HiPEAC info 62

JANUARY 2021

HiPEAC conference 2021 Virtual event

HiPEAC Vision roadmap 2021: taking a whole systems approach

Tulika Mitra on taking edge computing to the next stage

Evangelos Eleftheriou on the possibilities of in-memory computing

Brad McCredie on exascale and beyond

contents

IBM's Evangelos Eleftheriou on in-memory computing

AMD's Brad McCredie on HPC and exascale

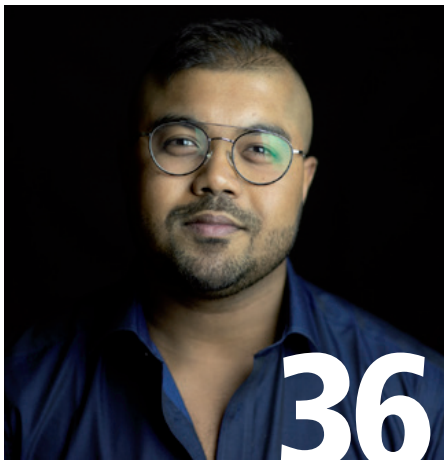
Tulika Mitra on the way forward for edge computing

20

The HiPEAC Vision 2021 technology roadmap

24

HiPEAC 2021 gold sponsor Huawei



36

Nosh: an app to manage changing food shopping habits

First of all, I would like to wish you a healthy and prosperous 2021, personally as well as professionally. The year 2020 was quite special, according to some an *annus horribilis*. I agree that it was a horrible year, but every dark cloud has a silver lining. In the past year, we learned many news kills, and several amazing things happened:

- For the first time ever, scientists were able to develop and produce hundreds of millions of doses of effective vaccines in less than ten months. This was completely unthinkable one year ago and it sets the bar for future developments.
- Central banks had been trying to replace cash transactions with digital payments for many years. Thanks to COVID-19, it happened overnight. I don't expect cash to make a serious comeback in 2021.
- Many European retailers saw e-commerce as a sideline to their brick and mortar shops. With the lockdown, some of them started experimenting with online sales. I am sure that many of them will keep it up after COVID-19.
- The computing industry tried for years to convince us to use videoconferencing for business meetings, with varying success. In 2020, we switched en masse to videoconferencing, and many of us have begun to appreciate the advantages.
- In 2019, few employers were encouraging employees to work from home. In 2020, they discovered that thanks to modern digital collaboration tools, remote working can be as good as office-based work. Employees discovered some of the advantages too. Telework will not disappear with COVID-19.
- Schools were forced to experiment with distance learning in 2020, and this has catapulted them in the 21st century. They are now investing in digital solutions, and many children have access to a laptop, which helps them with their school work. This might not disappear after COVID-19.

As a computing community, I believe we should therefore not feel too negative about 2020, but be hopeful that COVID-19 will lead to the permanent adoption of the technological solutions we have been working on for years.

This magazine comes with a copy of our biannual HiPEAC Vision. It explains how computing technology can help solving the grand challenges of the 21st century. I hope the HiPEAC community will take up these challenges, and create the technological solutions that will change the world for the better.

This is my new year's wish for 2021.

Koen De Bosschere, HiPEAC coordinator

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

hipeac.net

@hipeac hipeac.net/linkedin

HiPEAC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 871174.

Cover image: © Mon5ter dreamstime.com

Design: www.magelaan.be

Editor: Catherine Roderick

Email: communication@hipeac.net

The industrial policy of the European Commission is based on the twin green and digital transitions. Thirty years ago, these would have been two obscure and very different topics. European Commission Programme Officer Sandro D'Elia (DG CONNECT) explains that the green transition is now an urgent need, which is impossible to achieve without a robust contribution from digital technologies.

Two very different twins

My nieces are two wonderful small girls. One is dark-haired and has black eyes while the other is a blue-eyed blonde; if you look at them together, you would never think that they are twins. In this, they are just like the green transition and the digital transition.

We know that the green transition is not a choice: we have to reduce radically the footprint that our activities leave on the planet, and we need to do it quickly. The objective of the European Commission to make Europe a climate-neutral continent by 2050 is extremely ambitious and will be very difficult to meet, but it is a bare necessity to keep our planet in decent shape.

The digital transition is also an obligation because we know that, in order to keep the EU economy competitive, we need a lot of digital technology. We also need it, for example, to guarantee good healthcare for everybody, to reduce traffic congestion, and to make public services more efficient; all these objectives require a strong presence of digital technologies, but this is something we have known for a long time. The new element of the game, which emerged clearly only in the last few years, is that the digital transition is also a requirement to achieve the green transition.

"The competitiveness of European industry will be measured by its capacity to develop products and services which are both innovative and sustainable"

Green Deal

Let me clarify this point. The objective of the green transition is to reduce enormously our environmental footprint on this planet. This will probably mean quitting some of our most polluting habits, but we cannot go back to living like our ancestors did a thousand years ago. Therefore, the only possible strategy to make humans compatible with the planet we live on is to improve the sustainability of all the technologies which support our ways of life.

The interesting point is that today any complex machine, from the car to the boiler, the hospital ventilator or the manufacturing tool in the factory, has a computer inside. If we want to improve anything in any industrial or economic process, we have to use those computers inside the machines, and very likely design better and smarter computers that can make the best use of the data they have available.

Artificial intelligence (AI) can help us. It is very promising as a green technology because it can potentially help in any process where there is a lot of data to be managed. There are several applications we see today, including the optimization of the processing industry (food, oil, concrete production etc.), and preventive maintenance based on failure data (trains, industrial machinery etc.).

The most interesting aspect is that AI can help by enabling completely new solutions which are impossible with legacy technology, like generative design, autonomous robots, or climate change simulation. These are just a few of the many fields where AI can make a real difference.

Agriculture

As an example, let's look at agriculture. Currently, big agricultural machines work best on homogeneous soil and for extensive cultivation; this generates a loss of biodiversity and has, in general, a negative impact on the countryside. Smarter and smaller agricultural machines with a high degree of autonomy could work on uneven and difficult soils, reducing the impact on the ecosystem and adapting better to local needs and smaller productions, e.g. typical local food. By using feedback from sensors, the use of water and pesticides can be greatly reduced, making agriculture gentler on the environment.

This is just an example – there are many possible applications of AI in smart cities, in transport, or in monitoring and understanding the evolution of our planet, which is the objective of the European Commission's "Destination Earth" initiative.

We cannot forget that AI has its own sustainability problem. Think about GPT-3, the recent language model that uses deep learning to produce human-like text, created by OpenAI. I don't know exactly how much energy was used to "train" the model, but I have read some impressive numbers, equivalent to the amount of energy needed for the production of a thousand tons of steel in an electric furnace. Even if this estimation were wrong by one order of magnitude, it would still be clear that this type of AI is not sustainable.

The HiPEAC community is well aware of this problem – research on energy efficiency in computing has been going on for many years, with some very interesting results; in the coming years, the European Commission is planning to fund research on "frugal AI" and will develop a new generation of energy efficient and secure hardware and software. But this, of course, is only a small part of the work needed.

What is most important, at the beginning of the European Green Deal, is to identify the areas where AI can make our economy more sustainable, enabling innovative solutions that transform the products and services we use in something that the planet can afford. This is not only an altruistic objective aiming to leave a better planet to our grandchildren: the reason is also economic. The world has changed, and the competitiveness of European industry will be measured by its capacity to develop products and services which are both innovative and sustainable. There are countries where you can manufacture products without caring about pollution, and places where labour is cheap and workers' rights do not exist. Europe simply cannot compete on these grounds; our only choice is innovation. We now know that AI can be the "secret weapon" for a sustainable and competitive economy, and in the next few years we will see huge development of this area.

So, if you want to contribute to the future of digital technology, find out what is needed to make the products and services of tomorrow more sustainable. Take a look at the HiPEAC Vision 2021 roadmap published this month for inspiration. The possibilities are vast: from simulation / digital twins to frugal neural networks, low-energy devices, autonomous robots, generative design and many others; all these can have very interesting applications in a sustainable economy.

You will not become an eco-warrior for this work: you will be improving the competitiveness of Europe in the only possible way. As a by-product, my nieces will be grateful to you – they will have a lifetime to enjoy a better planet.

A postcard from Budapest

HiPEAC 2021 was due to take place in Budapest, Hungary but has been transformed into a virtual event. Co- General Chair and Master Lecturer at Budapest University of Technology and Economics (BME) Péter Szántó tells us a bit about his home city.

This year's conference is a virtual one as we've sadly had to postpone holding it in Budapest. BME still has an influence on the event; can you tell us a little about the most interesting projects taking place there?

Well, BME is a really large institution, so I would not try to give a comprehensive answer... On a more personal level, I would like to highlight two H2020 projects: BRAINE and SMART4ALL, because both of them are interesting from different perspectives.

BRAINE is about creating an energy-efficient, heterogeneous, AI-centric edge server architecture using multi-core CPUs, GPUs, and FPGAs. As we choose to follow a slightly different path in creating the architecture, it is an exciting engineering task both on the hardware and the software side.

SMART4ALL, on the other hand, is all about supporting innovation within the EU. The university's primary role in this project is to initiate and support technology transfers between SMEs; and create a pathfinder application with SME partners around Europe to accelerate digitization in several application areas, like agriculture, assisted living, etc. I find it important to have a common vision and to deepen the cooperation between EU countries.

What is the local tech scene like in Budapest? Are there any areas of specialization?

Fortunately, I can say that the tech scene in Budapest is in really good shape, and it is growing dynamically. There is a wide range of development areas, ranging from ASIC design and verification to consumer application development. If I need to highlight a few areas, it would be the automotive industry and AI development – most of the Tier 1 automotive suppliers have a development centre in Budapest, and there are quite a few companies involved in different aspects of AI processing.

The conference coincides with the publication of the 2021 HiPEAC Vision technology roadmap. What are the advances in embedded systems or digitization that you hope to see in the next five to ten years?

I expect that efficiency will be an even more important keyword than it is today – and in all areas, including embedded systems. As I see it, heterogeneous, application-specific accelerators play a key role in achieving our goals and, together with the rise of the use of the open-source RISC-V instruction set, it is a great chance for Europe to take a leading position. So, I hope that in five to ten years, Europe will be a strong player in acceleration and system integration both in the hardware and software aspects.

Ten spinoffs from BSC in just five years

In October 2020, Barcelona Supercomputing Center (BSC) announced having launched ten spin-off companies in just five years. In further cause for celebration, these companies have already created 93 high-skilled job posts, raised €2.5 million in external funding and secured contracts worth more than €8 million.

The spin-offs offer advanced services in a range of fields including biomedicine, aerospace and automobile industry security, and quantum computing. The services offered by the companies use results that have come directly from research and technology developed at BSC.

Associate Director Josep M. Martorell points out that the wide-ranging nature of the spin-offs reflects the centre's character: "The diversity of the technologies exploited by these companies (from life sciences to earth sciences, as well as engineering and computing) shows the wide scope of our research. Supercomputing technologies are one of the most important and enabling ones today, without which many disciplines could not exploit all their research potential."

The spin-offs include Energy Aware Solutions, created in collaboration with the Universidad Politècnica de Catalunya, which provides commercial solutions for the task planning and energy management for data centres, and Qbeast, a company that provides a cloud platform for the analysis of big data, based on high-performance technologies developed at BSC.

Read more and see the full list of spin-offs:
bsc.es/ZAs

The TransContinuum Initiative for cross-domain collaboration

HiPEAC is very proud to be one of the eight organisations that have come together to coordinate to best minds in Europe to design computing systems fit for the future.

The TransContinuum is the unison of related digital technologies which offers solutions for the operation of complex data workflow systems. In this continuum, HPC plays a central role as the engine propelling AI, big data, the IoT, cybersecurity, and mathematical components to work together.

The Transcontinuum Initiative, or TCI, is developing a vision of the characteristics of the infrastructure required for the convergence of data and compute capabilities in many leading edge industrial and scientific use scenarios. A paradigm change is needed: we will have to design systems encompassing millions of compute devices distributed over scientific instruments, IoT, supercomputers and cloud systems through LAN, WLAN and 5G networks.

The TransContinuum Initiative will focus on collaboration towards five objectives:

1. Elaborate joint recommendations for R&D to be carried out in EU- or JU-funded work programmes addressing challenges in the digital continuum.
2. Engage with EU Research & Innovation funding entities to promote our recommendations.
3. Generate and foster an interdisciplinary network of experts in science and industry.
4. Contribute to Strategic Research [and Innovation] Agendas or any other road mapping documents issued by participating partners, specifically on interdisciplinary technical aspects, with a view to extend the concept of co-design to cover the entire continuum.
5. Contribute to the five Horizon Europe missions (adaptation to climate change including societal transformation, cancer, healthy oceans, seas coastal and inland waters, climate-neutral and smart cities, soil health and food.)

"We are delighted that the collaborative work initiated with five other associations to prepare the latest edition of our Strategic Research Agenda now lives on and develops into the TransContinuum Initiative. This is a first step towards the necessary synchronization of European efforts across domains," said Jean-Pierre Panziera, chairman of ETP4HPC.

The eight organisations involved are:

- 5G IA, the 5G Infrastructure Association,
- AIOTI, the Alliance of Internet Of Things Innovation,
- BDVA, the Big Data Value Association,
- CLAIRE, the Confederation of Laboratories for Artificial Intelligence Research in Europe,
- ECSO, the European Cybersecurity Organisation,
- ETP4HPC, the European Technology Platform for High-Performance Computing,
- EU-Maths-In, the European Service Network of Mathematics for Industry and Innovation,
- HiPEAC project (High Performance Embedded Architecture and Compilation).

Read more:

etp4hpc.eu/transcontinuum-initiative.html

hipeac.net/vision

TCI TransContinuum Initiative

MonetDB/e: the new open source SQL engine for embedded data analytics

HiPEAC member company MonetDB Solutions announced in October the release of MonetDB/e, a high-performance embedded version of the MonetDB database system for data analytics. It brings the full power of the world-renowned column store MonetDB for business intelligence, data analytics and machine learning into the heart of users' applications, combining the best from both embedded and server-based RDBMSs into one system. It offers data scientists a lightweight and easy-to-use embedded database with fast analytical queries and support for transactions.

MonetDB/e is part of the open-source MonetDB software suite and is distributed under the liberal MPL2.0 licence. It's available on all major operating systems including Linux, Mac and Windows. The current release supports the Python and C language bindings. Support for R and Java is expected in 2021. The Python package monetdbe is available from PyPi and users won't need to install or compile the MonetDB code itself.

https://pypi.org/project/monetdbe/

www.monetdbsolutions.com

Congratulations! Prof. Martin Kersten, CEO of MonetDB Solutions became in November 2020 a Knight of the Order of the Lion of the Netherlands, an honour conferred on him by the King of the Netherlands for exceptional service to the community, to reflect his work as a leader in academia and industry.

monetdb

New in 2020: TECoSA, the Center for Trustworthy Edge Computing Systems and Applications

New applications of artificial intelligence, augmented reality and collaborative smart cyber-physical systems require a new paradigm where localized computing power (so called edge or fog computing systems) enable real-time solutions that contribute to an economically and environmentally sustainable industry.

TECoSA aims to create a dynamic multidisciplinary research environment on trustworthy edge computing systems relevant for multiple industrial domains. The trustworthiness aspects currently in focus include safety, security, predictability/availability and privacy. TECoSA brings together experts in these fields at KTH Royal Institute of Technology with partners from several industrial domains to create a long-term multidisciplinary research and innovation environment with a focus on secure, safe and predictable edge computing systems and applications.

Expected outcomes from TECoSA include scientific results, open source material and people trained in edge computing systems modeling, resource management, machine learning, lightweight security and architecting. TECoSA will systematically foster education and knowledge transfer to contribute to Swedish competitiveness in industrial digitalization in terms of R&D efficiency as well as new cost-efficient products and services.

Launched in March 2020, supported by VINNOVA, TECoSA brings together a strong composition of complementary partners relating both to research and industrial domain competences. To achieve the desired critical mass, the TECoSA strategy has been explicitly designed to accomplish active collaboration among its stakeholders, across disciplines and domains. Three main projects (Safety, Security, Predictability) are used to structure the research with strong ties to industrial use-cases, drawing upon their common challenges. Collaboration within TECoSA is driven through knowledge exchange, testbeds and co-location, and internationally through mobility.
TECoSA hosts an open Seminar Series, with monthly talks from key players in the world of edge computing. Attended by industry members, researchers and PhD students, the panel debates help open up discussion and new lines of thought.

tecosa.center.kth.se

First ever ITEM workshop

In September, the first edition of a new workshop on IoT, Edge, and Mobile for Embedded Machine Learning (ITEM) took place, co-located with ECML-PKDD as the premier European machine learning and data mining conference. Even though the workshop had to take place virtually, there was a lively discussion and interaction, aided by inspiring keynote presentations by Luca Benini from ETH Zürich ("From Near-Sensor to In-Sensor AI") and Song Han from MIT ("MCUNet: TinyNAS and TinyEngine on Microcontrollers"). Those keynotes created the right environment for a number of contributed talks in the areas of hardware, methods and quantization, coming from institutions including Universidade da Coruña, Heidelberg University, Bosch Research, University of Duisburg-Essen, Technical University of Munich, KU Leuven, Università di Bologna, and Graz University of Technology.

Early take-aways include, on the hardware side, observations on open-source digital hardware (PULP) as well as analog hardware (BrainScaleS-2) as promising emerging alternatives to established architectures; that code generation for specialized hardware can be challenging; and that designing processor arrays is more difficult than one might think. From a methodological point of view, 8bit seems to be a natural constant when it comes to quantization, and time-multiplexing and on-device learning can be viable options. On the compression side, observations include that predictive confidence can help dynamic approaches to switching among models, heterogenous uniform quantization as well as application-specific (radar) quantization.

ITEM2020 was organized by Holger Fröning (Heidelberg University), Franz Pernkopf (Graz University of Technology), Gregor Schiele (University of Duisburg-Essen), and Michaela Blott (XILINX Research, Dublin), and assisted by Benjamin Klenk as progam chair (NVIDIA Research, Santa Clara). It is planned for ITEM to continue in 2021, so if you are interested stay tuned for updates or ask the organizers to add you to the mailing lists.

https://www.item-workshop.org

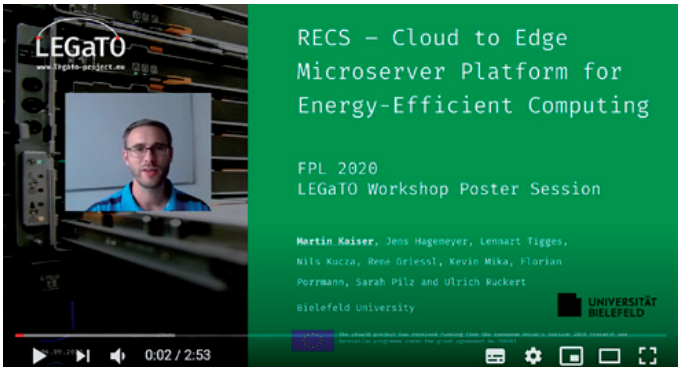
LEGaTO project draws to a close with low-energy heterogeneous computing workshop

The LEGaTO project celebrated its final event, the 'Low-Energy Heterogeneous Computing Workshop', on 4 September 2020 as part of FPL 2020. The aim of the workshop was to present the optimization low-energy techniques used in LEGaTO to other stakeholders and share project's latest results. It included a keynote talk by Stephan Diestelhorst (Xilinx), five panel sessions on the main topics of the project and a poster session.

Martin Kaiser (Bielefeld University), who presented the 'most voted-for poster', talked to us about his research and the experience of working on a European project like LEGaTO as an early career researcher.

Many congratulations on the success of your LEGaTO poster! Could you tell us a bit about the research you presented?
Thank you very much! First of all, it is not only "my" research, but part of the work of about ten enthusiastic and talented engineers here at Bielefeld University and the company Christmann. Together we are developing the hardware platform RECS (short for Resource Efficient Clustering System) that is being used for all applications within LEGaTO.

Martin Kaiser (Bielefeld University) demonstrating LEGaTO's hardware platform t.RECS, Universität Bielefeld/S. Jonek



Martin Kaiser giving his pre-recorded poster presentation, in true 2020 style

The unique feature of RECS is that application developers can quickly change among a variety of tightly clustered heterogeneous target architectures. From a research point of view, the RECS platform can be seen as a "sandbox", which allows you to test different hardware components with low effort and select the best match for your applications.

What were your responsibilities in LEGaTO? What was your experience working on the project?
Within LEGaTO, I worked on t.RECS, which is short for tiny RECS. t.RECS expands the RECS portfolio previously focused on cloud and data centre to include edge computing applications. The server comes in a compact form factor and aims for a high energy efficiency and connectivity and was designed to be used in edge applications like automation, automotive or smart home/city. As a part of the hardware development, we worked as an executive member of the PICMIG committee to define the upcoming COM-HPC standard, which is used for t.RECS. We succeeded in driving the standard towards LEGaTO's goals to support heterogeneous and energy-efficient computations.

What was the added value to your career of being in a pan-European project?
It was great to work in a group of such talented international scientists, who all have the common goal of improving the energy efficiency of computations in mind. Also, I gained lots of international experience by playing an active role in the standardization process of COM-HPC. But in the end, the best thing is to get the confirmation that the developments we worked on for years are being used in practice and provide added value in the daily development process.

Recording of the event: bit.ly/legato-workshop

Winners of the HiPEAC Tech Transfer Awards 2020

2020 marks the sixth edition of the HiPEAC Tech Transfer Awards. This year's award winners include a deep neural network used in a space mission and a scalable logic locking framework for hardware integrity protection. The range of their winning work serves as a demonstration of how HiPEAC research continues to resonate beyond the lab.

HiPEAC's Tech Transfer Awards recognize successful examples of technology transfer, which covers technology licensing, providing dedicated services or creating a new company, for example. In addition to a certificate, first-time winners are awarded the sum of €1,000 for the team that developed the technology.

This year, eight winners have been selected:

Petar Radojković, BSC: Performance, power and energy impact of Micron's novel HPC memory systems: hardware simulation and performance modelling

Barcelona Supercomputing Center has provided Micron Technology Inc. dedicated services to address the challenge of quantifying the impact of novel memory devices on the overall performance, power and energy consumption of memory systems for HPC.

Gianluca Palermo, PoliMi: GeoDock - a fast and configurable pocket-aware ligand pose generator for high-throughput molecular docking

GeoDock is a configurable and auto-tunable software library that was transferred from Politecnico di Milano to Dompé Farmaceutici for use in its proprietary drug-discovery platform.

Gianluca Giuffrida, University of Pisa: CloudScout segmentation neural network

The University of Pisa transferred to Cosine Measurement, a company specialized in the development of measurement systems, knowledge and expertise in the design and development of custom low energy hardware accelerator IP cores. The CloudScout DNN is in use in the European Space Agency's current PhiSat-1 space mission.

Matthias Jung, Fraunhofer: DRAMSys4.0 - a flexible DRAM subsystem design space exploration framework

DRAMSys4.0 is a flexible and fast DRAM subsystem design space exploration framework. It is being published as open source software under the BSD3 licence. Before the latest features of DRAMSys are released as open source, Fraunhofer IESE offers commercial licences to partner companies for early application of the simulation models and consulting.

Ricardo Carmona-Galan, CSIC: PHOTONVIS - a pilot for a scalable solid-state LiDAR

PHOTONVIS is a new technology-based company created to develop and commercialize CMOS-SPAD sensors. It uses results from the research group on Integrated Interface Circuits and Sensory Systems (I2CASS) of the Institute of Microelectronics of Seville, a group in a centre that spans the CSIC (Spanish National Research Council) and the University of Seville.

Silvia Panicacci, University of Pisa: SatNav E@syCare

IngeniArs S.r.l. is a spin-off of the University of Pisa. Its main product in the healthcare sector is E@syCare, an advanced system comprising a cloud-based web application and mobile applications for remote monitoring of vital parameters and other patient wellbeing metrics by medical staff, according to personalized healthcare plans.

Yannis Papaefstathiou, EXAPSYS: Very fast and accurate simulator of HPC and CPS systems

The main parts of the COSSIM simulator (the synchronization architecture and implementation and certain models) have been exclusively licensed from the Telecommunication Systems Institute at the Technical University of Crete, to Exascale Performance Systems - EXAPSYS Plc a start-up created mainly for the commercial exploitation of research results.

Dominik Šišejković, RWTH Aachen University: A scalable logic locking framework for hardware integrity protection and its application to a RISC-V processor

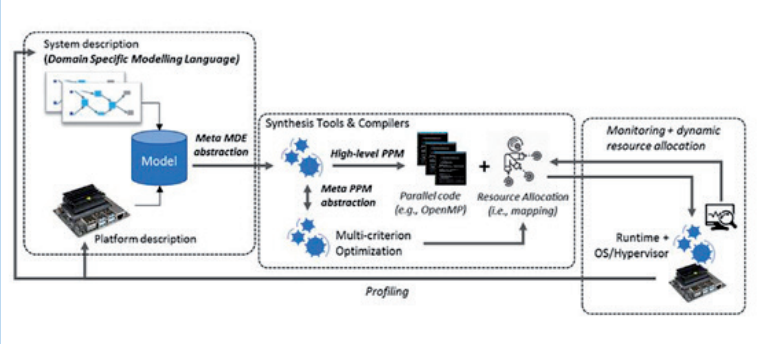
Two main contributions have been transferred to industrial partner HENSOLDT Cyber GmbH: an industry-ready logic locking software framework for the application of scalable locking schemes to protect hardware designs in the IC design and fabrication flow and a fully logic-locked 64-bit linux-ready RISC-V processor core.

READ MORE ON THE WINNERS:
hipeac.net/news

AMPERE project at the SoS integration with CPS workshop

The AMPERE project aims to develop a new generation of programming environments for low energy and highly-parallel computing, capable of implementing correct-by-construction cyber-physical systems (CPS). As a result, it has considered 3 different layers:

- model-driven engineering (MDE)
- code synthesis tools and compilers
- the runtime system, the OS and hypervisor.



AMPERE software components

Overall, AMPERE will provide a baseline for the future of CPS by bridging the gap between MDE and parallel programming models supported by the most advanced parallel embedded processor architectures. It will tackle long-term complex market challenges that need significant improvements to manage the needs of the competitiveness of European industries.

On that account, Eduardo Quiñones (BSC) - AMPERE's coordinator - presented the project on 10 September 2020 at a HiPEAC workshop on "SoS integration with CPS", SoS standing for System of Systems. The workshop provided an opportunity for coordination between experts in CPS and facilitated dialogue on CPS technology transfer and challenges. The group considered a model for CPS that is scalable and which will be easier to show where and how advancements are taking place. The model had been discussed and refined at several CPS workshops and provided as a report to the European Commission. The workshop resulted in a public report to the Commission, including advice on the engineering practices and tools needed to support the application of SoS to CPS.

AMPERE's future actions will be to keep its commitment to developing innovative software architecture for the development and execution of advanced CPS, taking into account the non-functional requirements inherited from cyber-physical interactions. This novel technology will be employed in the automotive and railway domains.

ampere-euproject.eu/
bsc.es/quinones-moreno-eduardo

AMPERE
A Model-driven development framework for highly Parallel and EneRgy-Efficient computation supporting multi-criteria optimisation

ELASTIC project attracts attention at EBDVF 2020!

The ELASTIC project contributed to the digital version of the European Big Data Value Forum (EBDVF 2020) this year with a parallel session dedicated to the novel ELASTIC software architecture and smart mobility use case. The session, entitled "Next generation smart mobility systems, leveraging extreme-scale analytics over a novel elastic software architecture", featured four expert talks and received questions and plenty of interest from the audience.

Eduardo Quiñones, ELASTIC project coordinator from the Barcelona Supercomputing Center (BSC) and one of the speakers, said that "it was a pleasure to be a part of EBDVF 2020, the flagship event of the European big data and AI Research and Innovation community, with a contribution on our innovative infrastructure developed for advanced mobility systems and autonomous transport networks."

Our speakers, Eduardo Quiñones, Jürgen Assfalg (ICT Manager at Città Metropolitana di Firenze), Anna Queralt (Senior Computer Scientist at BSC), and Elli Kartsakli (Senior Researcher at BSC), showcased the ELASTIC software ecosystem, an architecture capable of exploiting the distributed computing capabilities of the compute continuum of the smart city, while guaranteeing additional properties, such as real-time, energy, communication quality, and security.

Particular focus was given to the real-life mobility use case tested in the public tram network of Florence, Italy, aiming to improve safety, efficiency and maintenance of the transportation vehicles, and, therefore, the overall quality of life of citizens.

The full recording and presentation slides are available on the ELASTIC website.

READ MORE:

ELASTIC website: elastic-project.eu
ELASTIC at EBDVF 2020: bit.ly/ELASTIC-EBDVF
ELASTIC software ecosystem: bit.ly/ELASTIC-ecosystem
ELASTIC mobility use case: bit.ly/ELASTIC-usecase

ELASTIC
A Software Architecture for Extreme-ScaLe Big-Data AnalyticS in Fog CompuTIng Ecosystems

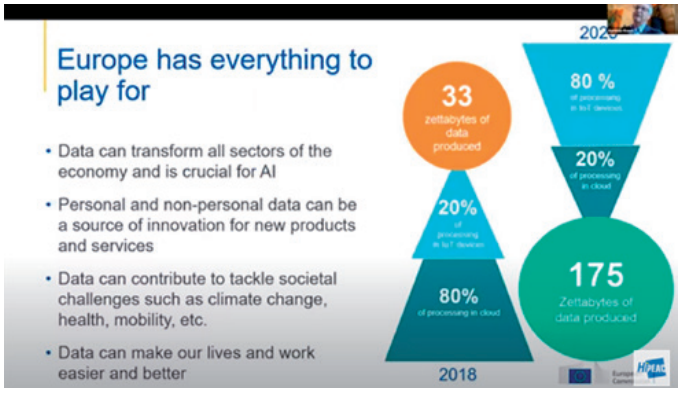
Photo credit: Arto Liskola

A virtual first for Computing Systems Week

Why Computing Systems Week in October took place as a fully online event needs no explanation. With a focus on cyber-physical systems (CPS) and innovation, the programme, organized by Tampere University, was a showcase for wide-ranging technologies and ideas, including an industry session focused on the burgeoning technology community in the Tampere region.

Keynote talks from Jürgen Niehaus (SafeTRANS), Markku Markkula (European Committee of the Regions), Maurits Butter (TNO) and Heikki Ailisto (VTT Technical Research Centre of Finland) covered the ECSEL JU funding scheme, innovation and its relationship with cities, business models for pan-European collaboration and computing on the edge and in the cloud, respectively.

The innovation landscape was explored further in sessions about the Smart Anything Everywhere initiative and the Digital Innovation Hubs (DIH). Technical sessions covered industrial IoT solutions for heterogeneous ecosystems, design of processor architecture and how the CPS technology-pull landscape is likely to evolve for sustainability. With presentations from Tampere University, the Robocoast DIH and local tech companies, the Inspiring Futures session helped current students to see the range of exciting career possibilities that await them.



The next HiPEAC Computing Systems activities are due to take place in Spring 2021. Follow us on social media for more.

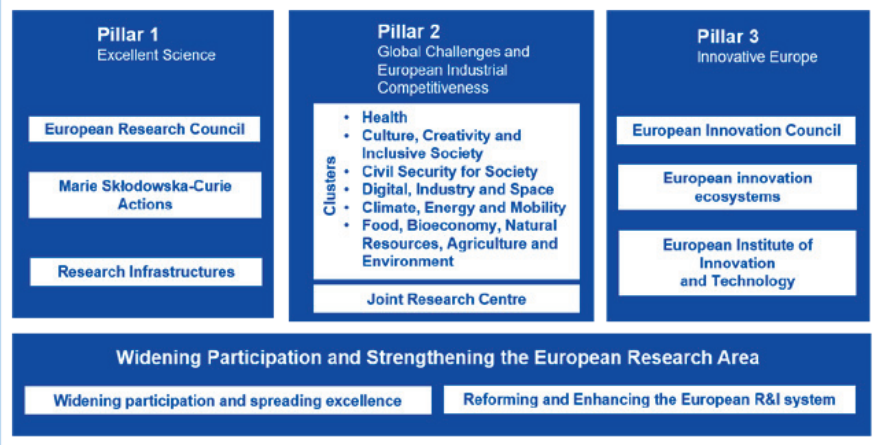
Videos of all sessions of Computing Systems Week Autumn 2020 are on the HiPEAC YouTube channel now: search #CSWAutumn20 You can also access videos of courses of the ACACES Summer School 2020 right in the same place, by searching for #ACACES20

Brighter days on the horizon

Set to be the most ambitious research programme in the world, the EU's Horizon Europe programme will support cutting-edge research and innovation to the tune of around €95.5 billion.

The EU institutions reached a political agreement on Horizon Europe on 11 December 2020. Although subject to formal approval by the European Parliament and the Council of the EU at the time of writing, the agreement paves the way for the largest transnational programme ever supporting research and innovation.

The budget of around €95.5 billion for 2021-2027 (current prices) represents a 30% increase vis-à-vis the current research and innovation programme, Horizon 2020 (comparing Horizon Europe against Horizon 2020 for EU27, in constant prices) and makes Horizon Europe the most ambitious research and innovation programme in the world.



Horizon Europe will promote excellence and provide valuable support to top researchers and innovators to drive the systemic changes needed to ensure a green, healthy and resilient Europe. The programme will support a range of schemes and models, including collaborative research relating to societal challenges. It reinforces technological and industrial capacities through thematic clusters that address the full spectrum of global challenges. For example, the Climate Energy and Mobility cluster and the Digital Industry and Space cluster will scale up R&I resources in climate-related domains and ensure that European enterprises have access to the technologies and data they need. In the latter, Quantum Research will be prioritized thereby expanding the European scientific leadership and excellence in quantum technologies.

The Health cluster will tackle challenges such as the coronavirus pandemic, the extension of clinical trials, innovative protective measures, virology, vaccines, treatments and diagnostics, and the translation of research findings into public health policy measures.

LEARN MORE:
@EUScienceInnov

HiPEAC EU projects: call to action!

As Horizon 2020 draws to a close and we look ahead to the new and exciting world of Horizon Europe, don't forget to take a look at some of the fantastic EU projects that the HiPEAC community has undertaken. The HiPEAC website currently lists 338 projects (yes, you did read that right) covering everything from excellence in simulation of weather and climate, via exascale programming models for extreme data processing, to scalable and power efficient HPC platforms.

In every issue, the HiPEACinfo features a selection of EU projects in the Innovation Europe section. We're also very keen to support your projects on social media, so don't forget to follow HiPEAC on Twitter and LinkedIn and tag us in your posts.

While online events and webinars don't offer quite the same networking opportunities as in-person events, they are still an excellent way to find out about who is doing what in the world of high performance and embedded architecture and compilation. Stay tuned for what's to come from HiPEAC in 2021. You never know, your next project partner might be just around the corner!

hipeac.net/projects
@hipeac
hipeac.net/linkedin

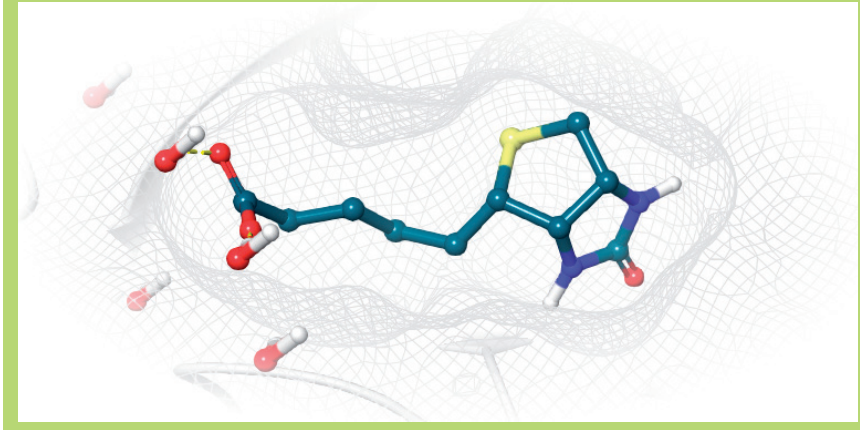
PELE Platform

Nostrum Biodiscovery, a spin-off from Barcelona Supercomputing Center which uses simulation based on computation and AI methods to accelerate the discovery of new drugs, has recently released a new version of PELEplat, a Python wrapper for PELE, its proprietary software for biomolecular modelling.

PELE is a molecular simulation program combining a unique Monte Carlo technique with structure prediction algorithms. It is based on a multiple-step approach consisting of localised perturbation followed by side chain sampling and minimisation cycles. Over time, many packages aiming to solve specific drug discovery problems were built around PELE. And that's how PELEplat was born! It currently consists of more than 10 packages for small molecule drug discovery, including allosteric pocket search, docking pose refinement and elucidation of entrance pathway in G-protein-coupled receptors. However, it is constantly being expanded and improved to address the challenges of enzyme engineering as well as make a number of machine and deep learning algorithms available to the users.

In the meantime, the PELE team introduced two major features in the latest version: AquaPELE and HT-FragPELE. AquaPELE includes an additional Monte Carlo step focused on explicit water molecules, allowing to explore water-mediated interactions as well as solvent displacement upon ligand binding. The feature is completely automated, so there is no need to tweak any parameters. PELEplat will even add water molecules to your system, if there aren't any! Additionally, significant work was done to enhance the performance of FragPELE giving birth to HT-FragPELE. From now on, the users will be able to grow a library of thousands of fragments onto a scaffold while exploring the conformational space of the protein in a matter of days.

Finally, the NBD team constantly works on improving user experience by adding nifty features like automatic detection of geometry around the metals and extended support for non-standard residues, making the modelling experience more straightforward and user-friendly.



WANT TO LEARN MORE ABOUT PELEPLAT? CHECK OUT THE WEBSITE!
Software: https://nostrumbiodiscovery.github.io/pele_platform/
NBD articles: https://nostrumbiodiscovery.github.io/papers/

EPEEC project work on parallelware technology awarded EU Innovation Radar Prize 2020

EPEEC consortium member Appentra was awarded the Innovation Radar Prize 2020 in October 2020 in the Innovative Science category for their work on parallelware technology carried out as part of the EPEEC project. The Innovation Radar, a European Commission initiative, identifies high-potential innovations and innovators in EU-funded research and innovation projects, and the Innovative Science prize category aims to recognize excellent scientific work that shows market promise.

The innovation in question, 'Parallelware Technology: analysis tool aiming to reduce the burden of making code parallel', was outlined by Manuel Arenaz of Appentra on 24 September 2020 during a special pitching session at the European Commission's Research and Innovation Days.

EPEEC's main goal is to develop and deploy a production-ready parallel programming environment that turns upcoming overwhelmingly-heterogeneous exascale supercomputers into manageable platforms for domain application developers. It will significantly advance and integrate existing state-of-the-art components based on European technology (programming models, runtime systems, and tools) with key features enabling three overarching objectives: high coding productivity, high performance, and energy awareness. EPEEC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 801051.

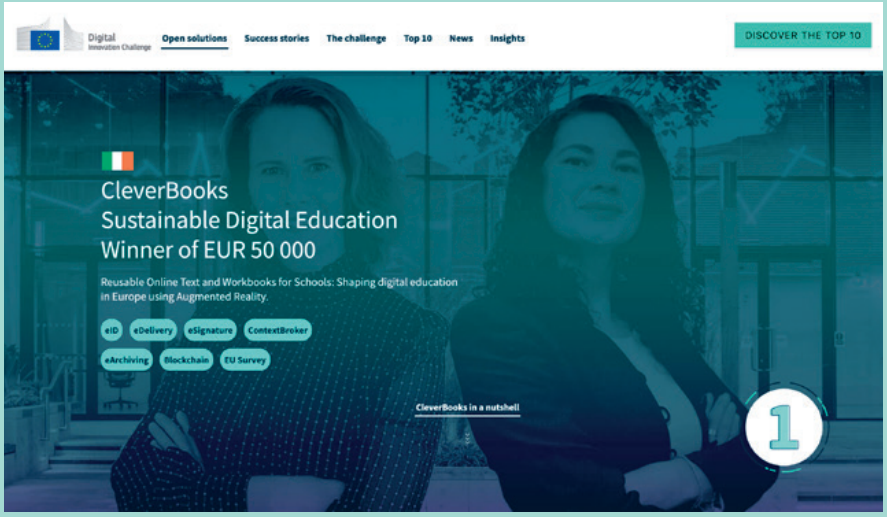
READ MORE:
epeec-project.eu/media/news

HiPEAC member SME CleverBooks wins EU Digital Innovation Challenge

Ireland-based EdTech company CleverBooks was named in November as one of just three winners of the European Commission's 'Digital Innovation Challenge'. The Challenge invites digital SMEs and startups from across Europe to innovate and grow their business using open and reusable solutions.

It was the company's Augmented Classroom product, a digital platform that provides mobile and web-based augmented reality (AR) solutions, that gained the judges' approval. Augmented Classroom allows students to experience more engaging and interactive lessons in the classroom as well as remote learning, a concept which has come to the fore in our now socially-distanced world. Through AR, students can hear, feel and touch knowledge, as well as gain vital digital skills.

Using Augmented Classroom, students can explore, co-create and collaborate globally with paired schools around the world in the AR environment. It's indeed the first Software as a Service (SaaS) platform that provides multi-user interactivity in real time for collaborative and co-creative projects, regardless of users' geographical location.



"Augmented Classroom is a truly flexible solution that uses the latest AR technology to support collaborative, cloud-based activities, allowing students to experience more of the world without leaving their classroom or home," says Dr Darya Yegorina, CleverBooks CEO. "We strongly believe that it's not about how educated you are, but how you are educated."

The European Commission sees the Challenge as an opportunity for small and new businesses to grow and help shape the 'Digital Europe', whilst inspiring public administrations across the EU to continue to adopt open and reusable solutions to create generic digital public administration components. These 'building blocks' can be used to facilitate the delivery of digital public services by public organizations.

FURTHER INFORMATION:
EU Digital Innovation Challenge bit.ly/EU-DIC
TEDx talk: Darya Yegorina bit.ly/DaryaTEDx
CleverBooks cleverbooks.eu

Dates for your diary

DAC 2021: Design Automation Conference
11-15 July 2021, San Francisco, CA, United States
HiPEAC Paper Award conference
dac.com

ISC High Performance 2021
June/July 2020, virtual
isc-hpc.com

DATE21: Design, Automation and Test in Europe
1-5 February 2021, virtual
HiPEAC booth
date-conference.com

ASPLOS 2021: Conference on Architectural Support for Programming Languages and Operating Systems
19-23 April 2021, virtual
HiPEAC Paper Award conference
asplos-conference.org

EuroHPC Summit Week 2021
22-26 March 2021, virtual
events.prace-ri.eu/event/1018/

PLDI 2021: ACM SIGPLAN Conference on Programming Language Design and Implementation
21-25 June 2021, Quebec City, Canada
HiPEAC Paper Award conference
dl.acm.org/conference/pldi

HPCA 2021: IEEE International Symposium on High-Performance Computer Architecture
27 February-3 March 2021, virtual
HiPEAC Paper Award conference
hpca-conf.org/2021/

What is certain in this uncertain world is that the volume of data collected, stored and processed will continue to grow, and at a rapid pace. HiPEAC 2021 keynote speaker Evangelos Eleftheriou of the IBM Zurich Research Laboratory explains how in-memory computing might open the way to improved performance that doesn't cost the earth.

'It's not that often that you get involved in developing and nurturing a new computing paradigm'

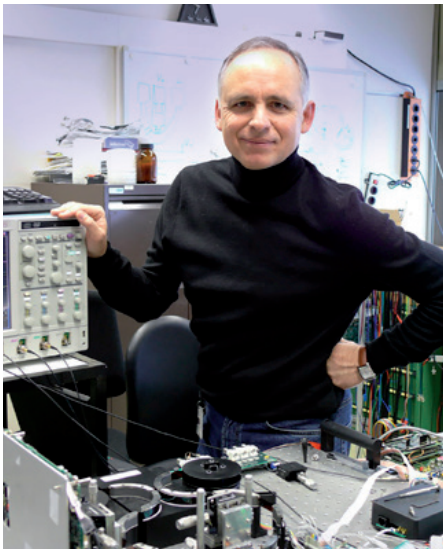


Image: IBM Research

"In-memory computing is at the crossroads of many diverse scientific disciplines ... At the IBM Lab in Zurich, we host dozens of masters and PhD students from universities around the world"

Why is this an exciting time to be working in in-memory computing?

Let me share a startling statistic. By 2023 hyperscale cloud data centre energy consumption is expected to nearly triple due to the enormous growth of data, which is expected to hit 175 zettabytes (ZB) by 2025. As the world looks to reduce carbon emissions there is an even greater need for higher performance at lower power. Compounding this challenge is the end of Moore's law. Thus, there is renewed interest to go in a sense "backwards" and explore analog computing with some inspiration from the most efficient computer of all: the human brain.

More specifically, traditional digital computing systems, based on the von Neumann architecture, involve separate processing and memory units. Therefore, performing computations results in a significant amount of data being moved back and forth between the physically separated memory and processing units, which costs time and energy, and constitutes an inherent performance bottleneck. The problem is aggravated by the recent explosive growth in highly data-centric applications related to artificial intelligence. This calls for a radical departure from the traditional systems and one such approach is analog in-memory computing. This sees certain computational tasks performed in an analog manner in place, i.e., in the memory itself, by exploiting the physical attributes of the memory devices.

Looking ahead to the next ten years, what are the key hurdles that need to be overcome in this area of investigation, and what is already in place to help tackle them?

There have recently been numerous hardware prototype systems from industry and academia that have successfully demonstrated in-memory computing, in particular for AI applications, using either charge-based or resistance-based memory devices. However, there are several challenges that need to be addressed so that this new technology becomes more robust and can be applied to an even wider range of application domains. For example, linear algebra computational kernels, such as matrix-vector multiplication, are common not only in machine learning and deep learning but also in scientific computing applications. However, both memristive and charge-based memory devices exhibit intra-device variability and randomness that is intrinsic to how they operate. They are also prone to noise, nonlinear behaviour and inter-device variability and inhomogeneity across an

array. Thus, the precision of analogue matrix-vector operations is not very high. Although approximate solutions are sufficient for many computational tasks in the domain of AI, building an in-memory computing system that can effectively address scientific computing and data analytics problems – which typically require high numerical accuracy – remains challenging.

Besides the challenges at the device level, there are several that need to be tackled at the peripheral circuit level. For example, a critical issue is the design of energy and area efficient high-precision digital-to-analog (analog-to-digital) conversion units. Finally, other crucial aspects are the hierarchical organization of in-memory computing cores in a way to efficiently tackle a range of applications as well as the design of a software stack that extends from the user-level application to the low-level driver that directly controls the computational memory unit.

What keeps you excited about this field of research and what words do you have for current university students who might be interested in getting into this field?

I've been in research for a long time and it's not that often that you get involved in developing and nurturing a new computing paradigm. The research field of in-memory computing has been gathering momentum over the past several years and memory manufacturers and system integrators are increasingly showing interest in commercialization – this is what keeps me excited.

In-memory computing is at the crossroads of many diverse scientific disciplines ranging from material science and memory device technology to circuit and chip design and from computer architecture and compiler technology to machine learning and deep learning algorithms and applications. As such it offers students a variety of topics for graduate studies and great prospects for future career opportunities in an emerging field. At the IBM Lab in Zurich, we host dozens of masters and PhD students from universities around the world. For any length of time, from six months to two years or more, they work at IBM with our expert scientists, publishing papers and doing research across multiple layers of the stack with an eye on commercialization. So my suggestion would be to apply for such positions at either IBM or other industry labs.

Which other technologies that you would like to see become a reality in the next decade?

Well some may argue that it's already a reality – quantum computing. Since 2016, when IBM put its first quantum computer online for free there has been explosive growth in this field and it's not just academic interest. Fortune 500 firms like JPMC, ExxonMobile, Daimler, Boeing, Samsung and GoldmanSachs are all looking at how quantum computing can address challenges previously considered intractable, for example, portfolio optimization and accelerating materials discovery. And within the next decade, as the hardware improves and as skills are developed, we are going to see tremendous improvements felt by society.

Image: IBM Research

HiPEAC 2021 keynote Brad McCredie, Corporate Vice President of AMD, talked to us about the path he envisages for high performance and exascale computing.

'Our biggest challenges in the world today are limited by what is computationally possible'

"AMD is today working with over twenty global institutions to deliver donations of compute resources to fight COVID-19"

Why is this an exciting time to be working in exascale computing?

Exascale is a milestone the HPC industry has been anticipating for a long time. As the need for compute power continues to grow, exascale is the next logical step in the evolution of supercomputing.

Our biggest challenges in the world today are limited by what is computationally possible. Drug discovery for novel viruses is one that is top of mind for all of us, for example. AMD is today working with over twenty global institutions to deliver donations of compute resources to fight COVID-19. Those researchers are looking into everything from attacking the spike protein of the virus (HLRS), to developing therapeutics (UCLA).

Predicting the impact of climate change, reducing our dependence on fossil fuels, finding ways to maximize crop yields to feed people, the list goes on and on.

As a chip designer, delivering the processors that will supply that next leap in processing performance is what gets me excited about going to work every day.

What are the key markers or indicators of progress in HPC?

For me, I look very closely at performance per watt and performance per dollar. As we continue to push the material science of semiconductors, the cost of putting more transistors to work without pushing power consumption to unreasonable levels is one of the greatest challenges we face. If we want to get HPC systems into the hands of more and more people so we can solve the big problems, those systems have to be as affordable as possible and use as little power as possible.

There are a lot of tricks that processor designers can employ to make computing more efficient. We are actively monitoring and managing power consumption across our processors now, shutting down complete parts of the chip, even for only a millisecond, to save as much power as possible.

Same with performance per dollar – by building a system in package, with multiple smaller chips connected together inside, rather than manufacturing one giant, monolithic processor, you are reducing the odds that you will have to throw away that very expensive piece of silicon if something fails on that single, large processor.

Those are just a couple of the leading indicators of progress for AMD.

Read the full interview with Brad:
hipeac.net/news/#/blog/
More about AMD:
amd.com

In her HiPEAC 2021 keynote speech, Tulika Mitra, Professor of Computer Science at the National University of Singapore, will tell us about software-defined accelerators for edge computing. We caught up with her about why this is an exciting time to be carrying out research in this area.

Pushing the frontiers of edge computing

Embedded computing systems are everywhere. They are the heart and soul of the internet of things (IoT) and cyber-physical systems (CPS) and these systems are transforming our everyday lives. For example, Singapore's Smart Nation initiative envisions the seamless use of IoT devices and data analytics as a measure towards enabling people to live sustainably and comfortably.

Until now, the processing of the data collected from the devices and sensors at the edge of the internet has mostly taken place in the cloud. But moving forward, that is no longer the best option for a variety of reasons. The transfer of rich image and video data from the edge devices to the cloud consumes tremendous network bandwidth and energy. As a result, a significant fraction of the data simply never gets used. Thus, it is essential to perform data analytics at the edge as much as possible. More importantly, edge computing also offers better security, privacy, and real-time latency for certain critical applications.

However, there are several hurdles to be overcome: edge computing is restricted at the moment by the limited compute power and battery life of the edge devices. I am interested in pushing the frontiers of edge computing to bring a lot of what happens in the cloud today to the smartphone or the smart watch, for example. If we can achieve that, the cloud could then be reserved for very large-scale and long-term analytics.

There is an evident consensus on the need for edge computing, but how do we do it? Fortunately, the technology trends are pointing towards bold innovations at the systems level to pave the way for edge computing. The end of Moore's law and Dennard scaling brings with it exciting opportunities, as general-purpose processors are no longer sufficient to meet the energy efficiency requirement of edge computing. This has led to a plethora of work in domain-specific accelerators, such as those for AI applications. These accelerators are mostly restricted to deep learning, which will evidently play an important role in analytics for a long time. Building a complete application, however, requires many other compute-intensive kernels to move to the edge. I am particularly interested in designing highly efficient universal accelerators that can support different computational kernels at different points through software-defined re-configurability.

We have recently embarked on a five-year research programme called PACE to create an innovative, reusable, versatile, ultra-low power, software programmable accelerator that achieves 50x improvement in energy efficiency for the edge devices. PACE would enable multi-sensory on-device analytics; it could, for example, replace current backpack-scale edge analytics solutions carried by individuals for rescue operations with tiny wearable edge devices. Our recent HyCUBE accelerator chip is the initial step in this direction.

Tulika's research interests

- Real-time embedded systems
- Design automation
- Low power design
- Power management
- Heterogeneous computing

Read the full interview with Tulika:
hipeac.net/news/#/blog/
More about Tulika's research:
comp.nus.edu.sg/cs/bio/tulika/

"There is an evident consensus on the need for edge computing, but how do we do it?"

The arrival of 2021 brings with it a new HiPEAC Vision roadmap. It describes advances and evolutions in the computing systems domain and outlines recommendations to guide future developments and funding or investment. In January 2021, we publish the eighth edition of the Vision and take a new approach so as to ensure that our vision keeps apace with the swiftly evolving world of computing systems.

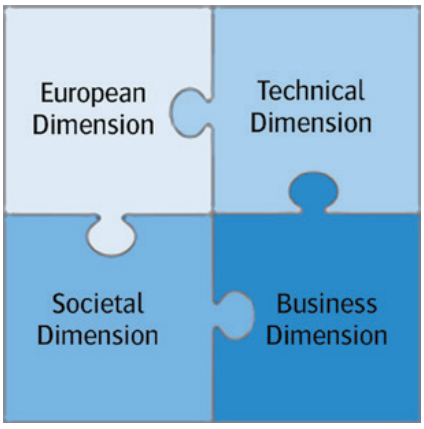
HiPEAC Vision 2021

In the thirty years since the advent of the internet, computing has changed how we live our lives. The HiPEAC Vision 2021 roadmap lays out that tomorrow's computing systems will build a continuum of computing that will stretch from tiny, interconnected edge devices to bigger and more complex systems distributed across the globe.

Our world is evolving very rapidly, as a result of both intentional scientific and societal developments, and unexpected events. The COVID-19 pandemic that began in early 2020 brought with it challenges as never seen before, yet has served to accelerate digital transformation beyond what could have been imagined when we published the last Vision in 2019. Computing and technology played a vital role in the ways in which we changed how we work, learn, shop and communicate with loved ones. We may never return fully to 'the old life'.

Even before COVID-19, the digital world was evolving rapidly and this has had an impact on the HiPEAC Vision: updating it every two years no longer seems in keeping with the speed of the evolution of computing systems. Therefore, we decided to stop producing a long roadmap every other year and create an agile and flexible magazine-like set of articles. The HiPEAC Vision 2021 therefore has two main parts:

- A set of recommendations for the HiPEAC community that will be updated periodically;
- A set of independent articles on various topics that will be continuously updated, improved and added to. This will guarantee that the HiPEAC Vision keeps evolving and remains up to date. Articles are grouped into four themes or dimensions: technical, business, societal and European. The editorial board has sought – and will continue to seek – various authors, experts in their field, to contribute to the articles. This allows for a greater diversity of point of view, and therefore better analysis of the computing systems landscape and robust quality of the recommendations.



So what are the key ideas of the Vision 2021?

ICT is expanding from cyberspace to interact with us directly, for example in self-driving cars, healthcare monitoring,

automated factories or even cities. We are now in the era of cyber-physical systems (CPS), and these systems will be increasingly enhanced with artificial intelligence, so that we could call them ***cognitive cyber-physical systems***, or C²PS. This evolution brings with it increased need for trust, autonomy, safety and efficiency. Complex C²PS need to be more pro-active, anticipate the actions of their environment, and therefore become Cognitive Cyber and Predictive Physical Systems of Systems or (CPS)². The 'systems of systems' part is driven by the fact that systems will need to work more and more with each other; software, applications and infrastructures will increasingly be aggregates of heterogeneous artefacts, including legacy ones, with a variety of deployment requirements.

Applications will be distributed, becoming a continuum of computing across platforms and devices, from the "deep edge" to the cloud, HPC and data centres. Programming has to be reinvented for this, with languages and tools to orchestrate collaborative distributed and decentralized components, as well as components augmented with interface contracts covering both functional and non-functional properties.

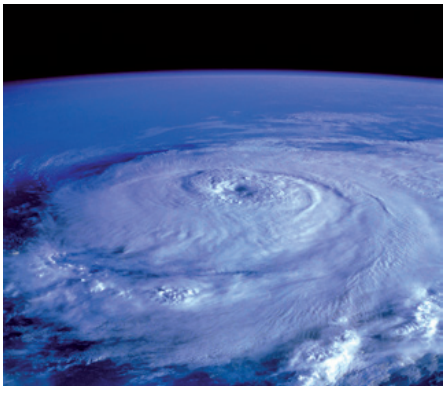
The continuum of computing, an idea pushed by several previous HiPEAC Vision documents, is now more widely recognized; the new "TransContinuum Initiative" involves a number of European organizations (ETP4HPC, ECSO, BDVA, 5G IA, EU-Maths-In, CLAIRE, AIoTI and HiPEAC) and will promote the ideas further through concrete actions.

The Vision 2021 explores changes and progress in the development of hardware and software, as well as the cultures and movements that contribute to their evolution. Key underlying themes include sustainability and the urgency to lower the environmental footprint of ICT, and the need for computing systems and devices that are secure by design and allow users to be in control of their privacy and assets. Indeed, the Vision covers a wide range of themes: interoperability, the widespread "as a service" trend, code migration, containerization, use of legacy systems, edge to cloud computing (the "continuum"), distribution of computing, secure and trustable platforms, and many more. These themes drive us to a more global vision of what the ICT infrastructure of tomorrow could be and we propose to bring this vision to life through a multidisciplinary "moonshot" **program** that we call **"Guardian Angels"**, orchestrating services and protecting people and businesses from the complexity and dangers of cyberspace. Because they understand humans' natural

ways of communicating, these guardian angels will allow everyone to access the full range of services of this new era.

HiPEAC proposes to help Europe to consolidate and coordinate its talents to create this paradigm, and unify and synergize technical solutions that already exist, for the benefit of humanity.

Read the articles published in January 2021:
hipeac.net/vision
Read more about the TransContinuum Initiative on page 7.



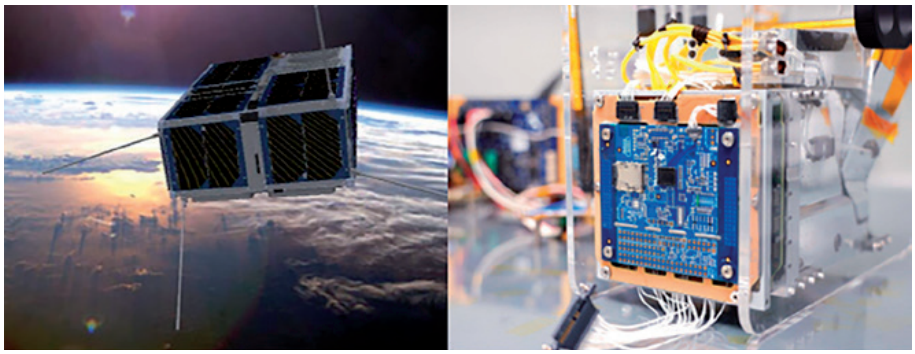
The final frontier, space exploration still captures the imagination of the young and old alike. Yet there are very practical solutions that satellite technologies can offer to mitigate the effects of the grand challenges that face humankind today, such as forest fires caused by changing weather patterns. We caught up with David Moloney, formerly of Intel and now based at Ubotica, to find out more.

The sky's the limit for Ubotica's edge AI platform

One of the most remote outposts of the EU, French Guiana in central America, is separated by 15km of shark-infested waters from the former French penal colony of Devil's Island which featured in the book and film "Papillon" and whose most famous inmate, Captain Alfred Dreyfus, changed the political landscape of France forever.

The former colony made history again on 2 September 2020, this time because the world's first edge AI-enabled satellite, PhiSat-1, took off for Low Earth Orbit (LEO) from the European Space Agency's (ESA) spaceport Kourou and is set to change the landscape of space exploration forever.

The involvement of Intel and a call out on social media by Bob Swan, CEO of Intel, has generated massive interest in the project with over 2,500 page views and 83k twitter impressions.



PhiSat-1 (L), and PhiSat-1's AI inference engine (R). Credit (R): Tim Herman/Intel Corporation

The tiny satellite is around the size of a cereal box and is packed with a hyperspectral sensor from Dutch company cosine and an AI engine from Irish SME Ubotica based on the Intel Movidius Myriad 2 SoC.

The original mission of PhiSat was to monitor polar ice and soil moisture but, as we'll see, the sky's the limit for AI-enabled applications with Ubotica's AI engine onboard PhiSat-1 as it whizzes around the Earth at 27,500km/h and around 530 km up in space.

The successful launch and commissioning of the satellite is the result of almost three years of work. This included testing the COTS (Commercial Off The Shelf) Myriad 2 device in the CERN LHC/SPS linear accelerator and other radiation testing facilities around Europe to determine its suitability for space applications and its susceptibility to soft errors. Subsequently the launch itself was delayed by over a year by a failed rocket, two natural disasters, and a global pandemic.

According to Dr. Gianluca Furano, the ESA data handling engineer behind PhiSat-1, although it is now common practice to build AI-enabled products for terrestrial use, it had never been done before for space applications, making it an enormous challenge.

"To use AI in a data-critical application is not straightforward even on Earth," he said. Flying an AI-powered chip like the one onboard PhiSat-1, 329 miles above the surface of the Earth, means that any repair, software patches, or upgrades have to be done via the satellite's comms uplink.

"Whatever has silicon inside is disturbed by ionizing radiation," Furano added. This means computer chips that would function perfectly on Earth would have unworkably high error rates, and could even catch fire (due to internal radiation-induced power latch-up) if sent into outer space without shielding or modifications.

But while lots of satellites carry custom-built spaceworthy chips, PhiSat-1 uses Intel's Myriad 2 chip, a commercially available chip found in DJI drones, Google Clips and surveillance cameras and even Magic Leap's AR goggles. The space processors typically lag more than a decade behind their commercial counterparts in terms of performance and the gap is widening every year; Myriad 2 provides a sudden performance 'step' for in-flight avionics of more than 100x with respect to comparable systems.

Approximately two weeks after launch, the satellite was commissioned and began transmitting its first images back to Earth.

The initial application is to filter out clouds, which obscure earth-observation imagery and which can account for 68% of such imagery on average. Filtering out clouds at source allows precious downlink bandwidth and power to be conserved.

The frontend of the PhiSat-1 imaging system is called HyperScout-2, and, unlike an RGB camera which processes three spectral channels, it processes 48 distinct spectral bands from visible light to infrared.

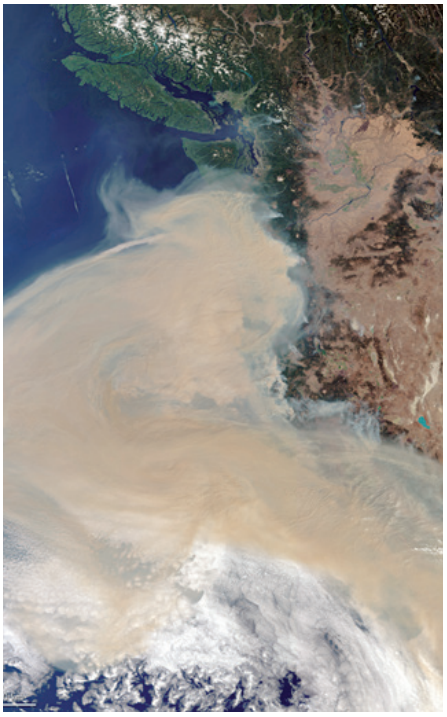
Selecting particular spectral bands to analyze for particular applications makes it possible to see things that no ordinary image would show - for instance the red carotene of drying leaves, which indicates areas at elevated risk of forest fires. Having AI inference onboard opens the prospect of identifying forest fires onboard the satellite by examining the IR bands from the sensor, with consequent reductions in the time to generate alerts for forestry managers.

The imager and associated AI engine are not limited to observing terrestrial areas but can automatically detect algae, oil spills and ships in the oceans and inland waterways. The fact that the platform is programmable allows AI inference networks to be uploaded over the satellite communications channel in a manner familiar to app store users, allowing existing satellites to support a myriad of new applications while in orbit.

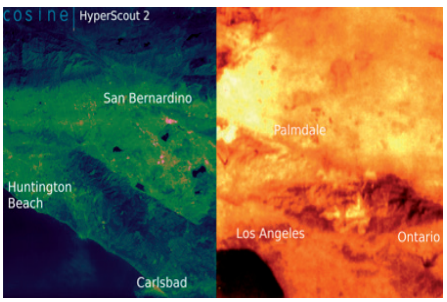
This could pave the way for future space-based innovations, from fighting fires and oil spills on Earth, to piloting spacecraft and landers, to driving rovers on Mars.

With a view to the future, Ubotica plans to launch a Myriad X based next generation platform with a space partner in early 2022 as well as a programmed mission to the ISS in 2021.

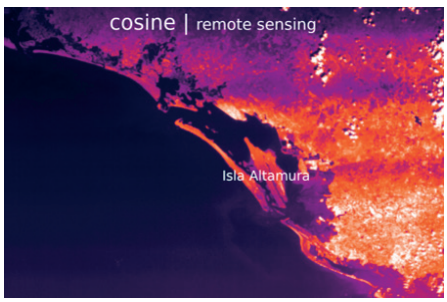
For any HiPEAC readers interested to get a closer look at this leading-edge technology, Ubotica has several internship positions:
hipeac.net/jobs



US west coast wildfires, September 2020. Image: contains modified Copernicus Sentinel data (2020), processed by ESA, CC BY-SA 3.0 IGO creativecommons.org/licenses/by-sa/3.0/igo/



PhiSat-1 hyperspectral and thermal infra-red images of California coast. Credit: cosine measurement systems



PhiSat-1 hyperspectral image of Gulf of California. Credit: cosine measurement systems

Huawei's Bill McColl, Director of the Future Computer Systems Lab at the Huawei Zurich Research Center, will speak at the conference Industry Session on 19 January. He tells us about the work that goes on in Zurich and the challenges that he and colleagues look forward to tackling.

HiPEAC 2021 gold sponsor: Huawei

Huawei is one of the world's largest and most successful technology companies, providing servers and storage systems, cloud services, communications and networking systems, smartphones and other consumer devices. Its HiSilicon department produces ARM-Based CPUs, AI Accelerators and many other types of chips and SoCs. It also produces world-leading high performance AI Accelerator Clusters.

In recent years, Huawei has opened new research centres in Europe and now has more than twenty across the continent. We have multiple research centres in each of the UK, France, Germany, Italy, Finland and Sweden. We also have research centres in Ireland, Israel, Poland and Ukraine, and the number of locations continues to grow. In addition to carrying out leading-edge research internally, a major goal of our research centres is to build academic research partnerships with leading universities across Europe. Our aim is to fund, and work with, professors, postdocs and students to drive research and innovation through specific projects, gift funding, and joint labs.

One of our newest research centres is in Zurich, Switzerland. The city is a major European hub for technology research and innovation, with major universities and a number of research centres of leading technology companies.

A major part of the research in the Huawei Zurich Research Center is within a new Lab devoted to Future Computing Systems. The Lab focuses on fundamental long-term challenges in architecture - hardware, software, and algorithmic challenges.

In the hardware area, we are exploring challenges in future manycore CPU architectures and in future AI accelerator designs. We are also looking at challenges in HPC and AI system architectures.

In the software area, we are working on many areas of research in parallel computing – programming models and tools, domain-specific languages, compilers, graph and tensor computing, resilience and predictability.

We are also investigating new directions in algorithms and theory, including the development of new theoretical frameworks for cost modelling and architecture optimization, and new communication-avoiding and memory-efficient parallel algorithms that can achieve high performance at massive scale.

In all of this research, a central goal is not to separate hardware, software and algorithmic research. We believe strongly that looking at these aspects together will be key to achieving future breakthroughs in performance at scale. One aspect of this is, of course, Hardware Software Co-Design, where we aim to build new systems in which the two parts are carefully designed to work well together.

Another related, but distinct, future challenge in computer systems research is to increase performance through what one might call "software-aided hardware". Today, quite often the compilation process throws away a lot of information that was available in the high-level software. The resulting low-level code that the hardware then has to execute will typically be missing vital information that could greatly improve performance. The hardware will then have only basic general methods such as prefetching and prediction to work with. Can we develop both new advanced compiler technologies and architectures with which we can overcome this problem?

These are just a few examples of the types of fundamental computer systems research that we are interested in, and working on, in the new Lab in Zurich. If these align with your own research interests then feel free to connect with us to explore how we might work together.

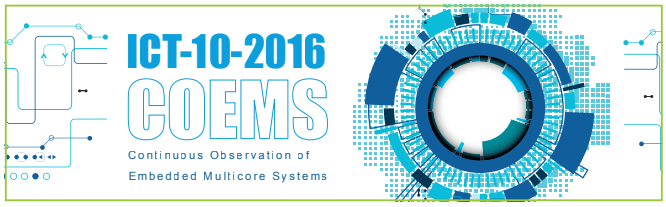
HUAWEI

The latest in our series on cutting-edge research in Europe showcases the range of computing systems research that is funded by the European Commission.

Innovation Europe

COEMS – HUNTING HEISENBUGS MADE EASY

Hannes Kallwies, Martin Leucker, Universität zu Lübeck; Alexander Weiss, Accemic Technologies



Infrequent bugs, i.e. those that seldom appear and are hard to reproduce, are a major challenge in program verification. This is particularly the case in systems which are influenced by non-determinism e.g. multi-threaded applications or embedded systems with sensor interaction. In these scenarios, traditional error detection approaches, like testing, are often insufficient, as the failures in many cases only occur during real-life scenarios.

The COEMS project, which ended in spring 2020, dealt with this issue. Its approach was to apply runtime verification techniques by directly analyzing the operations performed by the (multicore) CPUs of the embedded systems. In runtime verification, a system's execution is matched to the specified correct behaviour and any mismatch is reported as a failure. We relied on the built-in logging interfaces of today's CPUs, e.g. Arm®. For analyzing the execution, an FPGA-based, fast reconfigurable solution was developed. It can be attached directly to an examined system and observe it for failures. For specification of the expected behaviour the language TeSSLa (Temporal Stream-based Specification language) was used and refined. TeSSLa is especially suited for the simple and concise description of expected runtime behaviour, including timing constraints. COEMS also created a full tool chain to compile TeSSLa specifications to FPGA configurations automatically and supervise existent programs with low overhead.

A major benefit of this approach is that, by observing the logging interface of a processor, the execution itself is not influenced – the verification is non-intrusive. This is desirable since otherwise the timing behaviour of the supervised system could be affected, which would potentially lead to new errors. On the other hand, bugs in the system could also not occur because of an intrusive supervision and hence so called "Heisenbugs" would stay undetected despite intensive verification but reappear in an unsupervised system run. The cause of such "Heisenbugs" is, for example, branch prediction which behaves differently for an adjusted program with additional statements for verification purposes. COEMS provided the first comprehensive online observation approach that is non-intrusive.

Since the successful conclusion of the COEMS project, the team have been busy: development of the tool chain has continued, the TeSSLa language is now freely available as open-source project and the developed FPGA solution has been advanced to a marketable product.

NAME: COEMS: Continuous Observation of Embedded Multicore
START/END DATE: 01/11/2016 - 30/04/2020
KEY THEMES: observation, multicore, runtime behaviour
PARTNERS: Germany: University of Lübeck, Accemic Technologies, Airbus; Norway: Western Norway University of Applied Sciences; Austria/Romania: Thales.
BUDGET: €3.96M
WEBSITE: coems.eu

COEMS has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 732016

Read more: www.tessla.io.

MORPHEMIC – ADVANCED POLYMORPHIC AND PROACTIVE CLOUD COMPUTING RESOURCES ADAPTATION PROJECT

Alessandra Bagnato, SOFTEAM

MORPHEMIC covers areas including cloud computing, big data and open data, and artificial intelligence. twelve partners from seven countries are developing the innovative MORPHEMIC platform, which will provide a unique way of adapting and optimizing cloud computing applications for future specialized hardware configurations like GPUs, TPUs, AI chips, FPGA and HPC.

MORPHEMIC is an extension of the MELODIC multi-cloud platform and is a single universal platform and the simplest and easiest way for optimized deployment and management of applications cross-cloud.

MORPHEMIC introduces two novel concepts to cloud computing which will utilize cloud computing resources in the most optimal way. It uses the latest, state of the art methods, machine learning-based for time series forecasting like ES hybrid method, Tsetlin machine, echo state networks, temporal fusion transformer and others. For solving the optimization problems, the latest AI-based solvers are used, including the most advanced Monte Carlo tree search with neural networks. It is based on a similar concept to AlphaGO, the computer program that beat the world's best (human) players at GO. In MORPHEMIC, similar methods are adapted to optimize cloud computing deployments.

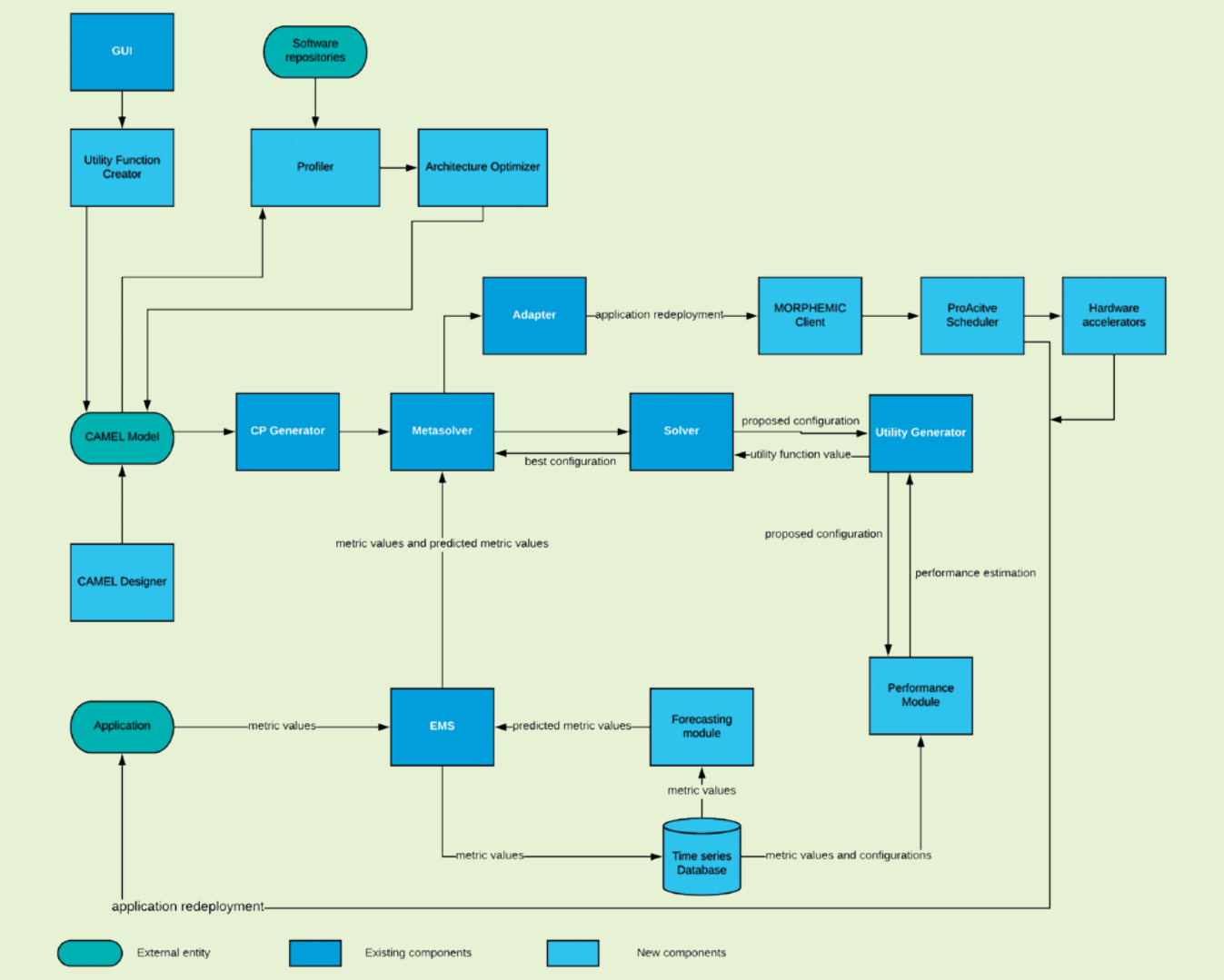


Figure 1 Initial MORPHEMIC architecture

Two innovative pillars of MORPHEMIC:

1 Polymorphing architecture will allow for dynamic adaptation of the architecture for application to the current workload. When a component can run in different technical forms (i.e. in a virtual machine (VM), in a container, as a big data job, or as serverless components, etc.), depending on the application's requirements and its current workload, its components could be deployed in various forms in different environments to maximize the utility of the application deployment and the satisfaction of the user. It will change the component type (VM, container, serverless, GPU or FPGA) based on the optimal solution found.

2 Proactive adaptation will allow the reconfiguring of the application to respond to forecast future usage and workload level, to adapt the application in the most optimal way to the future requirements. It aims to forecast future resource needs and possible deployment configurations. This ensures that adaptation can be done effectively and seamlessly for the application users.

Both concepts have not been exploited yet in the cloud computing area. Cloud computing is becoming the core paradigm for green computing and computing with no initial investment in computing resources, which is important for SMEs and start-ups. The MORPHEMIC team is expecting a significant number of research projects inspired by their project outcomes.

Nowadays the existing Cloud application modelling languages do not supply the polymorphic application components that can be deployed on different hardware and do not provide polymorphic infrastructure models. The goal is to simplify Cloud application modelling and continuously optimize and morph the Cloud deployment model to take advantage of beneficial Cloud capabilities.

NAME: MORPHEMIC: Modelling and Orchestrating heterogeneous Resources and Polymorphic applications for Holistic Execution and adaptation of Models In the Cloud
START / END DATE: 01/01/2020 - 31/12/2022
KEY THEMES: cloud computing, polymorph architecture, edge, data centre
PARTNERS: University of Oslo (Norway), 7bulls.com Sp. z o.o., Is-Wireless (Poland), Foundation for Research and Technology Hellas, CCG Inaccel Private Company, Institute of Communication and Computer Systems, University of Pireus Research Center (Greece), Softeam, Activeeon (France), ENGINEERING - Ingegneria Informatica SPA (Italy), Lausanne University Hospital (Switzerland), ICON Technology and Process Consulting Limited (UK)
BUDGET: €4.99M
WEBSITE: morphemic.cloud

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 871643.

SODALITE: AT THE FOREFRONT OF THE DIGITAL TRANSFORMATION OF EUROPEAN INDUSTRY

Daniel Vladušič, Elisabetta Di Nitto, María Carbonell, Paul Mundt, Joao Pita Costa, Lara López, Dragan Radolović, Nejc Bat.

The global market has seen a tremendous rise in utility computing, which serves as the back end for practically any new technology, methodology or advancement from healthcare to aerospace. SODALITE's vision is to support the Digital Transformation of European Industry through increasing design and runtime effectiveness of software-defined infrastructures, to ensure high-performance execution across dynamic and heterogeneous execution environments; and increasing the simplicity of application and infrastructure modelling to improve manageability, collaboration, and time to market.

SODALITE has released the first version of its open source toolset, including the initial versions of the SODALITE IDE, deployment blueprints, the xOpera orchestrator, and additional components with which to quickly get up and running.

A powerful solution fit for cloud, HPC and edge computing

SODALITE focuses on reducing the complexity of deploying and operating modern applications across heterogeneous HPC, cloud, and edge-based software-defined infrastructures by increasing design and runtime effectiveness and increasing the simplicity of application and infrastructure modelling.

SODALITE enables developers to describe their applications once, using high level semantic abstractions, assisted by the SODALITE IDE, that target generic and extensible abstracted execution platforms to bring-up a runtime bootstrap environment adaptable across a range of infrastructures.

▶

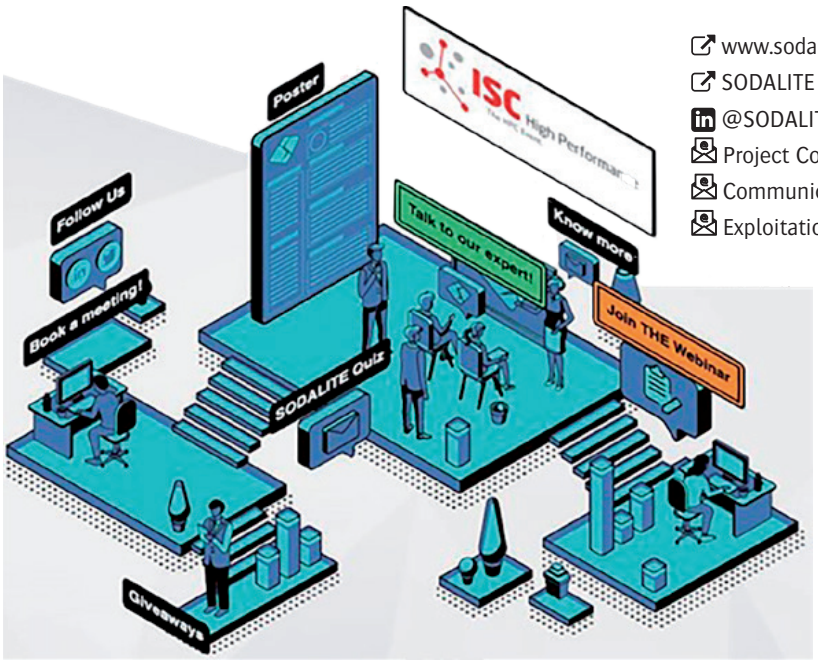
SODALITE use case domains

Immediate impact

SODALITE is achieving impact across a diverse range of use cases. In the health context, it is being used to improve performance of in-silico clinical trials. With regards to climate change, it is being used to leverage GPU and HPC resources to more effectively monitor snow levels in mountain ranges. In the automotive context, SODALITE enables adaptive application and deployment reconfiguration of connected vehicle services, leveraging heterogeneous compute resources in cloud/edge environments for more power- and cost-effective computation and processing of IoT and vehicular sensor data.

Showcasing sodalite advancements in COVID-19 times

SODALITE innovates not only technologically, but also in terms of impact and communication. During the first lockdown period, coinciding with the biggest HPC event in Europe, ISC HPC 2020, a new method of exhibition was needed to engage with virtual attendees. SODALITE developed a virtual booth, including a live chat, interactive games, and registration for follow-up events, such as webinars. See the project website for more.



Virtual booth deployed by SODALITE at ISC HPC 2020

NAME: SODALITE : SOftware Defined AppLication Infrastructures managemenT and Engineering
START / END DATE: 01/02/2019 - 31/01/2022
KEY THEMES: digital transformation, heterogeneity, HPC, cloud, application software
PARTNERS: XLAB (Slovenia), High Performance Computer Center University of Stuttgart- HLRS (Germany), ATOS (Spain), Politecnico di Milano (Italy), ADAPTANT Solutions (Germany), CRAY HPE (Switzerland and U.K), IBM (Israel), ITI-CERTH Information Technology Center (Greece), Jheronimus Academy Data Science (Netherlands).
BUDGET: €4.99M

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 825480.

SODALITE team members in Milano General Assembly

www.sodalite.eu
SODALITE toolset: github.com/SODALITE-EU
@SODALITE.EU
Project Coordinator: daniel.vladusic@xlab.si
Communication Manager: maria.carbonell@atos.net
Exploitation & Innovation Manager: paul.mundt@adaptant.io

DEEP-EST: A MODULAR APPROACH TO FUTURE HPC SYSTEMS

Jan Schulze, Leibniz Supercomputing Centre

Not all workloads are the same. Many combine data analytics, sheer number-crunching, machine learning – it is hard to optimize system usage, energy efficiency and execution performance with a single architecture. A modular system running workloads on hardware that's best suited for the specific case will provide superior potential for the next generation of applications. That's the goal of the DEEP-EST (Dynamical Exascale Entry Platform - Extreme Scale Technologies) project.

DEEP-EST is the third in a series of EU-funded research projects on Exascale technologies. It developed and built a prototype of the Modular Supercomputing Architecture (MSA), comprising three different modules that are connected by a high-performance fabric:

1. Cluster Module (CM): the CM is a general-purpose cluster with 50 nodes, using dual Intel® Xeon® Scalable processors with high single-thread performance and warm water cooling.
2. Extreme Scale Booster (ESB): the ESB provides 75 nodes, each equipped with a single Intel® Xeon® Scalable processor and one NVIDIA® Tesla® V100 GPGPU. It's also warm water cooled.
3. Data Analytics Module (DAM): the 16 DAM air-cooled nodes are each equipped with dual Intel® Xeon® Scalable processors, one NVIDIA® Tesla® V100 GPGPU and one Intel® Stratix® 10 FPGA per node. Besides a large DRAM complement, each node has 3 TBytes of Intel® OptaneTM Persistent Memory.

All modules are connected with a high-performance network. Besides standard Infiniband (CM) and Ethernet (DAM), the project introduces a new low latency fabric developed by Extoll. The Fabri3 network provides six 100Gbit/s links per node and enables 3D grids and tori to be configured.

On the software side, the focus is on ease of programming, resource management and job scheduling. The DEEP-EST SW determines optimal resource allocations for each combination of workloads, supports adaptive scheduling and enables dynamical reservation of the resources. Six innovative European projects provided co-design input and will validate the prototype system. Space Weather, Molecular Fluid Dynamics, and other real-life cases were ported to the MSA. Additionally, an Early Access Programme (EAP) enabled academic users to intensively test the new architecture with their applications.

With the prototype set up nearing completion, DEEP-EST is coming to an end. So far, the system has proven that many applications can benefit from the modular approach. But the project also reinforced that an optimized and highly scalable software stack is crucial for overall performance and energy efficiency. So, the next part of the DEEP project series is already pending: DEEP-SEA (Software for Exascale Architectures) will start in early 2021 and address this challenge.

PROJECT NAME: DEEP-EST (Dynamical Exascale Entry Platform - Extreme Scale Technologies)
START/END DATE: 01/07/2017 - 31/03/2021
Key Themes: Exascale, Modular Supercomputing Architecture
PARTNERS: Jülich Supercomputing Centre, Intel Deutschland GmbH, Megware, Extoll, Barcelona Supercomputing Centre, Leibniz Supercomputing Centre, Astron, CERN, EPCC - University of Edinburgh, KU Leuven, NCSA, NMBU, University of Heidelberg, University of Iceland, ParTec, Fraunhofer ITWM
BUDGET: €15.3M

www.deep-projects.eu

The DEEP-EST prototype with all three modules running was set up at the Jülich Supercomputing Centre in May 2020.

BUILDING THE FUTURE USING ALL OPEN SOURCE SOFTWARE AND HARDWARE

The next generation exascale machines need to exploit locality, and support legacy and new programming models and frameworks. What's more, data movement is a source of inefficiency in modern accelerator-based systems, as well as a source of complexity in program model and software implementations. The MareNostrum Experimental Exascale Platform (MEEP) is a hardware and software development vehicle. The novel idea for the accelerator underpinning this project is the ability to execute traditional host software on the accelerator along with the highly parallel code. This provides a single physical memory space which removes data migration and management between the host and the accelerator, as well as reducing programming model complexity. By using a flexible, FPGA-based emulator, a variety of hardware and software accelerator concepts can be emulated and validated. In particular, in MEEP our accelerator provides a flexible framework to integrate a vector accelerator with a high lane count and several types of systolic arrays, all based on the open RISC-V ISA.

The MEEP project sets foundations of the next generation of supercomputers that will work towards improving Europe's future in line with the European Commission's stance on 'Shaping Europe's digital future'. It will provide a foundation for building European-based chips and infrastructure to enable rapid prototyping using a library of IPs and a standard set of interfaces to the host CPU and other FPGAs in the system using a set of standard interfaces defined as the FPGA shell. Open source IPs will be available for academic use, and/or to be integrated into a functional accelerator or cores for traditional and emerging HPC applications.

As a new digital laboratory for exploring new system concepts, MEEP will provide a platform that can provide a glimpse into the future of hardware and software. It can be used to build specialized systems that can tackle the problems of today and of tomorrow. For example, with COVID-19 or other future pandemics, it could explore new systems to be built for drug or vaccine discovery, and personalized medicine. It will create a sketch of the future, giving us a glimpse of what we can build or how to improve systems for greater HPC capabilities, improving the quality of our lives.

Coyote, made by MEEP in Europe

Early development stages for MEEP require fast, scalable and easy-to-modify simulation tools, with the right granularity and fidelity, enabling rapid design space exploration. For this reason, the team developed Coyote, which is a new open source, execution-driven simulation tool, based on the canonical RISC-V ISA, that has a focus on data movement throughout the memory hierarchy. It provides a solid foundation for a fast and flexible tool for HPC architecture design space exploration. It will enable designers to make informed decisions early in the development of new architectures, based on data movement, a key limiting factor of performance and efficiency.

Coyote is an integration of existing RISC-V tools, leveraging previous community efforts and also producing a tool that users find familiar and easy to adopt. It is based on two pre-existing tools: Spike and Sparta.

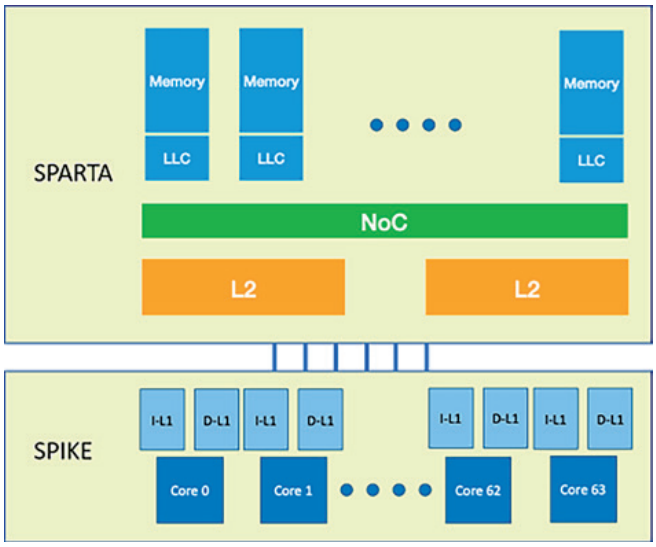


Figure 1: An example 64-core simulation target. Elements simulated by each tool shown in the figure.

NAME: MEEP: MareNostrum Experimental Exascale Platform
START/END DATE: 01/01/2020 - 31/12/2022
KEY THEMES: hardware development, software development, HPC, FPGA
PARTNERS: Spain: Barcelona Supercomputing Center; Croatia: University of Zagreb; Turkey: TÜBITAK BILGEM Informatics and Information Security Research Center.
BUDGET: € 10.38M
WEBSITE: meep-project.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme through grant agreement no. 946002.

Read more:
MEEP platform ecosystem: meep-project.eu/platform-ecosystem
RISC-V ISA: github.com/riscv/riscv-isa-manual
Spike: github.com/riscv/riscv-isa-sim
Sparta: github.com/sparcians/map/tree/master/sparta

WHAT IS SAE ALL ABOUT?

SAE
SmartAnythingEverywhere

The Smart Anything Everywhere (SAE) initiative supports various projects (so-called Innovation Actions) in different technology areas to provide small and medium enterprises (SMEs) and mid-caps with means to improve their products and services by incorporating innovative digital technologies. The SAE initiative builds on the European Commission's Digitising European Industry strategy. Its overall objective is to ensure that every industry in Europe – large or small, wherever it is located and in whatever sector – can fully benefit from digital innovation to improve its products, streamline its processes and adapt its business models to the digital age. Companies are directly supported by their local Digital Innovation Hub (DIH). DIHs are support facilities that help companies - especially SMEs, start-ups, and mid-caps - to become more competitive through the use of the latest digital technologies.

The SAE initiative started in 2015 and since then more than 200 Application Experiments (AEs) involving partners from 25 countries with a total of more than €18 million were funded in Phases 1 and 2. With the beginning of Phase 3 in 2020, a call for Smart Anything Everywhere projects in the priority area 'Digitisation and Transformation of European Industry and Services' was launched.

In this phase, €64 million will be available for Innovation Actions to stimulate the uptake of advanced digital technologies by European industry in products incorporating innovative electronic components, software and systems, especially in sectors where digital technologies are not sufficiently exploited.

How does it work?

Companies can apply for funding through Open Calls in their relevant technology area to receive financial support as well as access to digital technologies, infrastructure and networking opportunities.

A financing system called 'cascade funding' allows funds from the EU to be passed on to your company directly by the Innovation Action, without entering into a complex contract with the European Commission. Your company just signs a 'lean contract' with a representative of the Innovation Action. This reduces time, complexity and bureaucracy.

How can I participate?

If you are a European start-up, SME, or mid-cap and would like to benefit from digital innovation, check out the Open Calls on the SAE website. There, you can choose the technology area you are in and follow the instructions of the corresponding Innovation Action. Guidance such as proposal templates and instructions for the application procedure will be provided.

The Smart Anything Everywhere initiative is supported by Smart4Europe2, a Coordination and Support Action funded by the European Union's Horizon 2020 Research and Innovation programme under grant agreement No. 872111.

For more information on how to finally put your digital innovation project into action:
smartanythingeverywhere.eu
@SAE_Initiative
linkedin.com/company/smart-anything-everywhere-initiative/

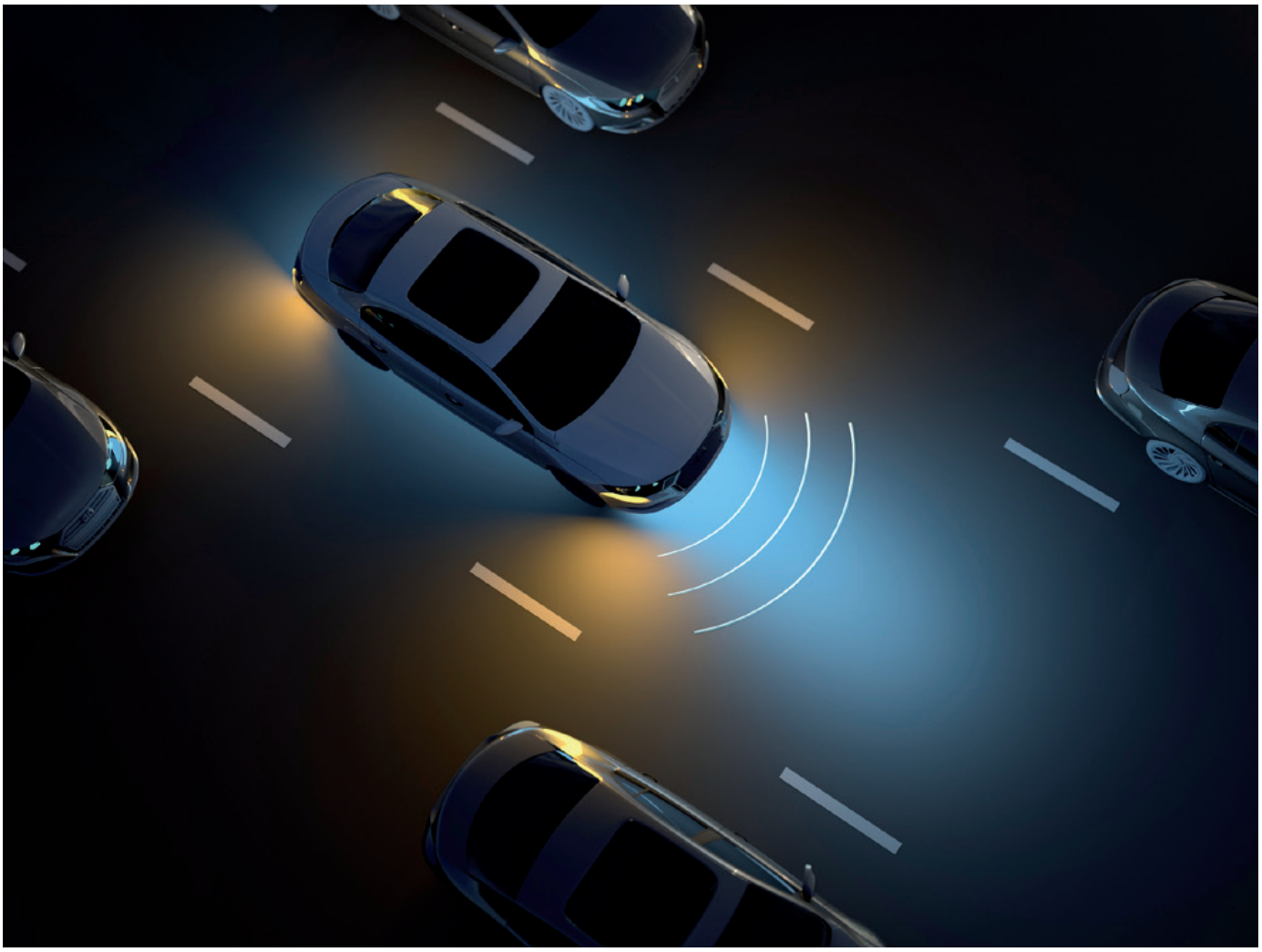
New techniques being developed at the University of Cambridge have the potential to enable a fundamental shift in the acceptability and trustworthiness of future autonomous systems, and have led to a spinout: Signaloid. Vasileios Tsoutsouras of the Physical Computation Laboratory tells us a little about it.

A new approach to computation that interacts with the physical world

Our research group investigates new ways to exploit information about the physical world to make computing systems that interact with nature more efficient. Our research applies this idea to new hardware architectures for processing noisy/uncertain data, new methods for learning models from physical sensor data, new approaches to differential privacy which exploit knowledge of physics, and new methods for synthesizing state estimators (e.g., Kalman filters) and sensor fusion algorithms from physical system descriptions.

Existing computing systems largely treat environmental sensor measurements as though they were error-free. As a result, computing systems that consume sensor data and which implement algorithms such as obstacle avoidance may perform billions of calculations per second on values that might be far removed from the quantities they are supposed to represent. When these algorithms control safety-critical systems, unquantified measurement uncertainty can inadvertently lead to failure of subsystems such as object detection or collision avoidance. This can in turn lead to injury or fatalities and the prospect and evidence of such failures reduces trust in autonomous systems.

With the ever more pervasive use of sensors to drive computation and actuation such as in autonomous vehicles and robots which interact with humans, there is a growing need for computing hardware and systems software that can track information about uncertainty or noise throughout the signal processing chain. The head of the Physical Computation Laboratory, Phillip Stanley-Marbell, saw an opportunity to enable a fundamental shift in the acceptability and trustworthiness of future autonomous systems. Previously, while working at MIT on methods for trading noise and uncertainty in sensing systems for lower power dissipation, he had taken part in an entrepreneurship/accelerator program which provided an opportunity for him to engage with potential customers at large semiconductor and consumer electronics companies. His research at the time had led to several patent filings (two of which were recently granted as US Patent 10,601,452 and US Patent 10,539,419) and, in talking with potential customers/licensees of those inventions, he realized there was an unmet need for new architectures for computing on uncertain data for robust and trustworthy machine intelligence and autonomous systems. After moving to the University of Cambridge as a faculty member in 2017, he started a new research programme to address these unsolved challenges. That research led quickly to a demonstration of a novel solution to the problem and a grant from the UK Engineering and Physical Sciences Research Council (EPSRC) to fund translation of the research into industrial impact. And so a spinout was born: Signaloid.



Signaloid technologies could contribute to the control systems of the autonomous vehicles of the future. Image: ID 192724551 ©Jackymkleung Dreamstime.com

Signaloid was incorporated in June 2019. We are currently bootstrapping and working towards releasing our initial product for our first paying customers. What makes this company special? Every member of the team is convinced that what we are doing in Signaloid is the most exciting problem any of us has ever worked on. We have a strong team with previous experience from industry (Lucent / Bell Labs, Philips Consumer Communications, NEC Labs, IBM, and Apple) and academia (University of Cambridge, National Technical University of Athens, TU Eindhoven, Carnegie Mellon, and MIT). The team is actively growing and our most recent team member, Orestis Kaparounakis, joined in September 2020.

Our vision is to enable sensing and computing on uncertain data for robust and trustworthy machine intelligence and autonomous systems. The potential benefits of our technology include, for example, safer and more trustworthy control systems in autonomous vehicles. Our methods would enable their underlying signal processing of sensor data to track the uncertainty of measurements (e.g. from LIDAR) and hence ascribe degrees of uncertainty to their computational results (e.g. whether or not there is a human in their path). These fundamental new capabilities have the scope for impact in other research disciplines as well as in a broad range of commercial applications.

READ MORE:

The research group investigates new ways to exploit information about the physical world to make computing systems that interact with nature more efficient.

PAPER:

Tsoutsouras, V., Willis, S., & Stanley-Marbell, P. (2020). Deriving equations from sensor data using dimensional function synthesis. To appear in Communications of the ACM.

https://signaloid.com
info@signaloid.com

Nostrum Biodiscovery, a spin-off of Barcelona Supercomputing Center, has begun using a technology transferred directly in from BSC. Ezequiel Mas Del Molino, company CEO, tells us about how the technology in question – pyDock – is being integrated into NBD's activities.

pyDock: specialized software enabling drug design and discovery

Nostrum Biodiscovery announced in late 2020 that a Technology Transfer Agreement had been signed with the Barcelona Supercomputing Center. The agreement includes the transfer of pyDock software [1] (10.1002/prot.21419) to NBD's cutting-edge computational methodologies portfolio, thus extending our expertise to an essential area of computational biology and drug design: the structural prediction of protein-protein interactions.

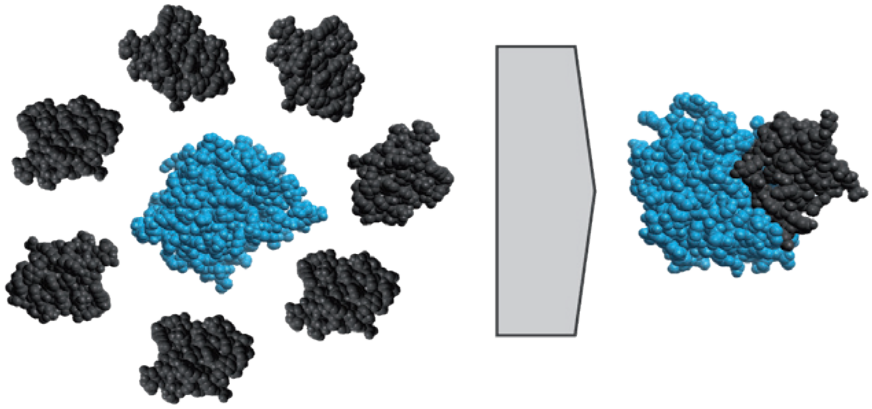
What is a protein-protein interaction?

Protein-protein interactions are involved in virtually all relevant biological processes and several pathological conditions [2] (10.1016/s0092-8674(00)80922-8) and are therefore highly attractive targets for drug discovery. Constant advances in X-ray crystallography and nuclear magnetic resonance (NMR) techniques have produced a wealth of structural data on protein-protein complexes, which has greatly extended overall knowledge on protein association mechanisms and has fostered rational drug design targeting protein-protein interactions. However, such structural data still covers only a tiny fraction of the estimated number of protein-protein complexes of the proteome [3] (10.1038/nmeth.2289, 10.1038/nmeth.1280). For this reason, it seems clear that protein-protein docking methods, such as pyDock, which allows the structural prediction of new complexes starting from the single component structures (generally much easier to be solved), would be highly useful in a large number of structural biology and drug design projects.

How does pyDock help?

pyDock is a protein-protein docking program originally developed by the Protein Interaction and Docking Group, led by Dr. Juan Fernandez Recio of the Barcelona Supercomputing Center. pyDock standard protocol consists of the generation of rigid-body docking poses based on both surface complementarity and basic electrostatics by means of FTDock program [4] (10.1006/jmbi.1997.1203) and their subsequent evaluation using a specifically designed protein-protein scoring function based on desolvation, electrostatics and Van der Waals energy. In addition to its standard protocol, pyDock allows the integration of supplementary information in order to guide the complex structure prediction by using optional modules. Some examples are pyDockTET [5] (10.1186/1471-2105-9-441) that enables the modelling of multidomain protein, as well as pyDockSAXS [6] (10.1093/nar/gkv368) and pyDockRST [7] (10.1016/j.jmb.2006.01.001), which allow the integration of SAXS profiles and external data-based distance restraints within the evaluation of docking poses.

Aside from the protein-protein docking framework, other additional prediction tools are perfectly integrated into the pyDock package: 1) ODA (Optimal Docking Area), for the identification of surface areas with higher tendency to be buried upon protein-



Pydock-figure 1

protein association [8] (10.1002/prot.20285); 2) pydockNIP, for the prediction of hot-spot residues in a specific protein-protein complex based on the identification of the higher interface propensity residues within a docking simulation [9] (10.1186/1471-2105-9-447); and finally, 3) pyDockEneRes for the prediction of the most energetically relevant residues within a complex by the estimation of the corresponding binding affinity changes upon mutation to alanine [10](10.1093/bioinformatics/btz884).

How has pyDock evolved and been evaluated?

Thanks to the work of Jimenez and co-workers [11](10.1093/bioinformatics/btt262), a new version of the pyDock scoring algorithm has been recently published, consisting of a new custom parallel FTDock implementation, with adjusted grid size for optimal FFT calculations, and a new parallel version of pyDock code, which dramatically speeds up its docking simulations by a factor of up to 40 and paves the way for the application of pyDock within HTC (High-throughput computing) and HPC (high performance computing) frameworks.

Since its first publication, pyDock has been periodically tested in CAPRI (Critical Assessment of PRediction of Interactions), a worldwide community contest aimed at assessing the performance of protein-protein docking methods in blind conditions, generally achieving excellent results. Along these lines, the all-round applicability of pyDock has been specially highlighted in the most recent editions of CAPRI, where in addition to the standard prediction of protein-protein complex assemblies, additional challenges have been proposed, including binding affinity predictions, free-energy changes upon mutation, as well as the prediction of sugar binding and interface water molecules, multimeric complexes and protein-peptide complexes. Indeed, pyDock occupied fifth position in the overall ranking in the Fifth Edition [12] (10.1002/prot.24387), and sixth position in the Sixth Edition [13] (10.1002/prot.25184). Furthermore, in CAPRI Round 46, a special CASP-CAPRI protein assembly prediction challenge [14](10.1002/prot.25838), pyDock hit the top position as scorers and second position as predictors.

As further confirmation of its prowess, in the last few years pyDock has also been successfully applied in several projects. These include initiatives relating to the effect of mutations upon p53 tetramerization process [15] (10.1016/j.molonc.2014.04.002), the structural prediction of Photosystem I in complex to different electron carrier proteins [16] (10.1016/j.bbabio.2015.09.006) and the structural characterization of the human Rab1 De-AMPylation by the Legionella pneumophila SidD protein [17](10.1371/journal.ppat.1003382).

Read about NBD's latest product launch in the News section.

CAPRI: ebi.ac.uk/msd-srv/capri

References (DOI)

1. 10.1002/prot.21419
2. 10.1016/s0092-8674(00)80922-8
3. 10.1038/nmeth.2289, 10.1038/nmeth.1280
4. 10.1006/jmbi.1997.1203
5. 10.1186/1471-2105-9-441
6. 10.1093/nar/gkv368
7. 10.1016/j.jmb.2006.01.001
8. 10.1002/prot.20285
9. 10.1186/1471-2105-9-447
10. 10.1093/bioinformatics/btz884
11. 10.1093/bioinformatics/btt262
12. 10.1002/prot.24387
13. 10.1002/prot.25184
14. 10.1002/prot.25838
15. 10.1016/j.molonc.2014.04.002
16. 10.1016/j.bbabio.2015.09.006
17. 10.1371/journal.ppat.1003382

The COVID-19 pandemic changed most people's food buying habits - from stockpiling groceries in the early weeks of the crisis to shopping less often, at different times and in different ways. However, these changes can easily lead to increased food waste. Somdip Dey of the University of Essex may have the answer, with the app, nosh.

Nosh Technologies

COMPANY: Nosh Technologies
MAIN BUSINESS: App for managing food stocks
LOCATION: Essex, UK
WEBSITE: nosh.tech

A startup has developed a novel artificial intelligence (AI)-based food management smartphone app, 'nosh - Food Stock Management' to help households keep better track of their food stocks during the pandemic.

"When the COVID-19 crisis started, everyone was in the same boat and many started to stockpile food in order to isolate at home. However, the biggest issue faced was managing a lot of food items in the fridge, especially the ones which have short expiry dates," explained Somdip Dey, then a PhD candidate in On-Device Artificial Intelligence at the University of Essex, who spearheaded the project. "With the momentum of the low- and zero waste movements, many people are becoming more aware of the need to reduce waste, and technology can make that happen more conveniently."

After the Essex team, alongside colleagues in India, identified that the market lacked an app to remind households of food expiry dates in order to manage their food supplies and track buying habits, the nosh app was born. One of the app's key features is the AI algorithm, which keeps track of the user's food buying and wastage habits so that they can make informed decisions on what products to buy or not to buy the next time so as to reduce food waste.

After its initial release in April 2020, the nosh app gained hundreds of active users within a month and also became one of the top 10 new apps in the UK Google Play Store. In June 2020, Nosh Technologies was formed as a tech spinout of the University of Essex with Somdip Dey as the CEO, charged with overseeing the development of nosh app as the all-in-one food management mobile application. With a motto of "Convenience comes first", the company has a vision to make food management easy by leveraging cutting edge technology such as edge computing, computer vision, machine learning, mobile resource management and blockchain.

Key achievements:

The app has:

- Over 5000 active users across iOS and Android platforms in 151 countries;
- Won the 2020 Spring Award - Best Mobile App Design by BMA for its innovation and user experience.

The company has:

- Developed four pieces of intellectual property (IP) in blockchain, machine learning (ML) & edge computing within six months of its initial launch. One such example is the state-of-the-art receipt scanning method using computer vision-based ML approach that is capable of achieving 85% accuracy in reading items on a receipt. This method is more sophisticated than Google Vision AI in this task, and is already implemented in the nosh app as a feature to promote convenience for the users through innovation;
- Worked with Plug and Play Ventures, a top early-stage venture capitalists, through funding rounds;
- Featured in numerous top media outlets including Business Insider, Business Weekly (Cambridge), CNBC, Yahoo News, MSN News and Outlook India.

Maspatechnologies S.L. is a Barcelona Supercomputing Center (BSC) spin-off that has weathered the storm in 2020 to provide vital services to industries that depend on real-time systems.

Maspatechnologies: supporting regulatory compliance in safety-critical industries

COMPANY: Maspatechnologies S.L.
MAIN BUSINESS: Consultancy services and tools for the design, verification, validation and certification of software running on multicore processors and accelerators.
LOCATION: Barcelona, Spain
WEBSITE: maspatechnologies.com

Maspatechnologies targets critical real-time systems primarily in the avionics and automotive domains, but also in other related areas including space, railway, and robotics.

Maspatechnologies was founded in early 2020 by BSC researchers Jaume Abella and Francisco J. Cazorla and despite challenging times due to the COVID-19 pandemic, it has managed to ramp up its activities by securing several contracts with TIER and OEM companies in the avionics market.

"We've also started a strategic collaboration with Rapita Systems Ltd and Rapita Systems Inc to provide integrated validation services to their customers," said Jaume Abella. "This is an exciting time for Maspatechnologies. Despite all that 2020 has thrown at us, we've created four jobs and are hoping to increase the team to six or seven in 2021."

Maspatechnologies offers tools such as microbenchmark technology that allows the adding of a configurable load on the desired interference channels to show the effectiveness of mitigation mechanisms, as required by CAST-32A in the avionics domain, or achieving freedom from interference as required by ISO 26262 in the automotive domain. Maspatechnologies tools have been ported to a plethora of processor families of interest for those industries including NXP's QorIQ T, P and LayerScape families, Xilinx Zynq UltraScale+, Infineon's AURIX TC27x/29x/37x/39x, Cobham Gaisler's LEON3/4/5, Intel's and AMD's x86 processors, Arm-based devices, NVIDIA GPUs, and some Texas Instruments' devices.

In addition, Maspatechnologies offers hardware consultancy services to support its customers in the avionics and automotive domains in the analysis and certification of multicore-based time-critical embedded systems, providing continuous support from planning to validation and verification, to fulfill the domain-specific regulations and practice.

Maspatechnologies employees have decades of experience in computer architecture, performance analysis and optimization, hardware design, real-time operating systems and low-level hardware development and tuning (e.g. assembly code). Value is added to that expertise by the team's additional experience in functional safety for avionics, automotive and space systems, software certification processes, timing analysis, and measurement protocols (including observability and controllability aspects).

Maspatechnologies offers highly-specialized hardware consultancy services and software analysis to assist industry by tackling all timing aspects of multicores and accelerators for their adoption in safety-related systems.

Alexander Weiss and Thomas Preusser of Accemic Technologies explain an important step made in observing and addressing software errors that can hinder the performance of cyber-physical systems

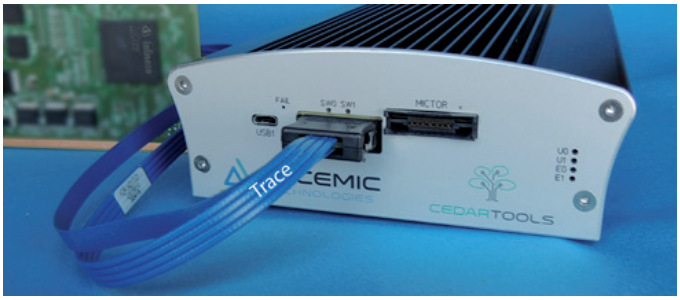
CEDARtools – look inside your processor (without affecting it)

Cyber-physical systems have become an essential contribution to our modern lives. Taking over more and more demanding and critical responsibilities, these systems are growing increasingly complex, executing sophisticated software on powerful processors. As complexity increases, so does the number of software defects, and disproportionately so.

Our new approach –pushed mainly by the Horizon 2020 COEMS project – enables the measurement of the timing behaviour of programs on multi-core architectures, the measurement of the coverage achieved by tests running on the target system, and the analysis of complex errors. The key leverage for attaining these capabilities comes from observability.

Almost all modern processors feature an embedded trace unit that provides information about the executed program flow. They expose the details of the operation of the CPU and its peripherals to the outside. However, they easily produce a few GBit of trace data per second. This quickly renders approaches combining storage and offline analysis infeasible.

We leverage embedded trace for development engineers in the best possible way. CEDARtools is a game-changing solution that analyzes the trace-data stream in real time over an arbitrary timeframe. However, there are two major technical challenges to be overcome:



1. The highly compressed trace data stream must be decompressed and the actual application control flow executed by the CPU(s) must be reconstructed. This demanding computation must often cope with the execution traces from multiple fast CPUs that are running at nominal clock frequencies far above 1 GHz. This decoding may be further challenged by additional abstractions and indirections introduced by different operating systems.
2. The reconstructed control flow must be analyzed into an apt event stream abstraction that is suitable to drive the desired of various possible backend tasks. For example, (a) branch information for the coverage analysis may be recorded or (b) dynamic properties over the event stream may be computed and validated against a temporal logic specification. The crucial step here is the provision of a convenient high-level language, in our case TeSSLa (Temporal Stream-based Specification Language), with which the engineers can specify the temporal behaviour to be investigated. These constraints are compiled for special dataflow processors implemented in an FPGA. This allows engineers to exploit the potential of high-end FPGAs without being experts in logic synthesis themselves.

The live analysis of trace data over arbitrarily long program runs enables (a) the measurement of the control flow coverage during the execution of integration tests and system tests, as well as (b) the dynamic constraints monitoring, which can be used to validate (b1) the correct coupling of the data and control flow as well as (b2) the runtime behaviour of an application.

This work was funded with funds from the EU H2020 project 732016 "COEMS", and from the BMBF research projects "ARAMiS 2" (FKZ 01IS16025) and "CoCoSI" (FKZ 01IS19044).

CEDARtool: accemic.com/cedartools/
TESSLA: tessla.io

Jorge Fernández-Berni, Associate Professor at the University of Seville summarizes the achievements and impact of research recently published in the IEEE Internet of Things Journal.

PreVIous: a useful tool for decision making in the deep learning jungle

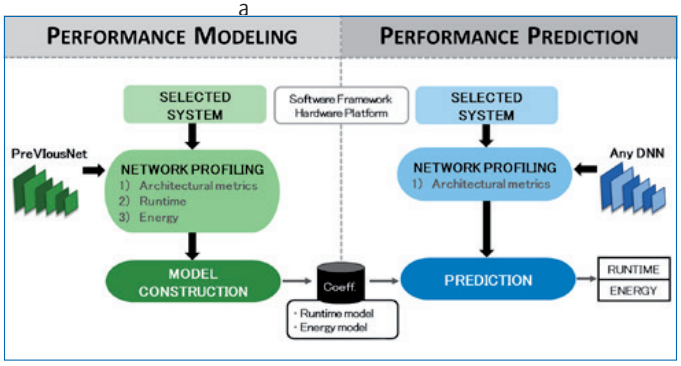
Back in 2012, AlexNet [1], a deep neural network (DNN) fundamentally based on trainable convolutions, accomplished an unprecedented boost in accuracy in the ImageNet Large-Scale Visual Recognition Challenge (a benchmark that evaluates algorithms for object detection and image classification at large scale). Prior to that milestone, vision algorithms were ad-hoc pieces of engineering demanding painstaking efforts from senior practitioners to achieve moderate performance in real-world scenarios.

AlexNet proved that highly accurate visual pipelines were possible by leveraging massive databases for training, the computational power of modern processors and new algorithmic techniques. Since then, both academia and industry have been successfully pushing the limits of DNNs and creating a vast technological ecosystem to make the most of them. In fact, the number of hardware platforms, software frameworks and model architectures tailored for DNN-based vision is now overwhelming.

From the point of view of application developers, the optimal selection of suitable components for prescribed specifications can be a really daunting task. It is in this scenario that PreVIous [2], illustrated below, can make a difference. This methodology can be summarized as 'one characterization to rule them all'. The 'one' characterization is that of PreVIousNet [3], a neural network specifically designed to build accurate per-layer performance predictive models.

The procedure to follow is simple: PreVIousNet is run on the target system and time and energy measurements are extracted; these measurements allow the construction of a prediction model that provides accurate performance estimates of any DNN considered for deployment on the system. The key is the architecture of PreVIousNet, which includes nine different layers with different configurations to cover most of the usual hyper-parameters of state-of-the-art DNN architectures.

a



General overview of PreVIous: the characterization of PreVIousNet on the target system feeds the construction of a model to predict the performance of any other DNN.

The prediction ability of PreVIous has been comprehensively tested. Seven popular DNN models were experimentally characterized on two different software frameworks and hardware platforms. These characterizations were compared with the predictions provided by the corresponding models resulting from PreVIousNet profiling. The average deviation of the predictions ranged from 3% to 10%, which represents state-of-the-art precision. All in all, we can conclude that PreVIous is a powerful tool that expedites the exploration of DNN components and subsequent decision making in practical realizations of artificial vision.

http://achieve-itn.eu/

READ MORE:

1 A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proceedings of NIPS, 2012, pp. 1097–1105.
2 D. Velasco-Montero, J. Fernández-Berni, R. Carmona-Galán, and Á. Rodríguez-Vázquez, "PreVIous: A Methodology for Prediction of Visual Inference Performance on IoT Devices," IEEE Internet of Things Journal, vol. 7, no. 10, pp. 9227-9240, 2020.
3 D. Velasco-Montero et al., PreVIous. Github Repository: https://github.com/DVM000/PreVIous.git

HiPEAC: Mobility across b

Since 2004, HiPEAC has been at the vanguard of European computing systems research. As a focal point for networking and collaboration between researchers, industry and policymakers, HiPEAC has grown year on year, and has been able to support vital researcher mobility throughout Europe. António Louro of Barcelona Supercomputing Center looks at what the trends have been and what's to come.

The HiPEAC mobility schemes have facilitated and improved collaboration between academia and research institutes and industry, with students having the opportunity to get hands-on experience and develop their work in new and different environments.

Mobility is extremely important for research, as it not only allows researchers to get feedback and input from professionals in other sectors and areas of expertise, but also provides an opportunity to apply theoretical knowledge in often real case scenarios or at least in a practical setup.

These ventures have proven to be fruitful. Since 2006, more than 350 people have been supported by HiPEAC mobility schemes. Many of the young PhD and graduate students who previously took part in these activities are now esteemed professionals, established senior researchers, and project managers. They therefore form an unofficial network of HiPEAC ambassadors, continuing to be involved in the network, and promoting its services and activities in their communities. Currently, HiPEAC offers two mobility activities: Internships and Collaboration Grants. While the Internships programme aims to develop recent graduate and PhD students through a first professional experience in industry, the Collaboration Grants are more focused on research, allowing students to continue and complement their research in a different institution. Participants in both schemes wrap up their placement by writing a report about the work developed, which in the case of Collaboration Grants often leads to publications.

Mobility in 2020

In 2020, the world stood still. Because of the COVID-19 pandemic, the travel industry came to a halt and the old European borders rose once again to restrain the free movement and mobility of citizens.

For this reason, HiPEAC increased efforts to provide value in spite of the difficult circumstances and launched a massive campaign to promote its mobility activities. By adopting a "remote-friendly" approach, HiPEAC gave students who were unable to leave their countries a

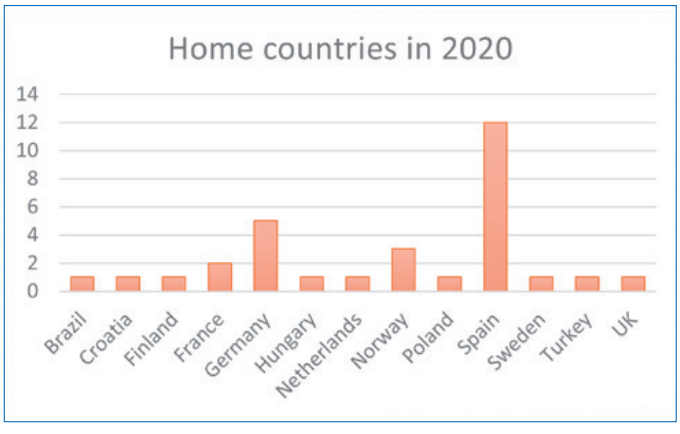
orders and across sectors

chance to enjoy the Internships programme, while launching a call for Collaboration Grants to foster mobility right after restrictions are eased or lifted. The result? 2021 is expected to be the best year for HiPEAC mobility of the last decade!

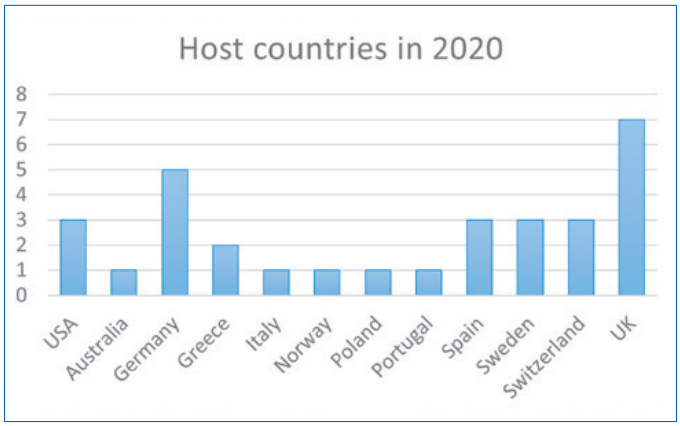
As a result of strong partnerships with industry, the HiPEAC Jobs portal had a record 78 internships posted by 29 different companies in 2020. Students can now do virtual internships without leaving their homes and still be supported by HiPEAC. Take a look at the Jobs portal to see the opportunities currently on offer.

The Collaboration Grants scheme also broke records with 56 students applying to receive a mobility grant. Due to the circumstances, HiPEAC extended the mobility period until 2022 and awarded 27 grants for a three-month onsite research collaboration in institutions all over the world.

This year, students from Spanish institutions were the most active and have shown a lot of interest in getting out in the world in order to grow professionally. Murcia University managed to send three students to research projects abroad. A Brazilian student from Rio Grande do Norte will also come to Europe to continue his work, proving that the HiPEAC network can identify and attract talent to Europe.



The United Kingdom and Germany are the preferred Collaboration Grant destinations for students. ETH Zürich, the University of Edinburgh and Uppsala University are going to host three students each, maintaining their status of excellence in the HiPEAC community. Also worth a mention are the students who were given the opportunity to go overseas to renowned institutions including Yale, Georgia Tech and Google AI in the United States and Melbourne University in Australia to complement their extraordinary work.



All in all, HiPEAC would like to thank its whole community of professionals, partner institutions and students for promoting cooperation and mobility. Even in these hard times, the HiPEAC community has managed to overcome the challenges faced and achieve better results than ever. Let's keep it up in the future and don't forget to head to the HiPEAC Jobs portal for the latest opportunities!

MORE INFORMATION

HiPEAC Jobs portal: hipeac.net/jobs

Internships: hipeac.net/internships

Collaboration Grants: hipeac.net/collaboration-grants

The HiPEAC network includes almost 1,000 PhD students who are researching the key topics of tomorrow's computing systems. This issue features a thesis by Anastasios Papagiannis, now a research assistant in the Computer Architecture and VLSI Systems Lab at the Foundation for Research and Technology Hellas (FORTH).

Three-minute thesis

NAME: Anastasios Papagiannis
RESEARCH CENTRE: Computer Science Department, University of Crete
ADVISOR: Prof. Angelos Bilas
DATE DEFENDED: October 2020

Memory-Mapped I/O for Fast StorageX

Applications typically access storage devices using read/write system calls. Additionally, they use a storage cache in DRAM to reduce expensive accesses to the devices (Figure 1 (a)). Fast storage devices provide high sequential throughput and low access latency. Consequently, the cost of cache lookups and system calls in the I/O path becomes significant at high I/O rates.

My thesis proposes the use of memory-mapped I/O to manage storage caches and remove software overheads in the case of hits. With memory-mapped I/O (i.e. Linux mmap), a user can map a file in the process virtual address space and access its data using processor load/store instructions. In this case, the operating system is responsible for moving data between DRAM and the storage devices, creating/destroying memory mappings, and handling page evictions/writebacks. The most significant advantage with memory-mapped I/O is that storage cache hits are handled entirely in hardware through the virtual memory mappings. This is important as it leaves precious CPU cycles for application needs.

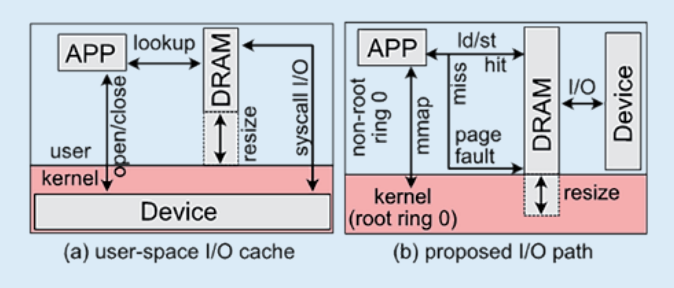


Figure 1 (a) common way to do storage caching and (b) our proposed design.

First, we design and implement a persistent key-value store that uses memory-mapped I/O to interact with storage devices, and we show the advantages of memory-mapped I/O for hits compared to explicit lookups in the storage cache. Then we show that the Linux memory-mapped I/O path suffers from several issues in the case of data-intensive applications over fast storage devices when the dataset does not fit in memory. These include: (1) the lack of user control for evictions of I/Os, especially in the case of writes, (2) scalability issues with increasing the number of threads, and (3) the high cost of page faults that happen in the common path for misses.

Next, we propose techniques to deal with these shortcomings. We propose a mechanism that handles evictions in memory-mapped I/O based on application needs. To show the applicability of this mechanism, we build an efficient memory-mapped I/O persistent key-value store that uses this mechanism. Subsequently, we remove all centralized contention points and provide scalable performance with increasing I/O concurrency and number of threads. Finally, we separate protection and common operations in the memory-mapped I/O path. We leverage CPU virtualization extensions to reduce the overhead of page faults and maintain the protection semantics of the OS (Figure 1 (b)).

We evaluate the proposed extensions using mainly persistent key-value stores, both in-house and widely used production-ready, that are a central component for many analytics processing frameworks and data serving systems. We show significant benefits in terms of CPU consumption, performance (throughput and average latency), and predictability (tail latency).

In this issue's second three-minute thesis, Florian Westphal tells us about new ways to make machine learning more efficient.

Three-minute thesis

NAME: Florian Westphal
RESEARCH CENTRE: Blekinge Institute of Technology (Sweden)
ADVISORS: Prof. Håkan Grahn, Prof. Niklas Lavesson
DATE DEFENDED: September 2020

Data and Time Efficient Historical Document Analysis

In machine learning, and in particular in supervised learning, the most expensive resource is user time, which is required to provide labelled samples for training. This thesis aims to reduce this time by focusing on approaches which allow users to guide the learning process by interacting with the learner. Facilitating this interaction is challenging. On the one hand, interfaces allowing users to efficiently make use of their cognitive abilities and domain knowledge may be compute-intensive. On the other hand, change should be immediate to allow users to perform their guidance in an iterative fashion.

This thesis explores these challenges in the context of historical document analysis. We focus on image binarization (the separation of text foreground from page background), character recognition and word spotting (the task of finding instances of a particular word image either based on a given sample image or search string).

Guided machine learning for document analysis

In a first step, we have explored heterogeneous computing, parameter prediction and throughput enhancement to speed up image binarization. With the help of heterogeneous computing and parameter prediction, we have reduced the execution time of a state-of-the-art binarization algorithm by a factor of 3.5 and 1.7 respectively. Furthermore, we have increased the prediction throughput of a recurrent neural network-based binarization algorithm to speed up training and inference by a factor of 4.

We allow users to guide the learning process of this algorithm by visualizing its current binarization quality together with its uncertainty on a new image. This aids users in choosing which pixels to label to improve the algorithm's binarization quality after re-training on the chosen pixels. This approach has shown a slight improvement over randomly labelling pixels.

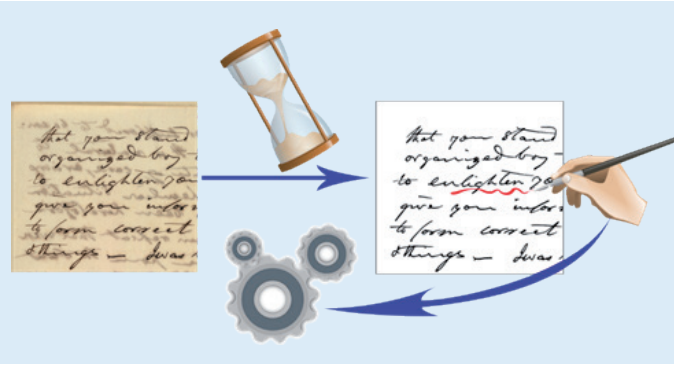


Illustration of the overall thesis goals for the example of image binarization.

Furthermore, we have explored the learning using a privileged information framework as a means to replace low-level label information with more high-level information in an attempt to reduce the number of interactions between user and learner. Lastly, we have proposed an active learning-based approach for automated sample selection to reduce the user's labelling effort by up to 69% for word spotting.

In conclusion, this thesis has explored guided machine learning as a possible way to help users to adapt a machine learning algorithm to their particular application context. This is a challenging approach, since it requires high-level mechanisms to transfer knowledge from the user to the learner in few interactions, as well as fast processing of the provided feedback to allow iterative improvements.

Photo: Nick Harrison

Thanks to all our sponsors for making #HiPEAC21, our first ever virtual conference, a great success!

arm DeepMind HUAWEI monetdb Raytheon

Reservoir Labs STREAM High Performance Computing Google MONT BLANC

European Processor Initiative epi Springer SUNDANCE THALES evolve

XILINX

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

Join the community

@hipeac hipeac.net/linkedin hipeac.net

This project has received funding from the European Union's Horizon2020 research and innovation programme under grant agreement no. 871174