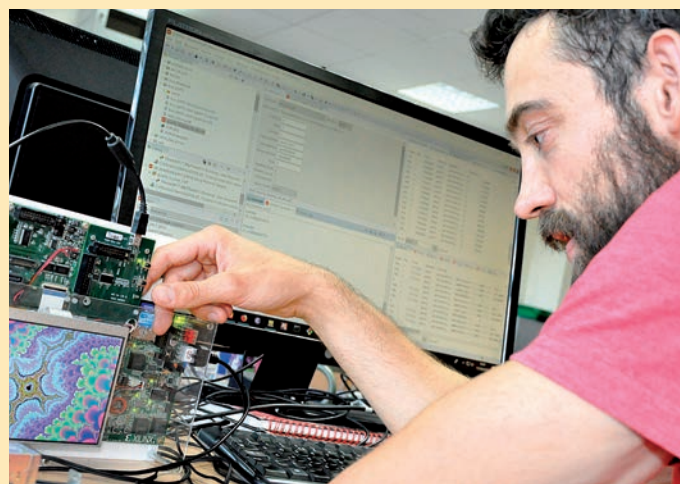
has a significant impact on capital cost: the classic example is the non-fatal bug that caused 25% of Google's entire disk fleet to run significantly more slowly than expected for three years. Fixing that problem – which was not sufficiently serious to stop the system working, but subtly undermined its performance – improved the performance of 25% of the company's servers – literally millions of compute nodes.

'Self-aware' chips provide a view into the hardware as it functions in real operating conditions, and that allows engineers to implement firmware or software changes to fix these 'long-tail' bugs much more quickly and easily.

Could self-aware chips provide improvements in other areas (security or energy efficiency, for example)?

Absolutely; it's possible for networks of servers to be consuming far more energy than they were meant to... without anyone knowing. The behaviour of the SoC impacts a broad range of performance criteria: compute power, power consumption, reliability, safety and security – all of these can be significantly enhanced by having this intelligence built in; ensuring that all parts of the SoC and wider system are doing exactly what they're meant to be doing.

Another great example of the benefits of this embedded analytics approach can be seen in functional safety. Intelligent self-monitoring systems make it much easier to implement redundant and lockstep architectures: a pair of duplicate processor cores runs the same operations, side-by-side. The embedded



UltraSoC creates IP for self-aware SoCs



Rupert Baines, UltraSoC's chief executive officer

intelligence watches over them to check they continue to respond in exactly the same way. If they are not aligned, an alert is raised.

What work is UltraSoC doing in this area?

UltraSoC is already enabling customers to build self-aware chips and to embed this type of intelligence within the heart of their SoCs – in fact our architecture is already deployed in end products from a leading disk drive manufacturer, and has been chosen by semiconductor companies including HiSilicon (Huawei), Intel, Microchip Technologies, Alibaba group company C-SKY, and leading IT systems integrator Kraftway.

The company's embedded analytics products are based on silicon IP and software, and provide intimate visibility and analysis of the operation of the hardware and software in any electronic system. UltraSoC's technology is processor agnostic, which makes it applicable to designs using any central processing unit (CPU) architecture, including Arm and the developing open-source RISC-V platform.

The embedded analytics architecture assists not only in improving end product performance and security, but also significantly benefits the design and development process: reducing risk, cutting time to market and slashing costs of design and development.

Given the growing need for improvements in performance and in applications controlling security and safety, we fully expect the applications to grow significantly from 2019 - hence we will see far more end products designed using this type of intelligent analytics and monitoring. ►

UltraSoC in brief

COMPANY: UltraSoC

MAIN BUSINESS: intellectual property and supporting software for self-analytic systems-on-chip

LOCATION: Cambridge, United Kingdom (UK)

WEBSITE: ultrasoc.com



UltraSoC provides semiconductor intellectual property and supporting software tools that allow development teams to incorporate intelligent self-analytic capabilities in the systems-on-chip (SoCs) at the heart of today's computing, communications and consumer electronics products. The company's embedded analytics technology helps solve the most pressing problems faced by the high-tech industries today – including cybersecurity, functional safety, and the management of complexity. Its solutions also allow designers to develop SoCs – the driving force behind both performance improvement and cost reduction in the electronics business – more quickly and cost-effectively.

The company's flagship product line is a suite of semiconductor IP that allows chip designers to integrate an intelligent analytics infrastructure into the core hardware of their devices. By monitoring and analysing the real-world behaviour of entire systems, engineers can take action to reduce system power consumption, increase performance, protect against malicious intrusions, and ensure product safety via UltraSoC's intelligent analytics embedded in the silicon. These capabilities address applications in a broad range of market sectors, from automotive and internet of things (IoT) products, to at-scale computing and communications infrastructure.



UltraSoC's IP is particularly valuable when used in SoCs designed for high-performance computing (HPC) applications. Embedded hardware-based analytics capabilities allow the non-intrusive collection of actionable insights on the real-world behaviour of entire systems. Such a hardware-based approach is uniquely capable of helping software and hardware teams to detect performance-impacting long-tail bugs and hard-to-identify issues in HPC systems – for example affinity management policies, contention and cache coherence.

UltraSoC's products are used by leading names including HiSilicon (Huawei), Intel, Microchip Technologies, Alibaba group company C-SKY, and leading information technology (IT) systems integrator Kraftway. Licensees have already taped-out devices at 40nm, 28nm, 16nm and 7nm process nodes. The company's partners include Andes, Arm, Cadence/Tensilica, CEVA, Esperanto, Imperas, Lauterbach, Mentor, MIPS, Moortec, Percepio, Segger, SiFive, Sondrel, and Teledyne LeCroy.

UltraSoC's technology also has a substantial impact on the increasingly-pressurized economics of the semiconductor industry. Traditional SoC development methodologies have failed to keep pace with escalating systemic complexity, creating a 'productivity gap'. UltraSoC's intelligent analytics close that gap, giving development teams actionable insights that shorten the total development cycle time, accelerate debug, and reduce risk and cost to ensure timely market entry. Analysis from SemiCo research demonstrates the bottom-line value of this approach: SoC design teams can double their profitability and reduce their development costs by a quarter by using UltraSoC.

UltraSoC is an active member of the RISC-V consortium, which aims to encourage the use of a new open source processor architecture which has been dubbed 'the Linux of semiconductors'.

Originally a spin-out from Essex University, the company was seed-funded in 2011, and now with £14 million funding, with investors including Indaco Venture Partners, Enso Ventures, Octopus and Oxford Capital. Headquartered in Cambridge, UK, UltraSoC today employs 30 people mostly in Cambridge and Bristol focused largely on innovation and product development. UltraSoC's global presence supports customers in the European Union, United States, China, Korea, Russia.

The company was named one of the 100 most exciting companies in the UK in the 2016 Mishcon de Reya / CityAM 'Leap 100' list, and by Gartner as one of its 2016 'Cool Vendors'. It was recognized as 'Best New Company' at the 2015 ELEKTRA Awards.

Find out how Europe is scaling up to the challenges of the data era, delivering powerful machine vision in small packages, managing hardware complexity and creating energy-efficient yet high-performance systems.

Innovation Europe

THE FUTURE OF MODULAR MICROSERVER TECHNOLOGY

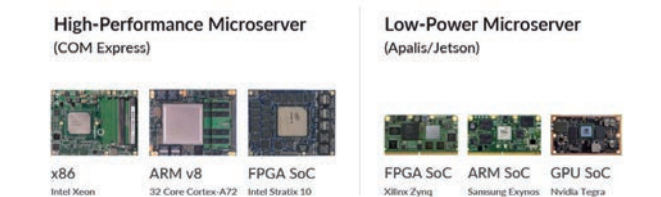
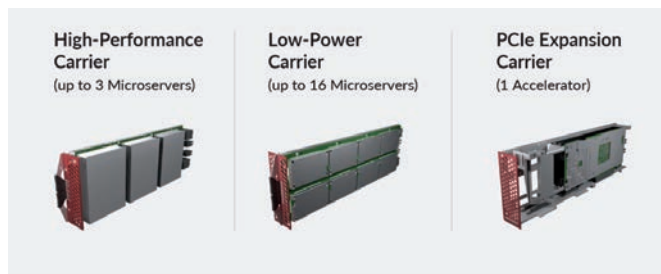


A new class of low-power appliances, with optimized total cost of ownership (TCO), built-in efficiency and dependability, has been developed by the European Union-funded project M2DC – Modular Microserver DataCentre. Thanks to a broad ecosystem of management software, the developed platform is easy to integrate, while being fully software defined, enabling cost-effective optimization for a variety of future, demanding applications.

New technologies including advanced mobile devices, the internet of things (IoT), 5G and machine learning are posing steadily increasing demands on the performance and energy efficiency of server platforms and data centres. Homogenous, commodity architectures will not be sufficient to meet the compute requirements of the next generation of information technology (IT) workloads. In response, heterogeneous hyper-scale data centres target these challenges with a combination of highly scalable server platforms and integrated hardware accelerators, based, for example, on graphics processing units (GPUs) and reconfigurable hardware.



Against this background, M2DC harnessed the expertise of 13 companies, research institutes and universities to investigate, develop and demonstrate a modular, highly efficient, cost-optimized server architecture composed of heterogeneous microserver computing resources. The architecture can be tailored to meet requirements from various application domains, such as image or signal processing, cloud computing, big data analytics, machine learning or even high-performance computing (HPC).



Advanced management strategies and system efficiency enhancements ensure high levels of energy efficiency, performance, security and reliability. Thanks to the M2DC middleware, the server's underlying heterogeneity – including new microserver designs based on Arm v8 64-bit processors and Intel Stratix 10 – is abstracted to a level suitable for data centres. The system design includes optimized middleware for the deployment of appliances.

M2DC customers can acquire: (i) a base appliance that can be tuned to a specific problem; (ii) a preconfigured appliance to the rising technological domains of IoT data analytics, HPC, cloud computing, and digital signal processing; or (iii) a custom

appliance for their business or research needs, and benefit from the full support of the M2DC team of experts. The platform chassis was exhibited at major international events such as ISC High Performance, Teratec and SC18, where visitors were very enthusiastic about its PCIe-based switching technology, middleware concept, integration of field-programmable gate arrays (FPGAs) for HPC, and the heterogeneous nature of the server.

To find out more, visit the M2DC website or follow the project on social media.

m2dc.eu [@M2DC_Project](https://twitter.com/M2DC_Project) facebook.com/M2DCproject
linkedin.com/company/m2dc [YouTube bit.ly/M2DC_YouTube](https://youtube.com/bit.ly/M2DC_YouTube)

NAME: M2DC: Modular Microserver DataCentre

START/END DATE: 01/01/2016 – 30/06/2019

KEYWORDS: microserver, data centre, high-performance computing (HPC), heterogeneous architectures, modular, scalable

COORDINATOR: Ariel Oleksiak, Poznań Supercomputing and Networking Center (PSNC)

PARTNERS: Poland: Poznań Supercomputing and Networking Center (PSNC); France: Commissariat à l'énergie atomique et aux énergies alternatives (CEA), REFLEX CES, Alliance Services Plus; Germany: Christmann Informationstechnik + Medien, Huawei Technologies Düsseldorf, OFFIS, CEWE; Italy: Politecnico di Milano; Slovenia: XLAB; Switzerland: Vodafone Automotive Telematics; UK: Arm

BUDGET: approx. €8 million

M2DC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 688201.

TULIPP: HIGH-PERFORMANCE IMAGE PROCESSING FOR EMBEDDED COMPUTERS



Question: What do advance driver assistance systems (ADAS), drones and medical X-ray imaging have in common?

Answer: They all need high-performance image processing and have tight size,

weight and power (SWaP) requirements. Thanks to TULIPP, there are powerful new tools to help them achieve this.

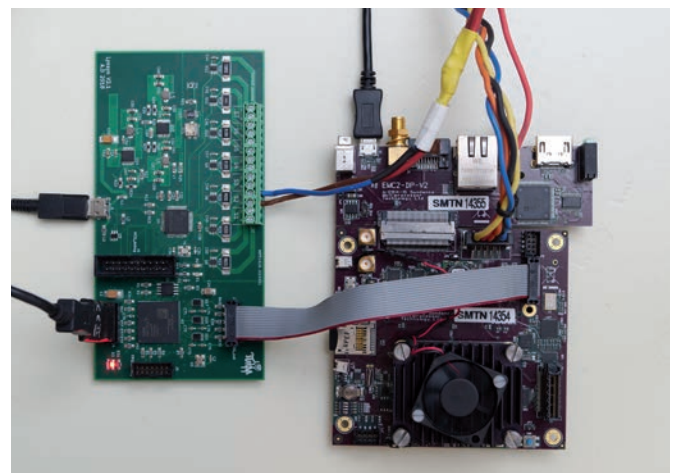
Coordinated by Thales, the three-year TULIPP project has designed a common reference processing platform – hardware, operating system and programming environment – capturing the real-time requirements, high-performance image processing and vision applications common to these three application domains. The project has also developed a set of guidelines to help select combinations of computing and communication resources to be instantiated in the platform while minimizing energy use and reducing both development costs and time-to-market.

TULIPP uses an abstraction known as the generic development process (GDP) to capture image processing application development. The input to this is a functionally correct implementation of the image processing system that runs on a desktop computer, before the application is moved to the embedded platform. GDP is an iterative process where each iteration is designed to bring the image processing system closer to system requirements. TULIPP aimed to minimize the number of iterations through the GDP and create a framework that efficiently supports developers carrying out the GDP.

To future-proof the reference platform, TULIPP has documented the design process in a book which will form the basis for future standardization efforts. The project has also created a development kit consisting of:

- a TULIPP hardware platform developed by Sundance, based on the Xilinx Zynq multi-processor systems on chip (MPSoCs) Ultrascale+™
- the TULIPP operating system, based on the HipperOS multi-core operating system
- STHEM, the TULIPP toolchain, which was designed by NTNU, the Norwegian University of Science and Technology, and TU Dresden

This development kit is available as a bundle and is available for purchase from Sundance.



It has been applied to three compelling use cases:

- **Medical use case** focusing on mobile C-arms, a medical system that displays X-ray views from inside a patient’s body, greatly enhancing the surgeon’s ability to perform surgery. By bringing the compute power of a personal computer to the hardware the size of a smartphone, TULIPP makes it possible to achieve the same image quality with a quarter of the radiation.
- **Unmanned aerial vehicles (UAVs):** TULIPP developed a real-time obstacle avoidance system for UAVs, also known as drones, based on a stereo camera. Disparity maps from camera images are used to locate obstacles in the flight path and help steer the UAV around them.
- **Advanced driver assistance systems:** TULIPP worked on ways to improve pedestrian recognition in real time. Modern image processing algorithms are capable of extracting almost all the information required about the vehicle’s surroundings, but there is often a big step between the prototypical implementation on a desktop PC and the integration into a small, energy-efficient electronic control unit (ECU). TULIPP’s optimizations, using FPGA logic in combination with Arm cores, resulted in processing time of 66ms per frame.

As it reaches its close, TULIPP is working on bringing more intelligence to the platform by adding new accelerated technology dedicated to artificial neural networks.

NAME: TULIPP: Towards Ubiquitous Low-power Image Processing Platforms

START/END DATE: 01/02/2016 – 31/01/2019

KEYWORDS: microserver, data centre, high-performance computing (HPC), heterogeneous architectures, modular, scalable

COORDINATOR: Philippe Millet, Thales

PARTNERS: Belgium: HipperOS; France: Thales, Efficient Innovation; Germany: Ruhr-Universität Bochum, Technische Universität Dresden, Fraunhofer-Gesellschaft; Norway: Norges Teknisk-Naturvitenskapelige Universitet (NTNU); Sweden: Synective Labs; UK: Sundance Multiprocessor Technology.

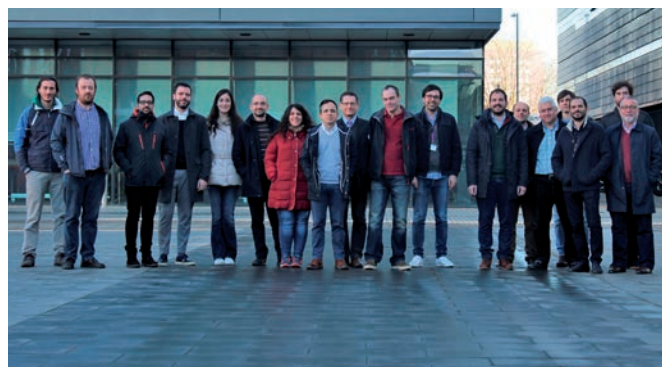
BUDGET: €4.73 million

TULIPP has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no 688403.

SOFTWARE FOR THE BIG DATA ERA WITH E2DATA



To address pressing scalability concerns, both big data users and cloud infrastructure providers such as Google, Microsoft, Amazon and Alibaba are investing in heterogeneous hardware resources. Combining a variety of architectures including central processing units (CPUs), graphics processing units (GPUs), field-programmable gate arrays (FPGAs) and manycore processors, the aim is to increase performance while minimizing climbing operational costs. In parallel, large companies are developing their own in-house application-specific integrated circuits (ASICs), with Google’s TensorFlow being the prime example.



E2Data is a new European project that will provide a new big data software paradigm to deliver maximum resource utilization for heterogeneous cloud deployments without requiring changes to the original source code. The proposed solution takes a cross-layer approach, allowing vertical communication between the four key layers of big data deployments (application, big data software, scheduler / cloud provider and execution runtime). E2Data dynamically exploits heterogeneous hardware – in this case GPUs and FPGAs – by:

- enabling dynamic heterogeneous compilation of arbitrary code
- following a full-stack vertical approach where state-of-the-art software frameworks will be enhanced
- designing an intelligent, elastic system where we can:
 - profile results, communicate to task and data schedulers and assess decisions to improve performance and energy efficiency
 - fall back and recompile on the fly to handle possible failures

To achieve its ambitious goals, the E2Data consortium brings together two different kinds of big data practitioner. On the one hand, big data users from the finance, healthcare, green building

and security domains – all with strict requirements regarding performance and infrastructure costs – are represented. On the other hand, four big data technology providers will implement the E2Data solution by extending cutting-edge European technologies.

NAME: E2Data: European Extreme Performing Big Data Stacks

START/END DATE: 01/01/2018-31/12/2020

KEY THEMES: heterogeneous computing, elastic resource provisioning, big data applications, data stream analysis, scalability

PARTNERS: UK: Exus Software (coordinator), University of Manchester

(technical leader), Kaleao, Spark Works, iProov; Greece: Institute of Communication And Computer Systems (ICCS), Institutouto Technologias Ypologistonkai Ekdoseon Diofantos (CTI Diophantus); Germany: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI); Luxembourg: Neurocom Luxembourg

BUDGET: €4.67 million

WEBSITE: e2data.eu

E2Data has received funding from the European Union's Horizon H2020 research and innovation programme under grant agreement no. 780245

A RECIPE FOR HPC SUCCESS

RECIPE

The high-performance computing (HPC) market is developing quickly, pushed by new application domains (i.e. computationally intensive data analytics), by new applications of massively parallel computations and by the increased ability of new customers to enter the market. In this scenario, it is necessary to develop middleware capable of making the system reliable despite the increasing number of resources and the decreasing mean time between failures.

To manage this complexity, the European Union (EU) Horizon 2020 project RECIPE will provide:

- A **hierarchical runtime resource management infrastructure** able to optimize energy efficiency and to minimize the occurrence of thermal hotspots. This infrastructure will also enforce the time constraints imposed by the application, ensuring reliability for both time-critical and throughput-oriented computations.
- A **predictive reliability methodology to support quality of service (QoS)** in face of both transient and long-term hardware failures.
- A **set of integration layers** to allow the resource manager to interact with both the application and the underlying deeply heterogeneous architecture.

RECIPE has three main goals:

1. To increase the energy efficiency of HPC systems by 25%, with an improvement of 15% of mean time to failure.
2. To improve the energy-delay product by up to 25%.
3. To reduce the occurrence of fault executions by 20% with recovery times compatible to real-time performance and full exploitation of available resources under non-saturated conditions.

The RECIPE project will assess its results against real world use cases, addressing key application domains:

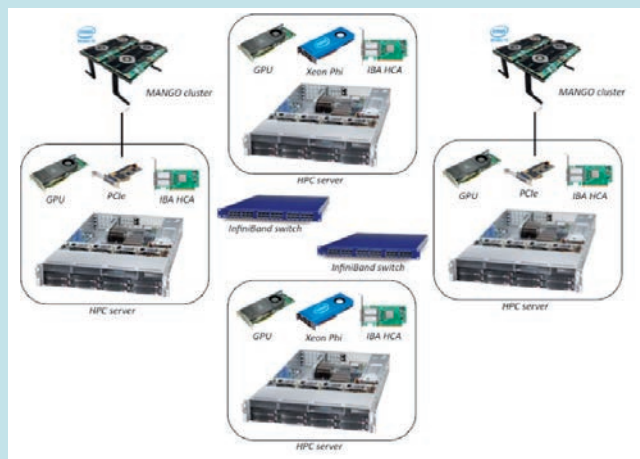
Geophysical exploration: thanks to the efficient implementation of the RTRM, the resulting Full Waveform Inversion tool reduces the uncertainty of current seismic exploration surveys;

Environmental monitoring and meteorology: the runtime resource management structure developed will improve the ability to keep the status of water basins under control and the behaviour of power plants exploiting renewable energy sources such as wind turbines.

Bio-medical machine learning and big data analytics: the software infrastructure developed will enable the deployment of epileptic seizure detection algorithms in a prototype platform able to manage a large-scale population while meeting the real-time requirements of the application.

Hardware heterogeneity in RECIPE

Scaled systems, including all types of heterogeneous accelerators, are interconnected in a disaggregated fashion through an HPC fabric. The host-side part of the prototype cluster includes commercial CPU+manycore hardware (e.g., Intel Xeon and Intel MIC/Phi). Disaggregated resources include GPU-rich server nodes, the Mango Communications FPGA cluster augmented with support for network protocol processing, and next-generation advanced interfaces demonstrating the concept of network-connected RISC many-cores.



NAME: RECIPE: REliable power and time-Constraints-aware Predictive management of heterogeneous Exascale systems

START/END DATE: 01/05/2018-30/04/2021

KEY THEMES: runtime resource management, high-performance computing, power management, real time, critical systems, mixed criticality, internet of things, thermal management, hardware heterogeneity, exascale computing.

PROJECT COORDINATOR: Professor William Fornaciari (Politecnico di Milano)

PROJECT TECHNICAL MANAGER: Professor Giovanni Agosta (Politecnico di Milano)

PARTNERS: Spain: Universitat Politècnica de València, Barcelona Supercomputing Center; Italy: Centro Regionale Information Communication Technology, IBT Solutions; Switzerland: École polytechnique fédérale de Lausanne, Centre Hospitalier Universitaire Vaudois; Poland: Poznań Supercomputing and Networking Center

BUDGET: € 3.29 million

WEBSITE: recipe-project.eu

CONTACT: william.fornaciari@polimi.it

RECIPE is funded by the European Commission under the H2020 Framework Programme for Research and Innovation FETHPC-02-2017 – Transition to Exascale Computing under grant agreement nr. 801137

'IF MANY SMALL ISSUES CAN BE SOLVED IN A COORDINATED WAY, WE CAN FIND AN OVERALL SOLUTION'



The Heterogeneity Alliance coordinates efforts by many different organizations to meet the challenge posed by heterogeneous architectures. Clara Pezuela (Atos Spain) tells us more.

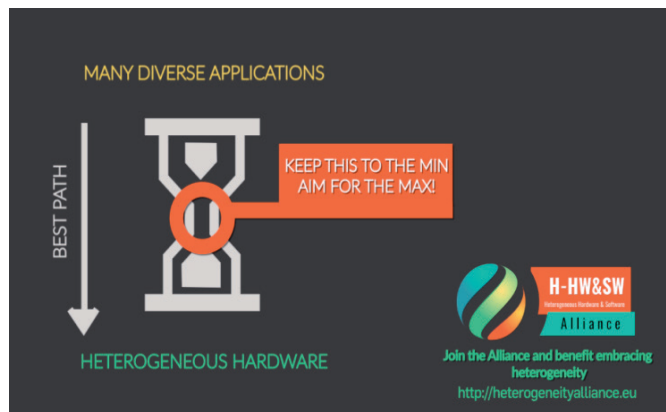
Why are heterogeneous architectures so important?

Combining different processor types in one system to improve absolute performance, minimize power consumption and / or lower costs, it's little wonder that heterogeneous architectures have received considerable attention. With the impact of this heterogeneity rapidly increasing, innovative architectures, algorithms, and specialized programming environments and tools are needed to efficiently use these mixed / diversified architectures.

The transition to multicore processors, graphics processing unit (GPU) computing, reconfigurable acceleration, and hardware-as-a-service (HaaS) cloud computing may be viewed as a single trend. As the market for heterogeneous architectures / multicore processors in embedded applications has begun to move into the product deployment stage, the need for software and underlying programming methodologies is also increasing in parallel.

Why did you decide to launch the Heterogeneity Alliance?

The challenges that heterogeneity bring – both in hardware and software – are so complex and broad that it is practically impossible for individual organizations to tackle them on their own. Many projects and organizations are starting to focus on solving these challenges, so why not join forces and do it together? One project or organization may solve only one minor



issue, but if many small issues can be solved in a coordinated and complementary way, we can find an overall solution for more ambitious challenges.

Members of the Heterogeneity Alliance bring their own methods and tools which solve a particular heterogeneity challenge – see the catalogue on our website for more information. They contribute to the definition of an architecture which can cope with all aspects of heterogeneity. In addition, they support the promotion and awareness of the Alliance mission by publishing news, organizing joint workshops and even contributing to our joint book about the challenges of heterogeneous architectures.

How would you like the field of heterogeneous architectures to look in 10 years' time?

Programming directly with this kind of heterogeneity is, frankly, a nightmare. In order to rapidly develop these kinds of solution and help companies use them for the construction of richer experiences and innovative business models, as software engineers we need to do something different. There

is an opportunity to build something in the middle to help us take advantage of high level abstractions – as in the past with traditional computing.

The information technology community therefore needs to start developing innovative architectures, algorithms and even specialized programming environments and tools to efficiently deal with very different architectures to be more productive, obtain rapid learning curves and to build more general solutions.

In the vision document sent to the HiPEAC roadmap working group, the Alliance highlighted some essential topics to be dealt with in next 10 years. You can read more about them in our book, but examples include:

- architectures and platforms for heterogeneity management
- heterogeneous execution support for managed runtime environments
- easy-to-use programming models for heterogeneous execution
- methodologies for predictability assessment on multi- and many-core systems
- hardware / software features to enable predictability in real time
- open-source tools for debugging parallel code running on ucore accelerators
- and many more...

FURTHER READING:

heterogeneityalliance.eu

ANTAREX: TECHNOLOGIES FOR ENERGY-EFFICIENT HPC SYSTEMS TOWARDS EXASCALE



The Horizon 2020 ANTAREX project due to concluded on 30 November 2018. We caught up with project coordinator Cristina Silvano (Politecnico di Milano) and project technical manager João M. P. Cardoso (University of Porto) to find out about ANTAREX's energy-saving techniques.

Why is energy efficiency so important in supercomputing?

Energy efficiency is of paramount for any class of systems, from embedded right the way up to high-performance computing (HPC) systems. Research on supercomputing is pushing towards the target of a 20MW exascale supercomputer – that is, one capable of a billion billion calculations per second – by 2021. Currently, the IBM Summit, equipped with IBM Power9 processors and NVIDIA VOLTA GV100 graphics processing units (GPUs), is the fastest computer in the world according to the TOP500 list. This computer has reached 143.5 petaFLOPS (million billion floating operations per second) and is ranked third on the Green500 list, with 14.7 gigaFLOPS/W.

To reach a 20MW exascale supercomputer would entail more than tripling the energy efficiency of today's supercomputers to around 50 gigaFLOPS/W. We believe that the next generation of supercomputers need a radically new software stack capable of exploiting the benefits offered by heterogeneity to meet the programmability, scalability and energy efficiency required by the exascale.

What are the main innovations delivered by ANTAREX?

In a nutshell, ANTAREX aims to provide a breakthrough approach for application self-adaptivity and to runtime manage, monitor and autotune applications for energy-efficient HPC systems up to the exascale level.

One key innovation of our approach is the concept of 'separation of concerns'. This enables the end-user application programmer to express adaptivity strategies and non-functional requirements, such as throughput and power constraints, in addition to application functionality. This was promoted by the design of the **LARA domain-specific language (DSL)**, inspired by aspect-oriented programming concepts for heterogeneous systems.



The ANTAREX group at CINECA (Italy)

Leveraging the LARA DSL, ANTAREX introduced a compile-time framework composed of the Clava source-to-source compiler, the **LibVC** dynamic compiler and function memorization, as well as machine learning techniques for compiler autotuning (**MICOMP** and **COBAYN**).

Another innovative technology is the ANTAREX runtime framework, based on the **mARGOt** application autotuner, to enable runtime tuneable approximations at the application-level by trading off accuracy versus throughput. The aim of this was to make HPC applications self-aware of their runtime behaviour.

The ANTAREX runtime framework also includes the **Examon** scalable, fine-grained energy-monitoring system and the **Countdown** power manager, which were recently integrated into the CINECA HPC production environment. The goal is to monitor and runtime manage the scalability and adaptability of a dynamic workload by exploiting the full system capabilities, including energy management, for emerging large-scale and extreme-scale systems.

What sort of challenges could we solve by exploiting the ANTAREX technologies?

ANTAREX addresses the autotuning and scalability characteristics of two important application scenarios in HPC: a biopharmaceutical application for accelerating drug discovery and an intelligent navigation system for traffic congestion mitigation in smart cities. These use cases were chosen for their direct economic exploitability and their major social impact.

Scaling up and accelerating molecular docking

For the biopharmaceutical use case, we worked with the biopharmaceutical company Dompé, part of the ANTAREX consortium. The focus was on molecular docking, an increasingly important application for HPC-accelerated drug discovery. We started by analysing the most computationally intensive kernels of the LiGen molecular docking application used by Dompé. We developed a runtime tuneable version of the molecular docking application for use in virtual screening experiments. This was deployed and scaled out to the full size of the 10.4 petaFLOPS Marconi supercomputer at CINECA to screen a database containing a billion ligands – biochemical substances which bind to biological molecules –with the aim of targeting unresolved infectious diseases. This represents the largest virtual screening experiment ever launched in terms of computational threads (up to one million) and size of the compound database (one billion ligands).

Participating in ANTAREX allowed Dompé to take advantage of HPC-accelerated and tuneable solutions, thus illuminating new

development paths not viable using conventional computing. Exploiting ANTAREX's HPC technologies supporting autotuning, scalability and energy efficiency, Dompé is now able to optimize molecular docking to reduce the virtual screening process for the identification of new active compounds by two orders of magnitude.

Mapping traffic congestion in smart cities

Slovak navigation application company Sygic and Czech supercomputing centre IT4Innovations worked on a smart navigation system to help mitigate traffic congestion in smart cities, using an innovative algorithm of road-balanced routing. Exploiting supercomputing power and ANTAREX code optimization technology, we reached the point of being able to calculate routes for tens of thousands of drivers simultaneously and perpetually towards a global optimum, i.e. providing traffic-flow-optimized navigation to reduce total travel time.

The project helped Sygic create solid foundations for a revolutionary product for municipalities, which is ready to be deployed in a pilot scheme immediately. Using the ANTAREX technologies supporting autotuning and scalability, Sygic is now able to adapt the product for a wide range of municipalities with their specific cost/performance requirements.

Sygic envisions pilot deployments in forward-thinking cities and to this end has already established contact with the municipalities of Milano, Ostrava, and Vienna. Thanks to reduced travel times and emissions, this system improves citizens' quality of life, hence we believe it will also be attractive to other cities. As the business assumption is that municipalities will bear the cost of the solution, providing the services to citizens for free, further improvements to the product will aim to improve the cost efficiency of the computation, exploiting ANTAREX optimization technologies to the full.

NAME: ANTAREX Autotuning and adaptivity approach for energy efficient exascale HPC systems

START/END DATE: 01/09/2015-30/11/2018

KEYWORDS: exascale computing, energy efficiency, application autotuning,

COORDINATOR: Professor Cristina Silvano, Politecnico di Milano, Italy

PARTNERS: Italy: Politecnico di Milano, CINECA, Dompé; Switzerland: ETH Zürich; Portugal: Universidade do Porto; France: INRIA; Czech Republic: IT4Innovations; Slovakia: Sygic.

EU CONTRIBUTION: € 3.12 million

WEBSITE: antarex-project.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 671623.



A recent survey of the HiPEAC community revealed that one of the most valued instruments for HiPEAC members, whether academic or industry, is the HiPEAC Jobs portal. But what is this value, exactly?

More than a jobs website: The value of HiPEAC Jobs

Over the coming years, there will be an expected shortfall of more than 500,000 information and communication technology workers in Europe, a fact which is painfully obvious for anyone who has spent several months trying to recruit for certain positions. This is critical, especially for small and medium enterprises (SMEs) lacking a well-known brand or personnel devoted to talent attraction.

Generalist platforms such as Monster, Indeed or LinkedIn usually attract hundreds of unsuitable applicants, leading to hours of time-consuming filtering to identify the genuinely appropriate profiles. Alternatively, there are literally hundreds of niche portals focused on academics, career levels such as PhDs and / or technological areas such as computer science that may allow users to reach out to candidates with specialist profiles. They are usually private platforms requiring payment ranging from €500 to €2,000 to post a limited number of jobs or to post for a limited time. It is unclear whether the investment pays off. For higher levels of specialization there are head hunters and external talent seekers, but their costs are considerably higher.

The HiPEAC Jobs portal has proved to be a good tool for attracting and retaining talent in Europe. The portal allows employers to promote opportunities in Europe free of charge and to reach the right community of specialists in computing systems. Hence, recruiters are able to find the right candidates without having to spend either too much time sorting out unwanted applications or thousands of euros on platforms or talent consultants. At the same time, it also helps members with deep, specialist knowledge to find a company that values their expertise, through contact with the right people at the hiring institution.

Recently the use of the Jobs portal has been extended to promote short-term collaborations and internships, under HiPEAC's mobility programme. The dynamism of the portal is indicated by the number of job openings posted and general usage statistics, as shown in the table below.

Number of jobs posted in the last quarter by institution	Jobs from selected institutions, sponsors, contributors and projects
44 jobs @ research centres	
31 jobs @ industry	
31 jobs @ SMEs	
52 jobs @ universities	
18 jobs @ H2020 projects	
16 internship positions	

To increase the reach of the portal, HiPEAC also organizes a number of complementary activities:

- A jobs wall displaying current vacancies is featured at every HiPEAC event.
- At our 'Inspiring Futures!' career mentoring workshops, organizations are invited to pitch their open positions and raise awareness of their brand.
- HiPEAC's annual ACACES summer school offers the opportunity to approach over 200 students in a more informal setting.
- The regular HiPEAC Student Challenge allows companies to see how students perform in action.
- Exhibitors at the HiPEAC conference have the opportunity to meet students attending the STEM Student Day and attract visitors to their booth even during quiet periods.

Thanks to the success of these activities, a number of external partners have started asking HiPEAC to organize a jobs wall or organize a career session at their university departments, conference, summer school or workshop.

Would you like to set up a jobs wall or a careers-based activity powered by HiPEAC? Contact us: [✉ recruitment@hipeac.net](mailto:recruitment@hipeac.net)

FURTHER INFORMATION:

hipeac.net/jobs/#/career-center

hipeac.net/jobs

Follow us on Twitter @HiPEACjobs





Ten years ago, a book made Vincent Hindriksen change tack in his career. Now the managing director of performance optimization company Stream HPC, Vincent explains how he ended up in high-performance computing and the lessons he's learned along the way.

Career talk: Vincent Hindriksen, Stream HPC

How did you get into computer science; more specifically, how did you get into high-performance computing (HPC)?

Initially I didn't look into computing science, but more into something like physics. After visiting universities computing science became my final choice unexpectedly, even though I couldn't always relate to others looking into this subject.

I chose the University of Twente in the Netherlands, a good technical university. However, because of my relationship at the time, I moved to Groningen. There were many good things about my time at that university, except the computing lab had a single server comparable to a 486 serving 40 thin clients. And we had to use those computers to learn to program, meaning a long loop already took a few seconds.

I still remember people shouting 'Who the beep launched Netscape?', as this would cause all other computers to freeze. The supercomputer where I learnt Fortran on was so old, it mostly burnt energy and was only turned on for learning Fortran classes when you reserved a spot. So did I choose HPC? Not really. Only during my thesis project at Astron, the Netherlands Institute for Radio Astronomy, (LOFAR project) did I learn real optimizations for a good cause.

"I still remember people shouting 'Who the beep launched Netscape?'"

The real decision to work in HPC came years later. During my first job I found I still enjoyed the performance optimizations I learnt at Astron. The requirement was only 1.5 transactions per second? I'd avoid a few memory-copies and get it a few factors faster. Two and a half hours to process data? Avoiding a lot of database-calls and keeping data in-memory brought it down to 20 seconds. The rest of my job I found utterly boring. I called it A-to-B programming – A is where we are, B is where we must go and the simple path towards B will be my boring job over the next three months. I missed something which would get my brain buzzing.

One rainy day in November 2009, I walked into a bookstore and saw a book titled *Boreout*. It described a sense of boredom and demotivation as a result of not being engaged at work and suggested that the effects could be the same as burnout. Several months later I registered a domain for StreamComputing and started a company. I focused totally on OpenCL, something that got my brains buzzing and thus made me happy.

I did not see programming graphics processing units (GPUs) as HPC initially. This was because we focused and still focus on a single node first, and multi-node is often an evolution of the developed algorithms and not a totally new thing. So I got into HPC without immediately realizing it, which might be why I like working with customers who try to solve a problem and don't realize that others would label the solution HPC.

What have been some of the highlights of your career so far?

It has been a wild ride, especially because I wanted to write software and put business second. This resulted in some pretty difficult years, until I started to become company leader full time and fully trusted the team for building great software. The biggest highlight has been this professionalizing of the company while keeping our promises on quality and performance. To make this happen, I simply made sure that I only hired better programmers than myself.



Vincent studied in Groningen

For the future we'll focus more on clusters, as we think that our way of building performance-optimal software is very useful for such projects. Of course we'll include more artificial intelligence (AI) and other easily accessible solutions.

What were some of the biggest challenges you faced when setting up the business?

Running a business is learning a very broad spectrum of specializations. You cannot just focus on your strengths to keep the company running. I knew that to make Stream a success I had to keep learning, pivot quickly and mostly just persevere. If I were to do it again, I would do it with a team. Most of the mistakes I made were because I was working under pressure for long periods of time.

What advice would you give students about to set out on their career today?

Four tips that get you into HPC much easier:

“Never put Java and .NET on your résumé (to avoid recruiters for boring office jobs)”



Photo credits: Hans Ravensbergen / Rudy and Peter Skitterians, Pixabay

- Performance programming is a puzzle you do in your head, so train yourself to approach the problem theoretically and don't just try and run variations of an algorithm. Point and click puzzle adventures are a good way to get good at problem solving; luckily, there are many of these games from the last 30 years that are still difficult to solve.
- Solve a very difficult mathematical problem on a computer. There are many tactics and approaches, so make sure you prepare. Many great programmers solved a big problem when they were 14 or 15. Many other great programmers solved a big problem when they were older. The rest only remember '14 or 15' and think they started too late. Ask friends, teachers or connections on social media for a difficult problem that suits your interests.
- After graduating, you have a good opportunity to end up in the industry others think you have the most experience in. So the best thing is to convince them that you're made for HPC, embedded and other interesting industries. Make sure you have lower level languages (C, OpenCL, Fortran) as your first language (to really understand hardware). Then learn the difference between scripting, pragmas and programming (thus never compare Python with OpenMP or C). Last but certainly not least, never put Java and .NET on your résumé (to avoid recruiters for boring office-jobs).
- Then the mindset. Looking back, I think rewarding yourself more for solving puzzles would really help, instead of getting mad at yourself if you can't find the answer.

How do you see the field evolving?

AI and hyperscale; there are many reasons this will define HPC over the next 10 years. Hyperscale is not HPC from my perspective – it is HPC-hardware optimized for a single purpose (16-bit GPUs, tensor processing units etc.). Hyperscale will therefore remain cheaper than HPC and with such a large workforce of AI-developers, this could mean it would be simply cheaper to throw a huge AI installation at any HPC-problem before trying the old-fashioned algorithmic way where the human has to think.

While researching new hardware, the HiPEAC Vision recommends investing in the development of accelerators. This issue's featured thesis is by Nuno Paulino, now an assistant professor at the University of Porto. It focuses on methods to help design application-specific hardware for embedded systems.

Three-minute thesis



NAME: Nuno Paulino

RESEARCH CENTER: Institute for Systems and Computer Engineering of Porto (INESC Porto)

SUPERVISOR: João Canas Ferreira

CO-SUPERVISOR: João MP Cardoso

DATE DEFENDED: July 2016

THESIS TITLE: Generation of Custom Run-time Reconfigurable Hardware for Transparent Binary Acceleration

With increases in application complexity and the amount of data being processed, computational power needs to increase in tandem. Until just over a decade ago, technology improvements allowed for the increase in clock frequencies in conventional processors. However, with the forecast end of Moore's Law, new architectures and computing paradigms are being explored, with a particular focus on scalability and power efficiency.

In the context of embedded systems, a straightforward way to achieve ideal performance and power consumption is to design custom hardware on a per-application basis. However, this solution requires advanced hardware expertise as well as implying a long development time and very high manufacturing costs.

For this thesis, we designed a transparent binary acceleration approach targeting field-programmable gate arrays (FPGAs) where instruction traces are used to generate specialized accelerators. A custom accelerator, capable of executing a set of previously detected loop traces, is coupled to a host MicroBlaze processor. This approach does not require the application source

code to be modified, ensuring transparency for the application developer. No custom compilers are necessary, and the binary code is not modified either offline or during runtime. The accelerators contain per-instance reconfiguration capabilities, which allow for the reuse of computing units between accelerated loops, without sacrificing the benefits of circuit specialization.

To increase performance, the accelerator can perform two concurrent memory accesses to the MicroBlaze's data memory, and the repetitive nature of the loop traces is exploited via loop-pipelining. By supporting single-precision floating-point operations via fully-pipelined units, the accelerator is capable of executing realistic data-oriented loops. Finally, the use of dynamic partial reconfiguration (DPR) allows for significant area savings when instantiating accelerators with numerous configurations, and also ensures circuit specialization per-configuration.

Several fully functional systems were implemented to validate the design iterations of the accelerator. First, control and dataflow graph representations of the traces were translated into a multi-row array of functional units, achieving a geometric mean speedup of 2.08x for 15 benchmarks. A second implementation augmented the accelerator with memory access to arbitrary addresses of the entire local data memory of the MicroBlaze, so that data-parallelism in realistic workloads could be exploited, improving the mean geometric speedup to 2.35x for 37 benchmarks.

The most efficient implementation supports floating-point operations and relies on loop pipelining. Accelerator instances are generated by modulo-scheduling traces at the minimum possible initiation intervals, achieving a geometric mean speedup of 5.61x for 24 benchmarks, with accelerators that require only 1.12x the FPGA slices required by the MicroBlaze. Through DPR, an accelerator with 10 configurations requires only a third of the lookup tables relative to an equivalent accelerator without this capability.

To summarize, the approach is capable of expediently generating accelerator-augmented embedded systems which achieve considerable performance increases at a low resource cost and without requiring manual hardware design.

HIPEAC

Thanks to our sponsors for making #HIPEAC19 a great success!



Sponsors correct at time of going to print. For the full list, see hipeac.net/2019/valencia

Join the community



@hipeac



hipeac.net/linkedin



hipeac.net

This project has received funding from the European Union's Horizon2020 research and innovation programme under grant agreement no. 779656

