Cyber security threats: HiPEAC strikes back

2017 H2020 Customised and low energy computing call

Opinion: the complex software/hardware interface
The 2017 call for proposals is coming

How HiPEACers handle hackers - Locking the back door

Innovation Europe

3 Welcome
Koen De Bosschere

4 Policy corner
The 2017 call for proposals is coming
Sandro D’Elia

5 News
A round-up of the latest news from our community

11 Security special
Locking the back door
Simha Sethumadhavan, Bjorn De Sutter, Philipp Miedl, Davide Basilio Bartolini, Vassilis Prevelakis, Sergey Tverdyshev

18 Technology Transfer
KALEAO introduces the first web scale “True Converged” platform
Giovanbattista Mattiussi

20 Innovation Europe
Getting high performance from low power
Alberto Scionti

21 Innovation Europe
Good timing: PROXIMA results to have wide-ranging impact
Liliana Caciu-Grosjean

22 Innovation Europe
Successful Conclusion to SCoRPiO Project
Nikolaos Bellas

23 Innovation Europe
A modular microserver data centre to power the data economy
Mariano Cecowski

24 Innovation Europe
Looking ahead to improved and flexible computer vision: Eyes of Things
Oscar Denis Suarez

25 HiPEAC Jobs: supporting recruiters and researchers

26 SME snapshot
From TV, to railways to insects: enabling innovation at Embedded Computing Specialists
Bertrand Rousseau, Philippe Manet

27 Industry focus
Movidius: leveraging EXCESS innovations for up to 50x power efficiency
Jack Dashwood

28 Industry focus
Embecosm: Towards HPC and data centres which don’t cost the earth
Dan Gorringe, Jeremy Bennett

30 Peac performance
Breaking the energy efficiency limit for conventional logic gates
Luca Gammaitoni

32 Peac performance
Technology opinion: Towards a style manual for hardware designers
Timothy Roscoe

34 HiPEAC futures
Career talk: Marco Cornero, ARM
HiPEAC internships: Victor Garcia, Barcelona
Supercomputing Center
HiPEAC collaboration grants: Konstantinos Maragos, National Technical University of Athens
Three-minute thesis: Marta Ortín Obón, University of Zaragoza
Postdoc funding focus: EU Marie Skłodowska-Curie Individual Fellowships: Xabier Iturbe, ARM
I hope you enjoyed a relaxing summer holiday and that you are ready to tackle the challenges of the coming year. Every other summer, HiPEAC works on its HiPEAC Vision, which describes the current computing systems landscape and makes recommendations to help the community and the European Commission improve the innovation potential of future research. Part of this effort is a SWOT analysis of computing systems in Europe. A recent report on science, research and innovation performance by the European Commission summarizes the European situation very well. The computing industry in Europe contributes less than 4% of the European GDP while the American and Japanese computing industries contribute more than 5% of their GDPs. European ICT companies spend less than 6% of their value added on research and development. American and Japanese ICT companies spend twice that amount. In the US, the amount of venture capital available in 2013 was six times larger than the amount available in Europe (while both economies are similar in size).

All these numbers should be a wake-up call. According to Commissioner Günther Oettinger, 150,000 extra jobs could be created in the European ICT industry per year. Yet this is only going to happen if we want it to happen: if companies start investing, if young people decide they want to be trained to work in ICT, if financial institutions are willing to invest in expanding companies. If this does not happen, these jobs will be created elsewhere in the world. Europe has the largest market in the world, it has one of the best educated populations in the world, it has a higher household consumption expenditure than the USA, and it is the safest large continent in the world. So, there is no reason why we cannot make this happen – we only have to want it. There is a role for the HiPEAC community here: spread the word, innovate more, create startup companies, train more young people. I hope that this edition of HiPEACInfo provides something to inspire you.

The next event we are organizing is the Autumn Computing Systems Week in Dublin. The event is a great networking opportunity for the community, and I am happy to see again that a number of European projects are co-locating their meetings with CSW. In January, there is the yearly HiPEAC conference, this time in beautiful Stockholm. We have again a large number of co-located events making the conference one of the largest networking events for the computing systems community in Europe. I hope that the event will be attended by a record number of people from industry because collaboration between academia and industry is one of the key ingredients for a strong innovation ecosystem in Europe.
A critical problem in these sectors is the cost and complexity of software development, especially because it is still difficult to write efficient programs for recent system architectures (heterogeneous, highly parallel). For this reason, the first version of the 2016-2017 Work Programme focused only on software ("Programming environments and toolboxes for low energy and highly parallel computing"). But 2016 has seen a stronger policy focus on the technologies for digitizing industry, and so it was possible to allocate extra funding for hardware research: the EC will fund one project in the range of €6-10 million to start the development of next-generation high-performance processors.

The main objective is to get a substantial and measurable improvement over the current state of the art in energy/performance ratio for high performance computing and server workloads. This is a step towards the “exascale” performance of tomorrow’s fastest supercomputers, but is also meant to support all the applications where high performance must be coupled with low energy consumption, and criteria like efficiency and space are relevant; in general, these are the requirements of many cyber-physical systems and of many applications requiring computing power also at the edge of the network.

The Work Programme text does not impose specific technical solutions: the only strict requirements are to improve significantly the energy/performance ratio compared to the state of the art, and to develop a design that can be manufactured in volume at a reasonable cost. We do not just want a computer design exercise, but something that will change the market in the coming years.

Here lies the challenge: the EU needs you to develop the computer chips of tomorrow. These are the chips which will drive your car, make your city’s subway faster, monitor your health and maybe even foresee the exact time and location of the next earthquake before it strikes. If you are reading these lines in the HiPEAC Info newsletter, you are part of the community that can do it.

**MORE INFORMATION:**
Learning from the world’s best at HiPEAC ACACES Summer School

Doctoral students, researchers and world-leading experts in computer architecture and compilation for high performance and embedded systems came together in the beautiful Italian spa town of Fiuggi for a week of technical courses, networking and knowledge exchange on 10 - 16 July.

This year’s School continued the tradition of blending classes and lectures on both entry-level concepts and highly advanced techniques so as to ensure that existing understanding is consolidated and updated in the context of exciting new developments and directions. Twelve international experts taught courses across the week on topics including Self-Aware, Variation-Aware, Cross-Layer Embedded System Design; Hardware-Up Security; and Technology-based Entrepreneurship.

The programme was reinforced by a keynote speech given by Professor Edward A. Lee (U.C. Berkeley) on *The Internet of Important Things* and an invited talk on *Building a software company in Europe - and lessons learnt doing so* by Knut Degen, Co-Founder and CEO of SYSGO AG.

For Radhika Jagtap, of ARM Research (UK), ‘attending ACACES is a great way to learn about new trends, about what’s coming up in academic research. It’s also a fantastic opportunity for networking and to pose questions to international experts.’

The twelve teachers stayed on campus all week, so as to allow participants ample opportunity to pose questions and discuss ideas with them. The poster session remains an integral part of the School as it underpins the aim to provide an informal forum for debate and sharing of ideas with a view to the formation of new collaborations and networks.

Poster presenter Daniele De Sensi of University of Pisa (Italy) says: ‘the best thing about coming to ACACES is the possibility to meet new people, to exchange ideas and viewpoints, and to have access to different people’s backgrounds. Secondly, this is a great opportunity to learn new things and to hear the lecturers’ perspective on technical issues and even on my own project at the poster session. As this is the only Europe-wide HPC summer school of this scale, it’s a unique opportunity; you can’t find such a community of scientists brought together to exchange ideas anywhere else.’

“You can’t find such a community of scientists brought together to exchange ideas anywhere else”
New horizons through HiPEAC

For ACACES2016 participant Mustapha Bouhali, PhD student at the École National Polytechnique d’Oran (Algeria), attendance represented something very special. Mustapha is in the first year of his PhD studies on applications of FPGA platforms in the design of unmanned aerial vehicles (UAVs) for transportation and other purposes.

‘Being at ACACES in Fiuggi represents my first trip outside Algeria. I am a self-funded PhD student and so being here is reassuring me that I am doing the right thing, that the time and the money that my family and I are investing in my studies are worth it. In Algeria, government funding for doctoral studies is extremely difficult to obtain, it is a very difficult environment in which to be a PhD student,’ Mustapha said. ‘Attending ACACES is also helping me to meet other people who work in my area, to grow my network and, hopefully, to forge new collaborations. At this Summer School, I’ve met many researchers who are open for future collaborations and who’ve seemed really interested in pursuing options with me.’

In countries where government and private funding for all levels of research is in short supply, it is a huge decision to choose a career in research: ‘I couldn’t do what I am doing without the support of my parents,’ Mustapha concluded. ‘Being here is making me feel confident that our commitment is worth it.’

What is your brain up to?

It is now a little easier to find out. HiPEAC SME member PLUX - Wireless Biosignals, S.A. launched in August a new version of its BiTalino board: the BiTalino (r)evolution. The board is offered as an all-in-one toolkit with a microcontroller, wireless transmitter and built-in biosignal sensing modules and the (r)evolution offers increased sensing and connectivity capabilities in a much smaller package when compared to the previous version.

The new board contains modules for measuring not only motion, skin conductance, heart signals and muscle signals but also brainwaves. What’s more, the cleverly designed packaging of the circuit board is customizable, allowing users to snap bits off so as to be able to form wearable units of the size and shape that meets their needs and to easily embed the device into another tool or gadget.

The BiTalino (r)evolution comes with everything you need to start your own bio-sensing project: the BiTalino circuit board, LiPo battery, electrode pads, cables and a quick-start instruction booklet.

As the device is intended for direct use on the human body, safety is paramount. The board runs on a small LiPo cell so as to avoid risk of electric shocks to the user. It communicates with the outside world using Bluetooth or BLE and the measurements taken are transferred to BiTalino’s OpenSignals software platform, which can be easily accessed through a web browser for viewing the recorded data.

For more information on the BiTalino (r)evolution, contact Hugo Silva hsilva@plux.info – www.bitalino.com

Keynote speaker Professor Edward A. Lee commented that he was ‘impressed by the students’ talent: they really have a lot to offer. This allows for classes with more of a two-way action, which is enjoyable and informative for us all. The poster session has been a very positive experience and I think it is a very important part of the Summer School. I have really enjoyed my time here in such a beautiful setting.’

The thirteenth ACACES Summer School will take place in 2017 and will be, as always, ‘open to all’: to students and researchers in relevant fields working in either academic institutions or the private sector.

Affiliated and former ACACES student Alexandre Aminot’s side-project app reached #5 in top iOS downloads in France this summer: http://tupreferesapp.fr/
Exploring all things CUDA at the seventh PUMPS summer school

As an NVIDIA GPU Center of Excellence, Barcelona Supercomputing Center (BSC) proudly hosts the Programming and Tuning Massively Parallel Systems (PUMPS) summer school every year. Organized by BSC, the University of Illinois, the Universitat Politècnica de Catalunya-Barcelona Tech and HiPEAC, the seventh edition, held on 11-15 July, once again offered the opportunity to learn more about graphics processing unit (GPU) architectures and programming languages for GPU computing. As usual, it was extremely popular, attracting 140 applications. Eighty students from 22 countries were finally selected based on their prior knowledge of CUDA and how PUMPS could be beneficial for their work.

Co-directed by Mateo Valero and Wen-mei Hwu, and organized by Antonio Peña and Victor García, the programme featured internationally recognized lecturers including Wen-mei Hwu (University of Illinois at Urbana-Champaign), David Kirk (NVIDIA Corporation) and Juan Gómez Luna (Universidad de Córdoba). Topics covered included dealing with non-uniform data, computational thinking and machine learning, higher-level programming on GPUs, dynamic parallelism and specialist classes on BSC’s OmpSs programming model.

The winners of the best poster award were Josef Michálek and Jan Vaník (University of West Bohemia), with their poster ‘Chunking SVM Training Implementation in CUDA’. The award for best clinic went to Athanasios M. Kintsakis (Aristotle University of Thessaloniki). All winners were presented with an NVIDIA Tesla K40 graphic card, kindly donated by NVIDIA.

Manager of the GPU Center of Excellence at BSC, Antonio Peña, commented: ‘This year’s is another highly successful edition of PUMPS. You feel it was worth all the work when you see how involved the attendees are: asking questions, working on the hands-on exercises, and even staying after the class is over. The poster and clinics sessions were also of very high quality - it wasn’t easy to select a winner. We are very proud to host the most prestigious GPU computing summer school in Europe.’

One face which was sadly missing from this year’s event was Nacho Navarro, HiPEAC partner and the ‘father’ of the summer school, who passed away in February 2016 at the age of 58. Paying tribute to Nacho, Mateo Valero said: ‘If I were to sum up my feelings about Nacho, I would say he was the most generous and extraordinary person I’ve ever met. He was always helping me, taking care of any needs me, my family or colleagues may have had…in two words, our “guardian angel”. I’m sure he is now in heaven looking after all of us.’
Building links in the EU’s newer member states

On 17 June, HiPEAC Coordinator Koen de Bosschere and Steering Committee member Rainer Leupers co-organized a workshop at AGH University of Science and Technology in Kraków, Poland, to promote and encourage membership of the HiPEAC network. The workshop took place as part of a long-term and ongoing effort to promote the network to potential members and stakeholders in EU ‘new member states’.

While HiPEAC membership numbers are going from strength to strength, new member states, that’s to say nation states which have joined the EU since 2004, continue to represent a relatively small proportion of total membership at around 10%. Reaching out to the high performance and embedded architecture and compilation communities in these regions is a key strategic aim for HiPEAC as increased participation will make our existing community more robust, diverse and interesting. It will also support the success of researchers in the new member states in obtaining funding from programmes such as the EU’s Horizon2020.

Opened by AGH UST Vice-Rector Tomasz Szmuc, the workshop saw HiPEAC staff present the project and the benefits of membership as well as selected member research activities. AGH UST staff reciprocated with presentations of key areas of work in the university, including Embedded systems in applications of fog computing, Quality of Experience: connecting network to user and Two speed embedded software development in double enterprise architecture environment. This allowed for potential collaborations and exchanges of knowledge to be identified.

Reflecting on the workshop, leader of the HiPEAC ‘Membership Management’ task Rainer Leupers of RWTH Aachen University noted that ‘the smooth organization and great hospitality offered by AGH UST allowed for a highly productive and enjoyable day which has already resulted in new memberships of the HiPEAC community’.

HiPEAC will continue to continue to vigorously encourage engagement and membership in new member states. The whole community stands to benefit from the increased participation of these countries and the new collaborations, partnerships and knowledge transfers which will come about.

HiPEAC membership is free and the application process is simple. For more details: www.hipeac.net/members/membership

---

Publication: Circuits and Systems for Security and Privacy

Editors: Farhana Sheikh and Leonel Sousa

How to design and develop security-aware circuits and systems

Co-edited by HiPEAC member Leonel Sousa of INESC-ID, IST, Universidade de Lisboa, Circuits and Systems for Security and Privacy begins by introducing the basic theoretical concepts and arithmetic used in algorithms for security and cryptography, and by reviewing the fundamental building blocks of cryptographic systems. It then analyzes the advantages and disadvantages of real-world implementations that not only optimize power, area and throughput but also resist side-channel attacks. Merging the perspectives of experts from industry and academia, the book provides valuable insight and necessary background for the design of security-aware circuits and systems as well as of efficient accelerators used in security applications.

For further information and to order a copy, visit the CRC Press website: www.crcpress.com/Circuits-and-Systems-for-Security-and-Privacy/Sheikh-Sousa/p/book/9781482236880
IT4Innovations researchers release ESPRESO –
the new massively parallel linear solver library

ExaScale PaRallel FETI Solver (ESPRESO) is a sparse linear solver library that the IT4Innovations National Supercomputing Centre in Ostrava, Czech Republic has been developing for the past three years. This summer, the alpha version was released to the public and can be downloaded from the project website espreso.it4i.cz.

ESPRESO contains not only the linear solver, but also several Finite Element (FEM) and Boundary Element (BEM) preprocessing tools designed particularly for FETI solvers. The BEM support was produced in collaboration with developers of the BEM4I library (bem4i.it4i.cz). The preprocessor supports FEM and BEM discretization for advection-diffusion equation, stokes flow, and structural mechanics. Real engineering problems can be imported from Ansys Workbench or OpenFOAM. In addition, a C API allows ESPRESO to be used as a solver library for third-party application. This has been used for integration with CSC ELMER. For large scale tests, the preprocessor also contains a multi-block benchmark generator. The post-processing and visualization is based on the VTK library and Paraview, including Paraview Catalyst for inSitu visualization.

The development of the library has been funded from several sources, each focused on the development of particular features. For instance, the EU FP7 EXA2CT project worked on the implementation of the Hybrid FETI algorithm. This method provides excellent numerical and parallel scalability that allowed the scientists to create this massively parallel library. A significant part of this work was carried out during a research internship at the Department of Aeronautics and Astronautics, Stanford University by two members of the IT4Innovations team.

Structural integrity of aircraft engine parts

Funding from the Intel Parallel Computing Centre supported the development of a new approach for acceleration of FETI methods in general, not just the Hybrid FETI, by Intel Xeon Phi accelerators. Meanwhile, the Oak Ridge National Laboratory (ORNL) Director Discretion Project supported work on the GPU acceleration of the solver. This project also allowed for tuning and scalability tests of the communication layer on the Lab’s Titan machine, which has 18,688 compute nodes, of which 95% were successfully used to solve a problem of up to 120 billion unknowns that arise from the discretization of the Laplace equation in 3D.
SAVE project delivers vital energy efficiency improvement technologies

The SAVE project drew to a close this summer having explored how complex hardware systems can more efficiently execute data intensive applications. By bringing together expertise from both industry and universities, SAVE has developed a number of innovations in hardware, software and operating system components. When integrated together, they can reduce application deployment costs and maximize usage of heterogeneous system computing units, resulting in energy efficiency being improved by up to twenty per cent.

A range of complex electronic systems stands to benefit from these innovations, including computer data centres, consumer electronics, automotive products and complex industrial electronics.

The computing units can be on chip, for example central processing units (CPUs) ranging from small and low-power to high-end and efficient, graphics processing units (GPUs), and dedicated accelerators. Alternatively, the units can be off chip, such as racks of dedicated accelerators or field-programmable gate arrays (FPGAs).

The prototyped technologies will enable performance and energy-efficiency gains in high-performance computing (HPC) and embedded heterogeneous systems.

Coordinator Cristiana Bolchini of Politecnico di Milano concluded that “dynamic trade-offs in performance and energy are becoming increasingly synonymous with heterogeneous system management. Within the SAVE project, researchers designed self-adaptive software components and associated optimization policies to manage at runtime the resources offered by the heterogeneous computing systems developed by our hardware vendor partners.”

SAVE was funded by the EC’s FP7 programme under grant agreement number 610996.

Read more at www.fp7-save.eu

HiPEAC news

International Conference on High Performance and Embedded Architectures and Compilers (HiPEAC 2017)
23-25 January 2017, Stockholm, Sweden
www.hipeac.net/2017/stockholm

22nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP 2017)
15th International Symposium on Code Generation and Optimization (CGO 2017)
4-8 February 2017, Austin, Texas, USA
http://hPCA2017.org/
http://conf.researchr.org/home/PPoPP-2017
http://cgo.org/cgo2017/

26th International Conference on Compiler Construction (CC 2017)
5-6 February 2017, Austin, Texas, USA

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

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

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

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

Dates for your diary

3rd IEEE/ACM International Conference on Big Data Science, Engineering and Applications (BDSEA 2016)
6-9 December 2016, Shanghai, China
http://computing.derby.ac.uk/ucc2016/

16th International Conference on Algorithms and Architectures for Parallel Processing
1st International Workshop on Theoretical Approaches to Performance Evaluation, Modeling and Simulation (TAPEMS)
14-16 December 2016, Granada, Spain
www.arcos.inf.uc3m.es/wp/ica3pp2016/
http://tapems.unex.es

22nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP 2017)
15th International Symposium on Code Generation and Optimization (CGO 2017)
4-8 February 2017, Austin, Texas, USA
http://hPCA2017.org/
http://conf.researchr.org/home/PPoPP-2017
http://cgo.org/cgo2017/

26th International Conference on Compiler Construction (CC 2017)
5-6 February 2017, Austin, Texas, USA

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

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

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

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

HiPEAC INFO 48
How HiPEACers handle hackers

Locking the back door

High-profile hacks are never far from the news, and security is something with which all computing systems engineers have to grapple. HiPEAC caught up with Simha Sethumadhavan, Björn De Sutter, Philipp Miedl, Davide Basilio Bartolini, Vassilis Prevelakis and Sergey Tverdyshev to learn more about this critical issue.

With computing systems playing an ever-greater role in many aspects of our lives, cybersecurity has never been so important. The digitalization of large swathes of industry and a huge number of services, along with soaring numbers of connected devices, means that the potential for mayhem caused by a security breach is of major concern to governments, businesses and individuals.

Little wonder, then, that this area is high on the agenda for the European Commission. In July this year, the Commission launched a public-private partnership on cybersecurity, expected to trigger €1.8 billion of investment by 2020, as part of a series of initiatives which aim both to better equip Europe against cyber-attacks and, at the same time, to strengthen the competitiveness of its cybersecurity sector. Indeed, cybersecurity was a topic important enough for discussion at September’s G20 Summit. A few days later, Yahoo revealed that 500 million user accounts were hacked in 2014 by a “state-sponsored actor”.

Concerns are justified when you begin to consider the devastating consequences which security breaches can have. One obvious example, noted by Columbia University faculty member, and founder of hardware security company Chip Scan, Professor Simha Sethumadhavan is hacking of banking systems, ‘which could compromise a region, country or organization’s entire financial infrastructure.’ Such a hack early in 2016 is reported to have resulted in the theft of US$80 billion from Bangladesh’s central...
bank through a series of transfers from its account at the Federal Reserve Bank of New York. Simha adds that ‘while an entire economy could be threatened by a successful large-scale hack, even more dangerous consequences are theoretically possible. The consequences of the hacking of a nuclear power plant’s ICT systems don’t even bear thinking about.’

**Increased vulnerability in an interconnected world**

The interconnected nature of today’s world and the ways in which we use and share information offer benefits and opportunities that we would not have thought possible even just a few decades ago. Yet the dangers are all too clear, and can be seen on many levels.

‘A monoculture, a system which is reliant on a particular technology or concept, something very common in computer systems, has an inherent vulnerability,’ says Simha. An attack can take down a large part of the system, such as the cryptosystems used in e-commerce. Yet at least such systems are designed with security in mind, he notes; in contrast, major infrastructures such as electrical power grids were originally ‘designed for reliability, not security, and so the retrospective addition of security features and techniques is a huge challenge.’ To make matters worse, the increasing interconnection offered to us by modern computing devices means that small problems can reverberate across the world; indeed, he emphasizes, ‘the hyperconnected nature of today’s world, and its potential consequences, should not be underestimated.’

This sentiment is echoed by Professor Vassilis Prevelakis of TU Braunschweig: ‘Society and individuals have come to rely on critical applications that sense and control system in the physical environment. The Medical Internet of Things, and the move towards Smart Cities and Smart Power Grids, for example, are driving the need for these Cyber Physical Systems (CPS) to function correctly and reliably. Until now, most CPS operated in isolation, that’s to say, their connectivity with the public internet was very limited. We are now linking everything together, and so systems without any appreciable security features are now accessible not only locally but from locations across the world.’
Evolution of attitudes on both sides of the battle

While national-scale security breaches may seem far off to most people, lower-level attacks on individuals are a well-recognized annoyance for end users everywhere. ‘Hackers’ motivations have evolved,’ Simha notes. ‘In the past, the hack itself was the objective and offered the satisfaction of having breached a system and the resulting bragging rights. Now, the incentives tend to be financial; the stealing and selling of personal data is a quick win.’

Email phishing scams are an obvious example of modern hacks: Simha gives as an example one which ‘might advertise a product and is customized to the recipient so as to make him or her more likely to click on a link contained within the email. This link will connect the user to a site which then releases malicious code, allowing a hacker to access the user’s computer. They could then lock the computer and demand a ransom in order to unlock it’. The problem is exacerbated, he argues, by the amount of personal information which individuals make available online today: ‘people’s perceptions and definitions of privacy have changed. We have lowered our guard, and this makes us more susceptible to being attacked.’

Professor Bjorn De Sutter of Ghent University, who coordinates the ASPIRE EU FP7 project on advanced software protection, agrees. ‘A decade ago, users’ mobile phones stored some text messages and their contacts’ phone numbers. That was it. Nowadays, our mobile devices offer access to our whole lives: our video collections, our location listings, the private and professional documents we store in the cloud. If you wear a sports watch, your phone can even store data as intimate as your heart rate over the last year. In addition, the phone contains apps to perform financial transactions and functions as an authentication device to access many paid services. It’s not hard to imagine how a compromised mobile device can disrupt a life, or, in the case of professional use, cause major economic damage.’

An additional pitfall posed by the Internet of Things, Simha points out, is that ‘people are providing personal data through devices such as smart TVs and smart thermostats which were not designed with security as a priority.’

Identifying and preventing attacks

Indeed, it is not only software that can present hackers with opportunities. The vast majority of today’s modern personal computing devices (including laptops and smartphones), as well as servers running cloud services, run on multicore processors. This means that huge quantities of sensitive and personal data are processed on and exchanged within such processors. Although multicore processors are not inherently more vulnerable to attacks than single-core versions, Philipp Miedl of ETH Zurich points out that ‘security always depends on the actual architecture of the multicore systems and on the type and nature of the resources or information being shared among the cores. This is where their main vulnerability can be found; often we assume that the cores in multicore systems can be treated as isolated elements, but they often share information and resources among each other’. The sharing of resources is driven by the need for efficiency, and not sharing would have a big impact on performance but, as Davide Basilio Bartolini of ETH Zurich adds, ‘processor designers need to be very careful in what structures are shared and how’. Work on this topic is currently being carried out at ETHZ and a number of other centres in Europe as part of the SAFURE EU Horizon2020 project.

Philipp, Davide and colleagues at ETHZ have recently published a paper which presents a thermal covert channel which would enable an attacker to leak a cryptographic key or a password. ‘This is a relatively small but very critical piece of data that works as a key for sensitive information,’ Davide explains. ‘A slow covert channel would be enough for an attacker to steal this key, resulting in easy and high bandwidth access to large amounts of sensitive data. In the paper, we exploit the fact that many modern multicore systems exploit thermal information, which is needed for thermal management, without strict security measures. Therefore, it is easy to use the thermal information to transfer data between two cores.’
If hackers’ motivations and the range of opportunities available to them have changed, as have our definitions of privacy, our attitude to security needs to change too. ‘It can be difficult to convince system users that security is essential, rather than an expensive luxury. Many only want to invest in it if there is a specific and defined threat,’ Simha points out. ‘This reactive approach leaves users vulnerable, unlike the proactive approach that ‘hardware-up’ promotes. Unprotected hardware leaves systems wide-open to attacks by hackers who are getting better and better and better at what they do. Users need to accept that attacks are a given, not random bad luck that might or might not strike.’

A focus on securing hardware as a way of securing a whole system is at the heart of the hardware-up research being carried out by Simha’s team at Columbia. ‘Hardware-up operates on a bottom-up, proactive basis. The guiding principle is to anticipate possible future attacks and to automatically build in defences. Using software to protect software just provides hackers with more opportunities to attack systems and outlets to exfiltrate information. Hardware-up permits the size and scope of a system’s protection software to be reduced, thus reducing the number of opportunities for hackers to attack the system.

It seems that our willingness to integrate computing systems into our lives and to share an ever increasing range of sensitive data through them may continue unabated, but computer architects and engineers are thinking ahead. Whilst hackers are becoming increasingly sophisticated in their skills and their targets, the tools and techniques being developed to ensure that all layers of a computing system are secure and resistant to attack will help us to remain one step ahead.

“One core runs an application which is functioning as the sender; as a side effect of running and waiting idle according to a well-defined pattern, this application causes heating and cooling of the silicon its core is built of. The second core runs an application which acts as the receiver by reading the thermal information either of the first core (if available) or of its own, which will still be affected since cores sit on the same piece of silicon. Thus, we show that the security principle of application isolation, which is broadly used on multicore systems, can easily be broken. This channel is not very fast, but is sufficient to extract passwords or cryptographic keys to get further access to sensitive data.’ By understanding how security can be breached, we can make ourselves better equipped to protect systems.

Simha provides a second example of hardware security vulnerabilities: the back door. ‘Hardware backdoors, an undocumented feature or a malicious circuit placed inside the hardware, can be extremely dangerous. Attackers use a specific code, for example one which repeatedly executes a set of instructions, so as to trigger malicious actions.’ However, research is making great steps towards tackling this: ‘by undermining the triggers and therefore changing hackers’ assumptions about the system’s design, we can protect against such attacks.’

Research prototypes in Simha Sethumadhavan’s group at Columbia. Photos: Columbia Engineering/Jeffrey Schifman

"Understanding how security can be breached, we can make ourselves better equipped to protect systems"
The transportation and manufacturing industries are some of the biggest winners from pervasive interconnections and adaptive services. Systems in these two industries often have safety critical functionalities: a plane/train/car cannot harm the passengers. Due to this criticality and responsibility, these systems are expensive to design, develop, deploy and maintain. Moreover, they are developed to run for decades. Security for these systems is very often an afterthought.

The big challenge is that interconnecting these systems with each other as well as to other less-or non-safety critical systems create a mixed-critical system without awareness of the security issues of the new deployment and newly created operating environment. The actors in these new environments can be anything from just unaware of the safety criticality to explicitly malicious. Recent headlines about Jeep hacking and manufacturing plants highlight just the tip of the iceberg.

Thus, it is becoming obvious that we all have to approach security as Security-By-Design for every chip, operating system, device, network element and, finally, integrated mixed-critical system. Only by having security in every system component (i.e. defence in the depth) and right at the heart of safety critical systems will we reach assurance that these systems are trustworthy, e.g. safe enough to drive us in autonomous cars.

I’d like to highlight two topics that go through the whole computing stack: concepts for security by design and for non-intrusive safety-aware attack prevention and detection.

There is a clear trend of increased security awareness: more new research ideas identified, a plethora of security start-ups, security-oriented companies acquired by established multinationals. This shows that the security field is huge and that this is just the beginning.

The SAFURE (SAFety and secURity by design for inter connected mixed-critical cyber-physical systems) project targets the design of CPS by implementing a methodology that ensures safety and security “by construction”. This methodology is enabled by a framework developed to extend system capabilities so as to control the concurrent effects of security threats on system behaviour. The current approach for security on safety-critical embedded systems is generally to keep subsystems separated, but this approach is now being challenged by technological evolution towards openness, increased communications and use of multi-core architectures.

The goals of the SAFURE project are:
- to implement a holistic approach to safety and security of embedded dependable systems, preventing and detecting potential attacks;
- to empower designers and developers with analysis methods, development tools and execution capabilities that jointly consider security and safety;
- to set the ground for the development of SAFURE-compliant mixed-critical embedded products.

The results of SAFURE will be a framework with the capability to detect, prevent and protect from security threats on safety, the ability to monitor system integrity from application level down to the hardware level including time, energy, temperature and data integrity;
- a methodology that supports the joint design of safety and security of embedded systems, assisting the designers and developers with tools and modelling language extensions;
- proof of concept through three industrial use cases in automotive and telecommunications;
- recommendations for extensions of standards to integrate security on safety-critical systems;
- specifications to design and develop SAFURE-compliant products.

The project leading to the application described has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 644080. This work was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0025. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Swiss Government.

www.safure.eu - coordination@safure.eu
Cyber-physical systems use a blend of embedded devices and traditional computing systems, and a variety of communication channels. The adoption of CPS offers a unique opportunity for revisiting the security stack to ensure that the new era of devices and services encompass lessons learned from the ICT cyber-security battles of recent decades. The Secure Hardware-Software Architectures for Robust Computing Systems (SHARCS) project aims to equip system designers with the tools to prevent malicious exploitation of software vulnerability or other security flaw by attackers or malicious software, by building a secure-by-design architecture spanning all system layers for end-to-end security. Consortium member Vassilis Prevelakis explains further: 'SHARCS is building hardware and software defence techniques as inherent components of their architectures. This way, computing systems built with SHARCS secure architecture will be very hard for attackers to compromise even if these systems run vulnerable software with common faults.'

The project’s work focuses on incorporating new security features in hardware, leveraging existing hardware features for building defence mechanisms, and making the necessary changes to the software stack for supporting new hardened applications and old legacy ones with increased security guarantees. SHARCS introduces the hardware design and implementation of two fundamental security concepts, namely Instruction Set Randomization (ISR) and Control-Flow Integrity (CFI), and the design and implementation of a new policy, which can be applied to legacy applications, where software recompilation is not possible. This new policy is enforced using an existing hardware feature, available in the x86 platform, which is not designed primarily for security. SHARCS will also take advantage of additional techniques, such as information flow tracking, secure hardware memory, fine-grained memory protection and type safety. While some demonstrators of the above techniques are based on particular architectures, it is important to note that SHARCS is platform-independent.

The consortium’s mix of academic research institutions and industry promotes the project’s agenda to have its technological innovations adopted and deployed via products and spin-offs. These innovations are being tested and demonstrated in three highly relevant ways:

• Creation of a secure, implantable neuromodulator for automatic prevention of seizures in patients. The implant will be enhanced with secure hardware and software so as to prevent malicious attacks from affecting implant functionality and/or from providing unauthorized access to the implant.

• Trusted execution of software on a potentially untrusted public cloud to prevent attacks on virtualized platforms using shared-tenancy hardware and hypervisor solutions provided by data centers to run their software applications in the Cloud.

• For smart cars, enhancing of Electronic Control Units (ECUs), as well as the onboard and off-board communication, with secure hardware and software to prevent attackers from manipulating functionality or gaining unauthorized access to ECUs.

Ideally, SHARCS pushes new functionality to the hardware level, and provides all necessary software-stack changes for
Special feature: security

ASPIRE: PROTECTION FROM ATTACKS ON MOBILE DEVICES

The ASPIRE project (Advanced Software Protection: Integration, Research and Exploitation) focuses on protecting assets embedded in software against man-at-the-end attacks, white-box attacks that are performed in the labs of malicious users. Developing trustworthy software for the thousands of different mobile devices and platforms in use is extremely challenging. What is more, whether for mobile or other platforms, providing security is highly complex because, in most scenarios, security issues can arise in any of the layers of abstraction at which computing devices are normally analysed and designed: physics, circuit level, computer architecture, operating systems, algorithmic level, protocol level, the list goes on. Capturing and addressing all of the possible problems is extremely difficult because those defending against attacks need to take care of all the vulnerable layers, whereas attackers only need to find an attack vector in one single layer or in a few layers combined.

Given this challenging context, the ASPIRE team has brought together market leaders in security sensitive ICT domains and academic institutions with the aim of establishing trustworthy software execution on untrusted mobile platforms that have at their disposal a regular or occasional network connection to a trusted entity.

One of the project’s main results is a protection tool chain that is convenient to use and that can deploy many different types of state-of-the-art software protection techniques and combinations thereof. It includes anti-reverse-engineering techniques, anti-tampering techniques and anti-debugging techniques. As Coordinator Bjorn De Sutter explains, ‘we have also developed strategies and tool support for making some techniques renewable. This means that the protected software and the protecting software components are frequently updated to shorten an attacker’s window of opportunity’.

The project has also developed a methodology to evaluate the strength of protections, including not only the ability of protections to slow down human attack activities, but also their resilience against automated attack techniques and tools. The methodology is used in an automated decision support system that can aid users of the protection tool chain in choosing the best combinations of available protections based on a simple annotation of the assets and security requirements in their source code. Bjorn adds: ‘although the decision support system is not yet as powerful as we would like it to be, we are making great progress and are already deploying our tools on three real-world use cases from our industrial partners. These use cases are a DRM protection library, a software license management library and a one-time password generator, all of which are critical pieces of software for the business models of many companies.’

The ASPIRE project has received funding from the European Union’s Seventh Framework Programme under grant agreement no. 609734.

www.aspire-fp7.eu – coordinator@aspire-fp7.eu

Dr. Sandro Gaycken, director of the ESMT Berlin Digital Society Institute, a strategic research institute for digital topics of the German DAX-companies, will give a keynote speech at HiPEAC17 on Cybersecurity - An unsolvable problem in the way of our IT-futures. Sandro’s research focuses on cyberwarfare, cyberdefence, cyberintelligence, and high security IT. He has worked for the German government as a strategist in the first design of a German Foreign and Security Policy on IT-matters, having been the lead author of the internet freedom and the cybersecurity/cyberdefence part of the policy. He serves as a regular commentator on IT-related incidents in the international press. www.hipeac.net/2017/stockholm/
Changing needs
In today’s era of digitally-powered innovation, service providers and IT organizations face the constant challenge of delivering ever-increasing IT services seamlessly, whilst adapting quickly to a variable business demand, retaining security, leveraging innovation and controlling costs.

These challenges are pushing IT organizations and service providers to look for more operational efficiency, IT simplification and automation, rationalization and real scalability of infrastructure, while also adopting solutions that can add the agility that business requires. At the same time, data centres need to deal with constantly increasing energy consumption, power capping issues and, in many instances, the scarcity of space, which may lead to costly new real estate developments.

First steps
A preliminary answer to these problems came from converged infrastructure, which brought compute, storage, networking and server virtualization into a single, centrally managed chassis. Often the hardware was pre-configured to run specific workloads, in some cases with also the software to manage those workloads. This family of solutions brought some benefits for the data centre, including:

• Accelerated and simplified data centre deployment with fewer errors
• Boosted performance and resource utilization. A common management interface and no trial-and-error tuning
• Single-vendor service and support

However, converged infrastructure has also its limits. While the vendor handles integration, users still pay for proprietary hardware and management software. The HW consolidation is de facto limited, and there are no gains in performance density or energy efficiency.

A new approach: hyperconvergence
An alternative to the converged system now comes in the form of hyperconverged solutions, which represent a converged infrastructure with a software-based and -driven architecture that vendors run with white box servers and other generic hardware. Hyperconvergence adds more simplicity, enables asset consolidation and brings a great deal of flexibility and agility to IT, but also brings the need for additional software to be purchased and added into the data centre’s stack, with potentially detrimental impacts on performance and energy efficiency.

KALEAO introduces the first web scale “True Converged” platform
In the latest instalment in our series of technology transfer success stories, we hear how KALEAO’s KMAX came out of European project EUROSERVER. Applying technologies and concepts developed in research projects helps companies to create products to meet rapidly changing customer demands.

KALEAO KMAX, recently launched at IP EXPO in London, is the first solution that implements the true convergence of compute, storage and networking in an extremely compact, scalable and low power platform.

Technology transfer: success stories

KALEAO

KALEAO KMAX

HiPEAC INFO

48
KALEAO recently introduced KMAX, the first web scale true converged solution that delivers hyperconverged results, together with a quite considerable increase in performance density, scalability and energy efficiency.

**True convergence**

True convergence is the native convergence of compute, storage, networking and integrated OpenStack in a single platform. KALEAO’s true convergence technology dynamically defines “physicalized” computing resources and directly assigns them to virtual machines and applications, without unnecessary software layers. An ultra-efficient microvisor works seamlessly with the innovative hardware platform to orchestrate global pools of independent software-defined and hardware-accelerated resources. This technology hence removes the performance overheads that are typical when layering applications over a virtualized, hyperconverged platform running on today’s undifferentiated commodity hardware, while bringing the appliance simplicity and the flexibility of a software-defined solution. Leveraging this advanced technology and the cost advantages of ARM 64bit hardware, the KALEAO solutions are capable of achieving significant improvements in energy efficiency, density and scalability.

KALEAO implements true convergence with the new **KMAX platform**, a hardware and software integrated solution, designed following the principles of low power consumption, data locality, high density and high performance.

**Technology transfer in action**

The architecture behind KMAX is the result of a different, more rational approach to server design. It has its root in the EUROSERVER EU FP7 project.

One of the goals of EUROSERVER was, and still is, to research and innovate key components towards ARM-based servers to ultimately design systems with very high energy efficiency, in line with the growing requirement of low power IT infrastructure coming from industry. KALEAO engaged in tangible R&D effort and built on the robust EUROSERVER conceptual grounds, to develop and release a fully-fledged off-the-shelf solution with the purpose of creating a product that was delivering value to customers, making their work better and their life easier.

In this respect, KALEAO succeeded. KMAX marks substantial advancements in:

- **Performance density**: 1536 CPU cores, 370 TB of all flash storage and 960Gb/s in 3U rackspace - to offer up to 10 times the performance density of today’s typical hyperconverged offerings, blades and rackmount solutions;
- **Energy efficiency**: less than 15W for each 8-core server with 10Gb/s I/O, providing over four times the performance per unit of energy consumed;
- **Scalability**: The improvement in latency and the lightness of the software architecture make the solution more suitable for maintaining performance at large scale, to grow with simple adds-on and to scale flexibly, thanks to the physicalization technology that creates disaggregated (but converged) resources that can be independently allocated to applications;
- **Cost**: KMAX further reduces cost of ownership by over 3 times by allowing the adoption of web scale, flexible and manageable infrastructure, paving the way for enterprises to obtain a more efficient and agile IT management that translates into bottom-line savings.

In summary, KMAX encapsulates an innovative approach to computing, conceived not to be pure technological fanciness, but to deliver real benefits to customers.

With headquarters in Cambridge UK, and offices in Crete, Italy and North Carolina (USA), KALEAO is a good example of how synergies between academic research, industry and the EU ecosystem can foster spin-offs and new ventures that can, ultimately, create value for society.

www.kaleao.com/Products/kmax
www.euroserver-project.eu/

EUROSERVER has received funding from the European Union’s Seventh Framework Programme under grant agreement no. 610456
How can we address the pressing need for energy efficiency of increasingly powerful computer systems? OPERA, a research project funded by the EU, aims to develop low and ultra-low power high-performance systems, covering the computing continuum through its work on servers and embedded boards tailored for surveillance/video applications. OPERA regards the growing importance of these systems as a way to respond to societal and industrial needs.

The consortium intends to deliver a solution which will reduce energy consumption by up to two orders of magnitude as compared to the state-of-art in 2013.

As shown in the diagram, energy efficiency can be achieved only by making all of the components of the computing continuum energy efficient. To address this challenge, OPERA wants to implement a high-density server which provides high-performance within a low power context, by exploiting heterogeneity. In addition, ultra-low power embedded devices will be developed around many-core processing and reconfigurable antenna technologies.
Innovation Europe

Three real-life case studies, all of which having important social and economic impacts, will show the energy efficiency and scalability of the proposed solutions:

1. providing a **virtualized desktop platform** to a very large number of thin clients;
2. developing a **highway traffic monitoring solution** with emphasis on detection of potential risks;
3. supporting operations of the Italian national emergencies and disasters agency by delivering **IT services on a truck**.

To achieve these goals, OPERA aims to exploit heterogeneous architectures by integrating ARM and Intel X86_64 micro-servers within the same HPE Moonshot-derived server chassis. Hardware acceleration will be delegated to FPGA modules, to offload networking functions as well as to accelerate computing tasks.

OPERA recognizes that programmability represents the major limitation factor in the adoption of such heterogeneous architectures. To solve this issue, OpenCL will be extensively adopted, along with an intelligent way of breaking the workload down into tasks that can be assigned to the most appropriate computing element for the execution, to maximize the energy efficiency of the whole system.

The whole system efficiency will be maximized by letting the external hardware appear cache-coherent with the POWER8 processors (used to run highly computational demand tasks) through the implementation of the Coherent Accelerator Processor Interface (CAPI) protocol. Ad-hoc policies will be embedded into the system to enable global monitoring of the energy efficiency, as well as to drive the task allocation.

OPERA is seeking to foster new collaborations with other initiatives. The consortium can be contacted by email: info@operaproject.eu

**NAME:** OPERA - Low Power Heterogeneous Architecture for Next Generation of Smart Infrastructure and Platforms in Industrial and Societal Applications

**START/END DATE:** 01/12/2015 – 30/11/2018

**PARTNERS:** STMicroelectronics (Italy, Coordinator), Hewlett Packard Centre de Compétences (France), IBM Israel Science & Technology (Israel), Istituto Superiore Mario Boella (Italy, Technical Coordinator), Nallatech (UK), Technion – Israel Institute of Technology (Israel), Certios (Netherlands), Département de l’Isère (France), Teseo Spa Tecnologie e Sistemi Elettronici ed Ottici (Italy), CSI Piemonte (Italy), Neavia Technologies (France)

**BUDGET:** € 6.5M

**WEBSITE:** www.operaproject.eu

The OPERA project has received funding from the European Union’s Horizon 2020 Programme under grant agreement no. 688386.

---

**GOOD TIMING: PROXIMA RESULTS TO HAVE WIDE-RANGING IMPACT**

The EU FP7 IP PROXIMA project has recently concluded having worked for the last three years on the development of probabilistic software timing analysis and tools for multicore platforms. Its goal has been to reduce the cost of software timing verification for mixed criticality and multicore systems.

PROXIMA focused on two main areas of work: timing analysis tools using probabilistic techniques to predict program timing behaviour; and methods to improve testability of real-time systems using injection of randomization into the timing behaviour of certain hardware/software. Covering both customized hardware designs and COTS technology, PROXIMA has been applied to several platforms including LEON3, AURIX and P4080.

It is having significant impact both at academic and industry level. The final PROXIMA Industrial Workshop attracted more than 40 expert attendees from academia and industry, representing automotive, avionics and space sectors and showing keen interest in the technologies presented. Several exploitable technologies and success stories resulted from the project; their impact will gain further strength as the PROXIMA concepts continue to expand their reach within industry. Project participants have been invited as keynotes or invited talks to 39 scientific and industrial events to disseminate the results which have been published in over 54 places (including 7 journal publications and 3 best paper awards). The most cited paper in the last five years from a flagship time-critical embedded systems conference - Euromicro Conference on Real-Time Systems - is one of the seminal papers of the PROXIMA technologies.

The main PROXIMA results are:

1. A set of hardware design principles, some implemented at FPGA level, showing how to achieve analysable software timing behaviour using randomization of hard-to-analyse resources and the removal of jitter from other resources.
2. The design of several software approaches for COTS architectures in which analysable software timing is achieved through a set of specialized randomization libraries that allocate program code and data at random memory locations across runs.
3. The development of several probabilistic timing analyses to provide tight and reliable probabilistic worst case execution

---

1. [https://scholar.google.co.uk/citations?hl=en&view_op=list_hcore&venue=tzuNspfpsn4J.2016](https://scholar.google.co.uk/citations?hl=en&view_op=list_hcore&venue=tzuNspfpsn4J.2016)
Innovation Europe

breakthroughs placing Europe at the forefront of research in the areas of approximate and error-resilient computing.

Significance-Based Computing for Reliability and Power Optimization (SCoRPiO) sought to introduce a new computing paradigm that exploits computational significance to design systems that are energy efficient and scale gracefully under hardware errors by operating below the nominal operating point, in a controlled way, without inducing fatal errors. This is possible because, in several application domains, not all computations and data are equally critical, with not all requiring to be performed or maintained at 100% accuracy. For example, in big data analytics, a user would be interested in statistical rather than bit-exact accuracy, in which case, a fast approximation would be preferable to slow and precise computations.

The SCoRPiO project followed a vertical, collaborative approach to the emerging areas of significance-based and approximate computing facilitated by the interdisciplinary set of skills of the consortium members. Using mathematically rigorous analysis based on interval arithmetic, algorithmic differentiation and compiler analysis, RWTH Aachen and CERTH developed dco/scorpio, a toolflow to automatically detect task-level code significance, and rank the contribution of each task to the quality of the final output. Dco/scorpio methodology effectively matches decisions of a domain expert in significance characterization for a set of benchmarks proving the potential of automating significance extraction. This significance information is expressed at the code source level using a task-based, OpenMP-like programming model developed by CERTH. This programming model conveys significance information to the runtime software, developed by QUB, which orchestrates all matters pertaining to the significance-aware execution management of the application tasks on top of a multicore platform. Analytical work by INRIA corroborates these experimental findings. The consortium evaluated this framework in an x86 multicore platform, using controlled fault injection to emulate timing errors due to aggressive voltage downscaling. An average energy dissipation reduction of 26% with minimal output quality degradation was achieved due to the Central Processing Unit undervolting to the point of eliminating design guard bands.

EPFL and IMEC worked extensively on fault modelling of static and dynamic memories and the OpenRISC CPU. These models were used to evaluate a plethora of hardware techniques to enhance the reliability of embedded many-core systems. For example, EPFL introduces the novel concept of predictive instruction-based dynamic clock adjustment to curtail dynamic timing margins in pipelined microprocessors. It exploits different timing requirements for individual instructions to selectively adjust clock periods on a per instruction granularity leading to an increase in speed of 38% or to a reduction in power

**SUCCESSFUL CONCLUSION TO SCORPIO PROJECT**

Coordinated by Nikolaos Bellas, Center for Research and Technology, Hellas (CERTH), a world-leading team of researchers across Europe is celebrating the successful conclusion of the three-year FET-Open project that resulted in a number of advances and
In today’s increasingly connected world, data-driven applications are gaining more and more importance. Tracking vehicle location accurately and in real time via sensors, or producing detailed meteorological simulations which allow us to predict weather events, are examples of data-hungry applications which can help improve efficiency, protect people and places or increase competitiveness. Simultaneously, more and more people, both individuals and businesses, are using the cloud to power their computing services, turning to infrastructure and platform providers to give them on-demand, reliable computing and storage for their day-to-day work.

Processing the vast amounts of data needed for such a variety of applications requires a new generation of data centre. Enter the M2DC project, which is developing turnkey appliances tailored to meet the requirements of different application domains which can be easily configured, produced, installed and maintained. This new class of low-power, specific purpose appliances offers built-in efficiency and dependability enhancements, will be easy to integrate with a broad ecosystem of management software and will be fully software defined to enable optimisation for a variety of demanding future applications in a cost-effective way.

Underlying these appliances will be the M2DC flexible server architecture, comprising heterogeneous hardware including ARM central processing units (CPUs) and field-programmable gate arrays (FPGAs). This architecture will enable customisation and smooth adaptation to various types of data centre, while...
advanced management strategies and system efficiency enhancements will be used to achieve high levels of energy efficiency, performance, security and reliability.

M2DC will demonstrate turnkey appliances tailored to meet requirements from various application domains such as photo finishing system serving (more scalable photo finishing), IoT data processing (data analytics for vehicles’ sensors), cloud computing (enhanced IaaS, PaaS solutions exploiting heterogeneity) or even high performance computing (efficient meteorological simulations).

NAME: Modular Microserver Datacenter
Start/End date: 01/01/2016 – 31/12/2018
KEYWORDS: energy-efficiency, low power microservers/processors, supercomputing, data centres
PARTNERS: PSNC (Poland), ARM (UK), Christmann (Germany), CEA (France), Huawei (Germany), Bielefeld University (Germany), OFFIS (Germany), XLAB (Slovenia), Vodafone (Switzerland), Politecnico di Milano (Italy), CEWE (Germany), Beyond (Poland), ReFLEX (France), AS+ (France)
BUDGET: €7.9M
WEBSITE: http://m2dc.eu/en/

The M2DC project has received funding from the European Union’s Horizon 2020 Programme under grant agreement no. 688201.

LOOKING AHEAD TO IMPROVED AND FLEXIBLE COMPUTER VISION: EYES OF THINGS

EYES OF THINGS

Halfway into the 3-year work plan, Horizon 2020 project Eyes of Things (EoT) is advancing towards a flexible low-power embedded vision platform. Mass-market mobile devices owe much of their success to their significant imaging capabilities. But could they be used as “eyes everywhere”? Vision is a device’s most demanding sensor in terms of power consumption and required processing power and so EoT’s objective is to build an optimized core vision platform that can work independently and be embedded into all types of device.

Since work began, two prototype boards have been developed; at present, the design of the final factor-form board is being finalized and manufacturing will start soon. The board is based on Movidius’s ground-breaking Myriad 2 SoC. The Myriad 2 chip has been specifically designed for high performance and low power. The EoT board also includes a low-power miniature camera, WiFi, IMU, flash memory, micro SD card, audio, JTAG, USB, GPIO, UART and MIPI connectors. An advanced PMIC unit is also included to allow for battery operated use and battery recharge.

Software for the platform has also been finalized. Communication with the platform via WiFi is based on the MQTT protocol. MQTT is a lightweight messaging protocol that also contributes to saving power. The EoT device has two operating modes. In Mode 1 the device expects and processes configuration commands from a client (typically a PC, tablet or smartphone), including the possibility to upload a user application to the device. In Mode 2, the user application is executed. A physical switch is used on boot to select between Mode 1 and 2. The developed software also includes libraries for computer vision (OpenCV, libccv, CNN, Quirc and native support for using Google Cloud Vision API), motor control, video streaming, audio input/output and support for scripting with the MicroPython language, including a remote IDE.

Since the platform contains a small, wearable camera, ethics and privacy issues have been taken into account. A number of features has been added to help in this direction, such as face blurring, secure communication protocols and an encrypted filesystem for the SD card. As the platform nears completion, the project consortium is exploring routes to market. An Early Adopters Program is planned for late 2016.

NAME: EoT - Eyes of Things
START/END DATE: 01/01/2015 – 31/12/2017
KEYWORDS: computer vision, embedded vision, deep learning, wearable camera, smart camera
PARTNERS: University of Castilla–La Mancha (Spain); AWAIBA (Portugal); Movidius (Ireland); THALES (France); DFKI (Germany); Fluxguide (Austria); Evercam (Ireland); nVISO (Switzerland). Partners Movidius, UCLM and THALES are HiPEAC members
BUDGET: €4.9M
WEBSITE: www.eyesofthings.eu

The Eyes of Things project has received funding from the European Union’s Horizon 2020 Programme under grant agreement no. 643924.
HIPEAC JOBS: MAKE IT WORK FOR YOU

Since HiPEAC was formed over a decade ago, we have learned a lot and one of the things that we have seen again and again is that organisations looking to recruit new staff in computing systems often have problems finding candidates with the specialist skills that they need. Likewise, researchers and engineers who have worked hard to develop skills and expertise in very specific areas often experience difficulties in finding out where and when suitable jobs are available.

HiPEAC Jobs therefore brings recruiters and candidates together through our online portal: www.hipeac.net/jobs/
The service is free of charge and is available to both members and non-members of HiPEAC.

**How does it work for recruiters?**
Log in to the HiPEAC website, go to the Jobs page and click on ‘Add a new job/internship’.
Fill in a simple form with the details of your Europe-based computing systems vacancy and your job will be advertised on the portal after a quick check by a member of our team.
• Your vacancies are made visible to thousands of computing systems researchers, doctoral students and professionals across the world through the portal and via our mailing lists, social media, events and newsletters;
• Post vacancies easily and for free. Non HiPEAC members just need to create an account on the HiPEAC website, which is quick, easy and costs nothing.

**How does it work for researchers?**
Just go to the portal and browse all vacancies or use the filters to search for roles according to location, area of technical expertise or career level.
• View vacancies relevant to your skills in both industrial and academic organisations and discover which organisations recruit in particular fields of expertise;
• No need to create usernames or accounts; click on the vacancy advert to apply directly to the recruiter.

HiPEAC Jobs will be at HiPEAC 2017, 23-25 January, Stockholm
www.hipeac.net/2017/stockholm
How has a start-up SME in the world of embedded systems made use of society and the economy’s growing reliance on automation, smart devices and data? Philippe Manet and Bertrand Rousseau of Embedded Computing Specialists explain.

Embedded Computing Specialists (ECS) was founded in early 2013 by Philippe Manet and Bertrand Rousseau. We decided to launch the company after obtaining our Ph.D. in embedded computer architecture from Université Catholique de Louvain in Belgium. We have been part of the HIPEAC community since its beginning.

ECS offers a set of competences and solutions that accelerates and secures the outcome of challenging R&D projects in computing systems and embedded electronics. Through ECS, we work in close partnership with our customers to help them build new innovative products, to improve their competitiveness and to extend their activity to new markets. Embedded electronics is everywhere, and the ever-growing need for more connectivity and automation means that it is becoming consistently more prevalent year on year. As R&D experts in embedded electronics, we are therefore active in many different markets.

For instance, in the broadcast industry, we helped a customer to build new storage solutions with a very high level of performance. Thanks to our collaboration, their new product is capable of reliably recording and playing hundreds of simultaneous low-latency high-definition video streams, all within a single computer.

In the railways industry, we designed the next generation of our customer’s M2M (machine-to-machine) computer that monitors locomotives’ health and relay status information directly to a business back-end hosted in the Cloud. For this, we created a complete computer architecture capable of processing thousands of parameters in real time in order to achieve preventive fault detection and maintenance.

Since its beginning, the main goal of ECS has been to bring the best of scientific-level expertise and methodology in embedded systems and computer architecture to industry. We use the substantial experience that we acquired through our Ph.D. to build and conduct efficient R&D strategies. We make use of high-value tools such as modelling, benchmarking and validation to ensure rapid development of new solutions. This approach has quickly given rise to strong interest from among our customers who are in a constant race to build new products by integrating complex hardware platforms and software stacks.

As a company, we decided to adopt a strong “business-first” attitude from the very beginning. We were very keen to avoid the trap that many high-tech start-ups fall into by focusing too much on the technology and not enough on the market. This is what led us to be present as soon as possible on the market; starting off as a consultancy company rather than only investing in our solutions and therefore spending a significant amount of time chasing funding. This path has proven to be successful since it helped us to get to know what the real needs of industry are and to learn how to build value together with our customers and grow both our revenues. Looking to the future, ECS aims to launch its own integrated embedded computer and OS to support the future needs of connected embedded applications including M2M and industrial IoT.

Do you work for an innovative technology SME? Contact communication@hipeac.net with your story.
In this edition’s Industry Focus, we take a look at how two collaborations between industry and academia have resulted in innovations that have the potential to lead to significant reductions in power consumption in systems ranging from huge data centres to small wearable consumer gadgets. Energy efficiency is a key factor in many industries’ ability to offer better, more competitive services and products. Here is some of what HiPEACers are doing to help.

**Movidius:**
leveraging EXCESS innovations for up to 50x power efficiency

Movidius, world leader in ultra-low power machine vision technology has implemented a new set of holistic power optimization tools to drive embedded performance and therefore bring significant power savings in battery-powered devices. The EXCESS project, an EU funded research program co-developed with Movidius a set of tools that allow programmers to optimize hardware and software simultaneously, bringing major power savings for devices that are reliant on battery power.

Energy consumption is one of the key challenges of modern computing, whether for wireless embedded client devices or high performance computing centres. As compute power increases, so too does energy consumption. The ability to develop energy efficient software is crucial, as the use of data and data processing keeps increasing in all areas of society. The need for power efficient computing is not only due to the environmental impact, but is also necessary to be able to deliver the ever increasing performance requirements.

To address this need for more efficient computing paradigms, Movidius in conjunction with the EXCESS EU research project led by HiPEAC partner Chalmers University of Technology developed a framework based on four pillars:
- Unified programming models
- Energy efficient libraries and algorithms
- Runtime & monitoring for energy analysis
- Hardware simulators for energy optimization

---

**Embedded computing for small-scale customised devices & applications (Movidius embedded platforms)**

**Execution models expressing key factors for energy efficiency**

**HPC computing for large-scale HPC applications (USTUTT- HLRS benchmarks & applications)**

**Iterative software/hardware co-design**
Industry focus

The resulting software tools allow programmers to develop hardware and software in a holistic manner from day one, yielding up to 50 times the efficiency of previous approaches. “The approaches developed in EXCESS have allowed Movidius to implement much more energy efficient algorithms in our commercial offerings” says Movidius VP of Engineering Brendan Barry. “This means our customers are getting maximum performance per Watt of power in the very smallest of mobile devices. Keeping power efficiency at the forefront throughout the development process means we’re able to deliver levels of machine intelligence unseen before in all sorts of consumer devices.”

As the commercial partner of the EXCESS project, Movidius has been using the technology to discover alternative methods to achieve low-power solutions, ultimately enabling more intelligence in devices such as drones, wearables and robots without sacrificing form factor or battery life.

EXCESS received funding from the EU Seventh Framework Programme under grant agreement number 611183.
Read more at www.excess-project.eu

Embecosm: Towards HPC and data centres which don’t cost the earth

With support from Innovate UK through the TSERO project, SME Embecosm has access to massive supercomputers, so why did they build their own from small board computers? HiPEAC member Dr Jeremy Bennett, Chief Executive of Embecosm, and Dan Gorringe, student at Brockenhurst College, explain.

The TSERO (Total Software Energy Reporting and Optimization) project is a collaboration between industry (Embecosm, Allinea and Concertim) and academia (STFC Hartree Centre, University of Bristol and Brockenhurst College), funded by Innovate UK, the UK government’s innovation agency. The aim of the project is to create an end-to-end system which significantly reduces energy used in HPC, using Hartree’s supercomputer facilities to demonstrate its effectiveness. Embecosm’s role in this is to develop compiler technology which can generate energy efficient code.

This project follows on from MAGEEC (Machine Guided Energy Efficient Compiler), which aimed to improve energy consumption for small embedded systems. Under the TSERO project, Embecosm is building on the ideas developed in MAGEEC to produce compiler tools for HPC systems, which can generate code optimized for energy efficiency. While the initial work is focused on HPC, the techniques are directly applicable to data centres. Using better compiler technology has the potential to cut energy consumption by 20% and, if applied nationally, could reduce the US data centre electricity bill by US$1.52 billion every year. Meanwhile in Europe, while there is limited information on the size of the data centre industry, the economy is comparable in size to that of the US, and so potential savings of more than €1 billion every year would be a reasonable estimate. The potential corporate beneficiaries worldwide include well-known names such as Google and Amazon, which, between them, spend $1 billion on electricity for data centres each year.
"Google and Amazon between them spend $1 billion on electricity for data centres each year"

For MAGEEC, Embecosm used a low cost board, to which energy probes could be attached (http://mageec.org/wiki/Workshop). A set of benchmarks with many different compiler settings were then run on these boards, creating a huge database of energy usage for different compiler settings. Machine learning, when applied to this data, allows the compiler to "learn" which are the best options for energy efficiency for a particular type of program. In TSERO, it is not permitted to intercept the power supply on processors and insert our own probes into multi-million dollar HPC systems. We are instead using commercial software from TSERO partner Allinea, specifically its MAP software, to supply the energy data. This uses interfaces such as Intel RAPL to acquire information about energy usage.

We have found that, while tools like Allinea MAP are extremely powerful, they do not provide some of the very detailed insight that can be obtained using a direct probe. Intel RAPL, for example, reports energy every millisecond, whilst the low cost probes used in MAGEEC sample twice per microsecond.

The solution is to build our own low cost HPC system, to which we can attach energy probes, and which we can afford to risk breaking. Dan Gorringe joined Embecosm from Brockenhurst College to construct a cluster of SBCs (single board computers). This also allowed us to fulfil a secondary objective of the TSERO project, which is to encourage the next generation of HPC engineers. Dan is about to start his final year of A-level (pre-university) studies in maths, further maths and physics, and hopes to go on to study engineering mathematics at university. Jeremy Bennett of Embecosm notes that a common complaint from organizations running HPC facilities is the difficulty of finding staff with the requisite skills and ability; there is ever increasing demand for very high performance systems, whether AI systems performing medical diagnoses, banking back offices minimizing global financial risks or the next generation of AI-based consumer facing software.

The result of this technical task is a Beowulf cluster of single board computers, connected by Gigabit Ethernet (see image). Individually, the boards used are relatively cheap and so easy to replace if broken, but are also easier to record data from. The cluster was built from sixteen Pine64s, each of which has an Allwinner A20 SoC with four 64-bit ARM processors and 2GB of RAM making, in total, sixty-four cores. These boards are very similar to the Raspberry Pi 3 but, unlike the Pi, they support 2GB of RAM and Gigabit Ethernet. The total cost of the system is well under $1000, the design is open to anyone and is documented in detail on the TSERO public wiki (http://tsero.org/wiki/Embecosm_Pine64_Cluster).

We have measured the performance of this system using both Linpack and High Performance Conjugate Gradients (HPCG). In comparison to the large systems at STFC Daresbury, this 20GFlop cluster is insignificant in terms of performance, but it models the same type of system which means that energy measurement data is comparable and therefore the conclusions we’ve been able to draw from it are also applicable to Hartree’s supercomputers. It will form a key part of our ongoing research into energy efficient compilers for high performance computer systems.

www.embecosm.com/  
www.tsero.org  
www.mageec.org
In 1961, Ralph Landauer, then working at IBM, published a paper where for the first time ‘information’, usually considered a purely mathematical quantity, assumed a role in physics. Specifically, Landauer’s paper aimed at identifying the minimum energy required to do computation using standard thermodynamics.

Landauer initially focused on a specific operation, today known as the ‘Landauer reset’, that consisted of putting a binary switch – which can be in each of the two possible logic states 0 or 1 – into a given logic state. Such an operation is sometimes interpreted as ‘information erasure’ because it reduces the amount of information that can be associated with the binary switch: before there are two possible states, whereas after the operation there is only one possible state. According to thermodynamic law, such a reduction in the number of available states for a physical device requires a minimum energy expenditure, easily computable using previous work done by Boltzmann.

In the same paper, Landauer generalized the result associated with the reset operation to cases where there was a decrease of information between the input and the output of a computing system. This is the case of so-called logically irreversible devices. Landauer wrote: ‘We shall call a device logically irreversible if the output of a device does not uniquely define the results. We believe that devices exhibiting logical irreversibility are essential to computing. Logical irreversibility, we believe, in turn implies physical irreversibility, and the latter is accompanied by dissipative effects.’ [IBM Journal of Research & Development, Vol. 5, No. 3, 1961]

In fact, most standard logic operations performed on ordinary computers show ‘logical irreversibility’. This is the case, for example, of the conventional ‘OR’ gate, where we have two bits at the input and one bit at the output. In this situation, simply knowing the value of the output is not enough to infer the actual value of the inputs.

Soon after the publication of Landauer’s paper, other scientists worked to deepen and extend Landauer’s principle to more general aspects of information processing. The most important result from this work is attributed to Charles Henry Bennett, also at IBM. In 1973 he published a paper titled ‘Logical reversibility of computation’ in which he proposed to introduce a model of computing, i.e. new devices, where there was no information decrease between the input and output of any logic operation. The motivation that led Bennett to introduce logically reversible operations was to overcome the minimum energy expenditure introduced earlier by Landauer.

Most notably, this limit was generally attributed to all logical irreversible devices, including traditional logic gates like ‘OR’, ‘AND’ and ‘NAND’. Landauer and Bennett’s works did not go unnoticed and a significant amount of scientific literature was produced to oppose or support the existence of such a minimum limit. It is no exaggeration to say that for more than 40 years this topic has been considered highly controversial.

An experiment has now put an end to this controversy. It clearly shows that there is no such minimum energy limit and that a logically irreversible gate can be operated with arbitrarily small energy expenditure. Simply put, it is not true that ‘logical irreversibility … implies physical irreversibility’ as Landauer wrote.

The experiment, carried out by scientists in the NiPS Laboratory at the University of Perugia and recently published in Nature Communications, aimed at measuring the amount of energy dissipated during the operation of an ‘OR’ gate. It shows that the logic operation can be performed with an energy toll as small as 5% of the expected limit. The paper concludes that there is no fundamental limit and reversible logic is not required to operate computers with zero energy expenditure.
What are the implications of this discovery? The ‘OR’ logic gate used by the scientists at NiPS Laboratory is created using a micro-electromechanical cantilever, acted on by electrostatic forces. Although this cannot be considered a promising new technology for substituting the energetically expensive transistors that make our computers today, the importance of the experiment is in the demonstration that there is no limit to how much we can lower energy consumption during computation. This will change our understanding of the energy dissipation process and push research forward.

We expect this result to impact future developments in at least the following aspects:

• It will promote research towards ‘zero-power’ computing: the hunt for new information processing devices which consume less and less energy. This is of strategic importance for the future of the whole information and communication technology sector, which was to deal with the problem of excessive heat production during computation.

• It will call for a profound revision of the ‘reversible computing’ field. In fact, the experiment means that one of the main reasons for this field’s existence – the presence of a lower energy limit – now no longer applies.

---

**ICT-Energy: reducing energy consumption across the ICT sector**

The LANDAUER project forms part of ICT-Energy, a Coordination and Support Action bringing together two EU-funded research communities focusing on energy consumption in information and communication technologies: ‘Towards Zero-power ICT’ and ‘Minimising energy consumption of computing to the limit’.

The ICT-Energy consortium comprises 10 partners, who between them coordinated 10 energy-related projects:

**LANDAUER**, coordinated by the University of Perugia
(01/09/2012–31/08/2015)
**Aim:** to test the fundamental limits in energy dissipation during the operation of physical switches representing the basic elements of logic gates.

**NANOPOWER**, coordinated by the University of Perugia
(01/08/2010 –31/07/2013)
**Aim:** to identify new directions for energy-harvesting technologies at the nanometre and molecular scale.

**ENTRA**, coordinated by Roskilde University
(01/10/2012–30/09/2015)
**Aim:** to promote ‘energy-aware’ software development using advanced program analysis and modelling of energy consumption in computer systems.

**Exa2Green**, coordinated by Heidelberg University
(01/11/2012–31/10/2015)
**Aim:** to develop new energy-aware computing paradigms for future exascale computing.

**ParaDIME**, coordinated by Barcelona Supercomputing Center
(01/10/2012–30/09/2015)
**Aim:** to create a processor architecture for a heterogeneous distributed system that delivers significant energy savings in data centres, for example.

**PHIDIAS**, coordinated by École polytechnique fédérale de Lausanne
(01/10/2012–30/09/2015)
**Aim:** to achieve minimal energy consumption in wireless body sensor networks by developing power-efficient biomedical sensor nodes.

**SENSATION**, coordinated by Aalborg University
(01/10/2012–29/02/2016)
**Aim:** to promote self-supporting systems by balancing devices’ energy harvesting with the energy consumption, through energy-centric modelling and optimization tools.

**TOLOP**, coordinated by Hitachi Europe
(01/09/2012–29/02/2016)
**Aim:** to research disruptive approaches to push significantly beyond existing power limits, by creating devices, evaluating their low-power capabilities and designing architecture to enable overall low-power circuit operation.
Computer hardware is a mess. It is faster, more power-efficient, and more capable than most could have conceived when Unix was written in 1970s. But the baroque complexity of the hardware/software interface is now a major, if largely unrecognized, obstacle to progress.

The software programming manual for a modern phone system-on-chip (SoC) runs to several thousand pages of register, interrupt, descriptor, and power documentation which is often ad-hoc, incomplete, and imprecise. Really well documented SoCs (they do exist), however, present a bewildering array of intersecting address spaces, heterogeneous cores, and complex interrupt routing and power control (which may or may not be configurable, or discoverable). Millions of lines of code in the Linux kernel, for example, are there to deal with this tangle of functionality.

However, once the operating system (OS) can correctly initialize, access, and program these devices, the problems are only starting. Getting optimal performance or energy efficiency out of modern hardware appears to be beyond the state of the art of systems software. For example, a paper this year shows the wasted CPU cycles in Linux due to the scheduler’s failure to adapt to different memory hierarchies.

The consequences of this are a huge programming effort spread across chip and device vendors, OS programmers, and application...
developers which nevertheless results in brittle, inefficient, and hard-to-maintain code which is, at best, optimized for a single point in the design space.

How did computer science get into this hole? We are supposed to be good at managing complexity. The responsibility lies with both hardware and OS designers.

Some complexity is inherent in the tremendous capabilities of modern hardware, and the need to extract performance from it. Operating systems, however, mostly stick stubbornly to the “everything’s a VAX” model of hardware. The weapon of choice in the fight against software complexity remains C. Some blame lies with the language community’s failure to create a better alternative for low-level programming (perhaps Rust or Swift). Nevertheless, OS designers have avoided most complexity-management techniques from any area of Computer Science.

However, much blame must lie with hardware folks, for whom OS design is a secondary consideration at best, often ignored completely. Bluntly, most hardware is poorly designed from a software interface perspective.

For example, the PCIe express standard does not specify an algorithm for online (re)configuration of PCIe bridges, and the problem is astonishingly difficult: IBM has a recent patent on genetic algorithms for this, and most OSes simply don’t even try to get it right.

This results in an impasse. Hardware designers carry on regardless. OS designers don’t intervene, and just write more C. Systems programmers are often good at critiquing a given device software interface (they know better than anyone), but often go silent if asked what it should look like.

What can be done? A first step is to explore the OS design space more. Multikernels, for example, are inherently better at handling core heterogeneity, partial coherence, non-shared memory, etc. Another is to embrace more well-known techniques. Constraint satisfaction, for example, turns out to handle PCI configuration quite well (in particular dealing elegantly with hardware bugs and “quirks”). Logic programming is a disciplined way to express policies that are light years ahead of text files or hardcoded C.

However, what we really need are frameworks for formally modelling hardware. Interrupt routing, for example, is a good match for recent techniques used to model forwarding in data networks, and the multiple layers of physical addressing in modern systems also turn out to resemble this model.

To be useful, such a formal model should capture all existing hardware features. Ideas like Device Trees are a step in the right direction, and show that there is a short-term value in such descriptions, but they lack any well-defined semantics.

A model, with a suitable syntax, has immediate benefits: OS code can be autogenerated, and the OS can reason online about devices and cores to better optimize for performance.

However, in the longer term, we can start to make judgements about which statements in this language are “tasteful”, and which are not. From this, we can finally derive principles for good design of hardware from an OS perspective. Such a language would finally enable OS designers to explain what they want to hardware people.

Moreover, it would finally put the field of operating systems proper – how low layers of software manage the hardware resources of the machine – on a sound logical foundation. This is an exciting time.

HiPEAC Member Timothy Roscoe is a Full Professor in the Systems Group of the Computer Science Department at ETH Zurich. He received a PhD from the Computer Laboratory of the University of Cambridge, where he was a principal designer and builder of the Nemesis operating system, as well as working on the Wanda microkernel and Pandora multimedia system. After three years working on web-based collaboration systems at a start-up company in North Carolina, Mothy joined Sprint’s Advanced Technology Lab in Burlingame, California, working on cloud computing and network monitoring. He then joined Intel Research at Berkeley in April 2002 as a principal architect of PlanetLab, an open, shared platform for developing and deploying planetary-scale services. After four months as a visiting researcher in the Embedded and Real-Time Operating Systems group at National ICT Australia in Sydney, he joined ETH Zurich in January 2007. His current research interests include network architecture and the Barrelfish multicore research operating system.

He was recently elected Fellow of the ACM for contributions to operating systems and networking research and was a course teacher at HiPEAC ACACES Summer School 2016.
Can you tell us a bit about your professional background? How did you come to be working at ARM and where have you worked before?

I graduated in Electronic Engineering from the University of Genova in Italy, where I also did my PhD in High Level Synthesis for VLSI Systems, during which I spent one and a half years at IMEC laboratories in Leuven, Belgium in the Cathedral 2nd team lead by Gert Goossens. After my PhD I spent some time at the University of Genova, and then I joined ST-Microelectronics where I remained for the next 12 years in different locations: initially in Grenoble, France for 5 years working on compilers for Application Specific Instruction-set Processors (ASIPs) in the group of Pierre Paulin (well known in the high-level synthesis community for the widely applied Force Directed Scheduling algorithm), then I moved to Boston, USA for 2 years to work on a joint project between ST-Microelectronics and the HP-Labs group led by Josh Fisher (among the pioneers and influential proponent of Very-Long Instruction Word (VLIW) processors). I then moved to Lugano in Switzerland to found and lead the ST-Microelectronics Advanced Computing Lab. In 2007 I joined ST-Ericsson in Milan, Italy as Fellow of Advanced Computing. When ST-Ericsson closed in 2013 I moved to ARM in Cambridge, UK, where I'm currently working on compilers for next generation GPU architectures.

Describe a typical day at work for you.

When I joined ARM, I moved back to my original passion for compiler development, so I spend a lot of my time developing software. Beyond development I interact with hardware/system architects with whom I closely cooperate for the development of next generation GPUs (Graphic Processor Units), and I also spend a considerable amount of time learning, given the very wide context of Graphics.

What’s the best part of your job? Is there anything that you’re less keen on?

I have the privilege to contribute directly to what I find among the more interesting topics in computing today – GPUs – and in one of the most successful companies in computing in general - ARM – meaning that what I do ends up literally in the hands of hundreds of millions or even billions of people, which is obviously very gratifying.

The downside of being focused on challenging software developments is that it often gets too focused and narrows on specific aspects and for long periods of time. Sometimes I’m tempted to move back into a wider, yet less technical role, but up to now in ARM I’ve rather kept taking advantage of the too-interesting-to-refuse technical challenges.
What are you most proud of in your career?

I have covered different roles in my career. As a technical contributor, the work I've done in ARM has been the most intense and impactful so far, as I had the opportunity to start a new project from scratch as an initial investigation, which then had a major impact in production.

As a manager I’m particularly proud of having started and led the ST-Microelectronics Advanced Computing lab in Lugano Switzerland. We managed to attract brilliant researchers from across the world, including some from HP Labs like Giuseppe Desoli and Stefan Freudenberger, others from STM like Erven Rohou and Benoit Dupont de Dinechin, and several others. It was a dream-team – too bad that it didn’t last long enough to fulfil its potential... This has also been the period where I managed to establish a lot of fruitful cooperations with several universities and with HiPEAC.

As a technology strategist in ST-Ericsson I led the introduction of the first ARM-based multiprocessing (SMP) solution - a very challenging task at the time when nobody was ready to adopt multi-processing in mobile phones. It seems strange now that it is mainstream to have at least 8 processors in your smart phone, but the tough part back in 2007 was to convince the industry to move from one to two...

Beyond the specific achievements, what I’m particularly proud of is that I had the privilege to work or interact with some of the most influential personalities in computing, and HiPEAC has been instrumental in establishing some of these relationships. If I had to choose a common pattern that I've learned from these “giants”, it is to always strive to distill the fundamentals in what we do, which often leads to beautiful and elegant simplifications – the alternative is to drown in complexity...

You’ve been part of HiPEAC since the beginning. What career advice would you give to the students coming up through the network?

I find it very difficult to give “career advice” to others because each context is so different, plus my evolution was driven mostly by my technical interests more than a “career optimisation” process. I’ve been very stubborn in pursuing my technical interests in computing and compilation even in situations when it was not a main priority from a wider business perspective. I sincerely cannot decide if it has been a limitation on my side to be unable adapt to evolving situations, or a more “noble” drive for a genuine technical interest – probably both - the fact is that I’ve rather changed jobs (and countries!) to continue to work in computing, rather than adapting to other roles that might have been more profitable career wise in the short term.

What matters I think is to understand what we are good at, which often coincides with what we like to do, and try to do it and keep getting better at it... There is nothing worse than a good technician who is forced into management for career purposes, or vice versa a potentially good manager trapped in a technical role for lack of opportunities... We spend so much of our time and energy at work that if we don’t do what we like we are bound to be unhappy. So my advice: find out what you like and try to adapt your career accordingly, rather than the other way around. In the end you’ll likely be happier and, as a less important side effect, you might even get a better career.

WHERE WILL YOUR CAREER TAKE YOU NEXT?

Where will your career take you next? Do you have a job that you’d like to advertise to people with the right skills? Check out or add to the numerous job opportunities on the HiPEAC jobs portal: www.hipeac.net/jobs

If you’re passionate about your career and would like to share it with the HiPEAC community, we’d love to hear from you. Email communication@hipeac.net with your story.
HiPEAC internships are a great way to get a funded work experience placement at a leading technology company. They provide an excellent opportunity to build long-lasting networks with experts in your field working outside academia. Interested in applying? Further information is available on the HiPEAC website: www.hipeac.net/mobility/internships

HiPEAC internships: a new dimension for your PhD

NAME: Victor Garcia
RESEARCH CENTRE: Universitat Politècnica de Catalunya / Barcelona Supercomputing Center
HOST COMPANY: ARM Research, UK
DATE OF INTERNSHIP: 01/10/2015 - 31/01/2016

BUILDING LASTING LINKS

I am a fourth year PhD student in computer architecture. My main research interests fall within the memory hierarchy, on topics such as data prefetching, memory consistency and cache coherence, especially in heterogeneous architectures. Thanks to a HiPEAC industrial PhD internship, I spent four months at ARM Research in Cambridge where I joined the Software and Large Scale Systems group.

Data prefetching has been proven to be an efficient technique to reduce the ever increasing “memory wall” by reducing the average access time that the cores of a processor need to wait for data from memory. In-order cores suffer especially from this problem, since accesses cannot be reordered to attempt to hide memory latency.

While prefetching is a well-known and extensively studied topic, prefetching on in-order cores in an HPC context has its own problems and limitations. Nowadays, most prefetching research assumes out-of-order cores that can hide the latency of a L1 data cache miss, and therefore focus on prefetching at the L2 cache. In-order cores on the other hand, cannot hide this latency and are therefore much more reliant on successful data prefetching.

The goal of the internship was to analyse the effect of data prefetching on in-order cores for a set of High Performance Computing (HPC) workloads, understand its limitations, and propose novel mechanisms or improve current ones, if possible.

My initial task was to evaluate current prefetchers on in-order cores and understand how suitable they are for HPC workloads. Next I implemented an oracle prefetcher to quantify the maximum performance achievable with data prefetching in our target architecture.

From the lessons we learned in these initial steps, I implemented a dynamic prefetching scheme to regulate the behaviour of the prefetcher based on a number of metrics obtained at runtime. This dynamic scheme was able to match, and sometimes even out-perform, the best static configuration, while reducing cache pollution and bandwidth usage.

With some further analysis of different prefetching implementations and these initial results, we expect to publish our findings in a relevant conference or journal.

Looking back, this internship was a great experience. I learned a lot and met many colleagues with whom I will continue collaborating in the years to come. I obtained really valuable first-hand experience of how industrial research is conducted, and how it differs from academia. I would like to thank HiPEAC for this opportunity, and the good folks at ARM Research who made me feel at home during those four months.

Victor’s positive sentiments about the value of the internship are echoed by his supervisor at ARM, Roxana Rusitoru: ‘Having Victor at ARM Research not only enabled us to pursue novel research areas (in-order core prefetching for HPC), but his expertise in prefetching brought new insights and helped deliver excellent results by the end of the internship. The work Victor started will be continued by future HiPEAC interns’.
If you are a doctoral student or junior post-doc researcher and your research would benefit from spending time at another institution in the HiPEAC network to learn from and work with a new team on key challenges in computing systems, a collaboration grant might be for you. Get more information at www.hipeac.net/mobility/collaborations

NAME: Konstantinos Maragos
RESEARCH CENTRE: National Technical University of Athens
HOST INSTITUTION: University of Manchester
DATE OF COLLABORATION:
20/09/2015 - 21/12/2015

WORKING TOGETHER TO ACHIEVE NEW INSIGHTS FOR MAINTAINING FPGA PERFORMANCE AS NODES GET SMALLER

name is Konstantinos Maragos and I am a PhD student at the National Technical University of Athens (NTUA). My research interests focus on reconfigurable and parallel architectures and variability in FPGAs. Thanks to a HiPEAC grant, I had a great three-month experience at the University of Manchester, which helped me to broaden my knowledge in the field of reconfigurable computing.

Nowadays, FPGAs have dominated a wide range of application domains, including data centres (e.g. cloud computing), medical (e.g. medical imaging) and defence (e.g. surveillance). They have experienced particular popularity in the last few years and have established a prominent market position. FPGAs provide a variety of key features, such as employment of highly parallel architectures, reconfiguration according to the environmental stimuli and low-power processing, rendering them a very attractive solution for a plethora of applications. However, as the technology node keeps scaling-down, it is becoming more and more difficult for performance advancements to keep up with the well-known Moore’s Law. This arises from the fact that the cmos technology reaches its physical and power-thermal boundaries which greatly impacts on the performance and reliability of the circuits.

The purpose of the collaboration was to study variability and power noise (IR-drop) phenomena in the present-day 2D and future 3D reconfigurable architectures, with the aim of developing novel approaches to address these issues. These phenomena become rapidly more pronounced as the technology node scales down.

For the purpose of our study, we used two Virtex 7 xc7vx485tffg1761 FPGA devices at 28 nm. We developed an infrastructure with on-line sensors on the FPGA fabric that allows the exploration of variability and power noise at run-time based on the measurement of the operating frequencies over various areas of the device. In our experiments, we used this infrastructure to investigate variability and power-noise by employing various computationally intensive designs implemented on the FPGA. The findings of this research give insights for the development of new novel techniques/algorithms in the context of electronic design automation (EDA) tools considering these phenomena.

I would like to express my sincere gratitude to HiPEAC for giving me the opportunity to collaborate with the Manchester group. I am deeply appreciative of the support and the great collaboration. Thanks to this grant I had the chance to exchange ideas about my research field with many worthy scientists and to carry out experiments on high quality equipment. It was a great pleasure for me to lay the foundations for cooperation between the two research groups (in Manchester and Athens) and I believe that this cooperation will lead to mutual benefits in the future.
HiPEAC futures

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

Three-minute thesis

NAME: Marta Ortín Obón
RESEARCH CENTRE: University of Zaragoza
ADVISORS: Víctor Viñals-Yúfera (Professor) and María Villarroya-Gaudó (Associate Professor)
THESIS: Networks-on-Chip: from the Optimization of Traditional Electronic NoCs to the Design of Emerging Optical NoCs

FEATURED RESEARCH: NEW STYLE NETWORKS-ON-CHIP FOR BETTER COMMUNICATIONS

As technology improves, memories and processors become faster, smaller, cheaper, and more energy-efficient, enabling computer architects to include more of them in a single chip. These memories and processors are nowadays used in a variety of devices ranging from smartphones, tablets, and laptops to supercomputers and data centres. Now that Moore’s Law is reaching its limit, the replication of simple cores is being used to continue improving performance while minimizing fabrication costs. As a consequence, performance now faces a bottleneck not only in computing power and memory access, but also in the communication of the chip elements. In this context, interconnection networks have emerged as the prevailing solution to provide fast, cost-effective and scalable communications. They are the key for the success of future digital systems, both chip multiprocessors composed of tens of identical cores and heterogeneous systems-on-chip.

The thesis starts by analysing the state-of-the-art of electronic networks-on-chip and detects that, even though the fundamental purpose of the interconnect is to exchange information among processors and memories, it is often designed and optimized without taking those essential components into consideration. Based on that observation, we propose a mechanism called Reactive Circuits that successfully leverages the information provided by the coherence protocol to dynamically build circuits for replies, reducing network latency and power.

As multiprocessors continue to scale, it is more challenging for electronic networks-on-chip to meet their communication demands within the power budget. As a consequence, silicon photonics are coming to the forefront with the objective of providing higher bandwidth and shorter latencies with reduced energy consumption. We present an algorithm to automatically generate the communication matrices for optical rings with any number of nodes and minimum number of waveguides and wavelengths, and calculate their power consumption. We also introduce the first complete network interface architecture for optical networks and demonstrate that it is responsible for most of the complexity of the optical NoC, both in latency and power.

We then analyse the feasibility of the optical interconnect technology when integrated into industry-relevant objects: a chip multiprocessor and a general purpose multicore accelerator. By performing an accurate cross benchmarking against an optimized electronic NoC, we determine that the optical network is better in latency and energy-per-bit, but still needs to be optimized in terms of static power. With the objective of tackling that issue as well as adapting the optical networks to the virtualization paradigm frequent in multicore accelerators, we design the first algorithm to partition an optical network while minimizing the number of used wavelengths, thus reducing power consumption.

Access the public online version for free:
Marta is now carrying out postdoctoral research in the Graphics and Imaging Lab at the University of Zaragoza.
Individual Fellowships fund researchers looking to enhance their career development and prospects by working abroad. This action is meant to support the best, most promising individual researchers from anywhere in the world. Researchers must have a doctoral degree or at least four years’ full-time research experience. Fellowships benefit not only the researcher but also the organisations which host them. See http://ec.europa.eu/research/mariecurieactions/ for more.

Postdoc funding focus: EU Marie Skłodowska-Curie Individual Fellowships

FROM EUROPE TO EUROPA: BUILDING THE NEXT-GENERATION SPACE-GRADE PROCESSORS AT ARM by Xabier Iturbe

The focus of my two-year Marie Skłodowska-Curie Fellowship was to pioneer research on innovative radiation and fault-tolerant ARM processors to enable their use in space missions. This research was conducted in collaboration with NASA’s Jet Propulsion Laboratory (JPL) - California Institute of Technology in Pasadena CA, where I spent my first year of the Fellowship from September 2014. During this time, I gathered performance, power budget, radiation hardness and fault-tolerance processor requirements while working with JPL’s on-board science instruments in the scope of a NASA mission to explore Jupiter’s moon Europa. This knowledge proved to be of utmost importance during my second year of the Fellowship at ARM Research Cambridge, where I worked in the design and assessment of a Triple Core Lock-Step (TCLS) processor to meet NASA’s requirements. The TCLS processor was built on the success of ARM Cortex-R5 CPUs and includes the capability to detect, tolerate and automatically recover from errors provoked by space radiation without interfering with any critical operation that might be being executed when the errors manifest themselves. As part of my Fellowship I also promoted the TCLS processor in technical conferences and space technology centres, and helped improve ARM’s position in the fault-tolerance market by generating Intellectual Property (i.e., patents) in the context of lock-step processors.

“\textit{This exciting experience has allowed me to strengthen my technical skills, develop management and leadership skills and create a network of contacts}”

During my Fellowship I was exposed to two different yet complementary environments. First, the interdisciplinary science environment at JPL-Caltech, where I collaborated with NASA’s engineers, physics, biologists and geologists to accomplish such a complex task as it is to send a probe to explore a distant planetary body. Second, the business-oriented high-technology environment at ARM, where I participated in the definition and development of a new processor solution. Overall, this exciting experience has allowed me to strengthen my technical skills, develop management and leadership skills and create a network of contacts both in Europe and the US. By the end of the Fellowship, my professional profile had turned out to be very interesting for ARM and the company has offered me the position of EMEA University Program Manager. Hence, from now on I will manage ARM technology outreach to universities and higher education institutions in the EMEA region.

Xabier’s Fellowship received funding from the European Union’s Seventh Framework Programme under grant agreement no. 627579
Scientific and technological expertise / exposure

❄ With 700 participants, and 80 companies present, the HiPEAC conference is the largest computing systems research event in Europe.

✓ A compelling programme and extensive range of associated events makes this a unique opportunity to connect with researchers, engineers and graduate students.

Strengthening industry-academia relations

❖ 20% industry participants, a figure set to increase.

❖ A global event, 20% of attendees have affiliations outside Europe.

❖ Media presence, auditorium for presentations, demonstrations and product announcements.

An excellent recruitment opportunity

❖ Graduate students represent up to 25% of attendees.

☐ Online job portal (hipeac.net/jobs) complemented by on-site recruitment booth and advance support with arranging interviews.


Tailored sponsorship plans

<table>
<thead>
<tr>
<th></th>
<th>Gold</th>
<th>Silver</th>
<th>Bronze</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>From €4,000</td>
<td>From €2,000</td>
<td>From €1,000</td>
</tr>
<tr>
<td>Free registrations for the conference</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Acknowledgement on website and media</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Presentation slot during the industry session</td>
<td>20 min.</td>
<td>20 min.</td>
<td>10 min. upon availability</td>
</tr>
<tr>
<td>Free mini-booth in the industry exhibition</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>Privileged booth location and customized booth</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eligible as a co-sponsor of student activities: travel grants, prizes, student poster session, best student presentation...</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

For even greater visibility

- Sponsor the student poster session reception or the social event.
- Hire alternative booth shells and more space.
- Extend your booth’s presence to two days or the whole conference.
- Target your donation towards subsidizing student attendance or promoting the involvement of underrepresented groups in engineering.

Contact us today to discuss how you can maximize your company’s visibility or make the most of recruitment possibilities.

<sponsorship@hipeac.net>

hipeac.net
hipeac.net/sponsorship
hipeac.net/twitter
hipeac.net/linkedin