HiPEAC info 65

JANUARY 2022

HiPEAC Winter Science Festival

Reaching for the stars: Sky-high ambitions for European tech

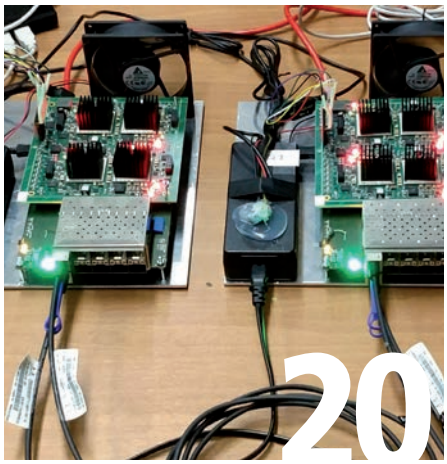
EU Missions for a more sustainable future

The real impact of European funding in deep tech

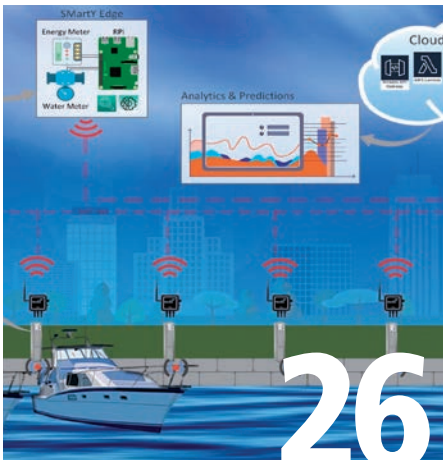
contents

EU Missions

Reaching for the stars in European tech



How EU investment is nurturing new companies



26

SMARTY4ALL-powered technology transfer

29

Modelling and simulation of COPD patients

36

Why it's time to embrace RISC-V

First of all, I would like to wish you a healthy and prosperous 2022, personally as well as professionally. 2021 was a more normal year than 2020, but it also taught us that the coronavirus will stay with us for the years to come. The vaccines do not protect us from getting and transmitting the virus, but they protect us very well from getting ill. There is a real chance that we will get our freedom back soon. That would be very good news for the start of 2022.

In November, the COP26 climate change summit took place. There is now more of a sense of urgency than in previous years, and more governments are convinced that climate change is real, and that it is caused by the use of fossil fuels. With the European Green Deal, Europe is clearly one of the global forerunners in the fight against climate change. I am very pleased to see that citizens and businesses are actively working on decarbonization projects. The task is herculean. In a couple of decades, the whole energy system of the planet will have to be converted into a carbon-free system, we will have to rethink how we house ourselves, eat, relax, work, travel, etc. This will require a combination of lifestyle changes and technological solutions. The problem is that lifestyle changes are difficult to sell, and that some technological solutions still need to be developed. Hence, there is a lot of work on the table.

Lifestyle changes are not part of the core business of HiPEAC, but technology clearly is. We create the technology that enables not only the digital transition but also the green transition. My hope for 2022 is that we all embrace the European Green Deal and start working on technical solutions that help make it a reality.

Where should we start? As a first step, we could try to apply the goals of the European Green Deal to our personal lives and make a plan to reduce our carbon footprint by 45% by 2030 compared to 2010 – equivalent to a yearly structural reduction of 7% between now and 2025. My new year's resolution for 2022 is that I will draft such a plan for myself, start executing it in 2022 and no longer buy oil and gas by 2030.

Who wants to join me?

Koen De Bosschere, HiPEAC coordinator

HiPEAC is the European network on high performance 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: © ESA/NASA - STScI
Design: www.magelaan.be
Editor: Madeleine Gray
Email: communication@hipeac.net

With the introduction of 'Missions' in the Horizon Europe programme, the European Union (EU) is aligning research efforts behind ambitious goals. In this article, Anna Krzyżanowska, Adviser to the European Commission on 'Mission Innovation', sets out what the Missions are and why digital technology plays a pivotal role in their realization.

Mission accepted

Finding solutions to our greatest challenges by 2030

A new feature of the largest public research programme in the world, Horizon Europe, EU Missions take on the biggest challenges of our time in the areas of health, climate and the environment. Missions will support research, innovation and entrepreneurial projects to deliver responses, by 2030, to fight cancer more effectively, help communities adapt to climate change, protect and restore oceans, seas and waters, help cities transit to climate neutrality, and ensure healthy soil and food.

Missions combine policy, regulatory and finance actions to drive innovation in a variety of sectors and deliver on the EU's main priorities, including the 'European Green Deal', 'Europe fit for the Digital Age', 'Beating Cancer' and the 'New European Bauhaus'. The approach targeting cross sector and multi-sector projects will also seek synergies with the EU's green and digital transitions.

Stella Kyriakides, Mariya Gabriel and Virginijus Sinkevičius at the EU Missions launch in September 2021. Photo credit: © European Union, 2021

EU Mission implementation will go far beyond research and innovation to develop new solutions that improve the lives of Europeans. Their added value is in operating as a portfolio of actions involving different instruments, business models and public and private investments at EU, national, regional and local levels. For Missions to be successful, support from other European and national programmes will be crucial.

This new collaborative approach across policy areas, involving citizens – including young people – is a chance to move together towards a healthy, green and digital future. The abundance of data and the dropping cost of gathering and processing information have allowed the Missions to establish a clear starting point, choose a measurable ambitious goal and build a robust system for measuring progress.

Based on recommendations from data driven analysis and the collective international experience of Mission Boards, in September 2021 the European Commission announced five Missions:

Cancer

If no further action is taken, the number of people newly diagnosed with cancer every year in Europe will increase from the current 3.5 million to more than 4.3 million by 2035. This Mission intends to improve the lives of those affected by cancer by 2030 through better prevention, cure and quality of life. It will achieve a thorough understanding of cancer, prevent what is preventable, optimize diagnosis and treatment, support the quality of life of all people exposed to cancer, and ensure equitable access to the above across Europe.

Adaptation to climate change

Climate adaptation is the process of adjustment to actual or expected climate change and its effects. The Mission will support this process by connecting citizens with science and public policy, maximizing the impact of the EU's support for research and innovation and demonstrating its relevance for society.

Its focus will be on solutions and preparedness for the impact of climate change to protect lives and assets. It will include behavioural changes and social aspects by addressing new communities beyond the usual stakeholders, which will help lead to a societal transformation.

This Mission will support at least 150 European regions and communities to become climate resilient by 2030.

Climate neutral and smart cities

Over 70% of Europeans live in cities, and the number of people living in cities worldwide is expected to grow exponentially. Cities are already responsible for 75% of global emissions and net neutrality is not achievable without supporting their green and digital transformations.

Yet cities are also natural crossroads where climate change policies meet real people and thus excellent living labs not only for a variety of technologies, including energy, mobility and digital, but also for imagining and co-creating the future. The objective of this Mission is to deliver 100 climate neutral and smart cities in Europe by 2030 and inspire all other cities to follow suit by 2050.

Restoring our ocean and waters

Life on Earth depends on a healthy hydrosphere: the single connected system of the ocean, seas, coastal and inland waters, which sustains ecosystems that provide oxygen, drinking water and food. With the ocean representing one of the planet's most important carbon sinks and its resources – wind, tides and waves –providing clean energy, healthy waters also support the transition to climate neutrality, as well as being of great economic importance. Human-caused, climate-driven changes are putting our waters, and consequently our societies, at serious risk.

The Mission will help protect aquatic ecosystems and biodiversity, for example by protecting 30% of the EU sea area, as well as preventing and eliminating pollution, for example by reducing plastic litter at sea by 50%. It will identify and use similar measures for 25,000 km of free flowing rivers. In parallel, it will make the blue economy climate neutral and circular with net-zero maritime emissions.

In addition to public mobilization and engagement, a digital ocean and water knowledge system will connect and integrate data and models from the physical, chemical, geological biological and socio-economical domains. It will transform them into knowledge and make them readily available to citizens, entrepreneurs, scientists and policymakers via an innovative set of user-driven and interactive tools.

A soil for Europe

Soils are the basis of 95% of our food and other essential ecosystem services such as clean water, biodiversity and climate regulation. However, an estimated 60-70% of soils in the EU are unhealthy, making us more vulnerable to food insecurity and extreme weather events.

The Mission will put in place an effective network of 100 living laboratories and light-house projects to lead the transition towards healthy soils by co-creating knowledge, testing solutions and demonstrating their value in real-life conditions. The labs will engage with people and create effective partnerships for soil protection and soil restoration across various sectors and territories. They will contribute to sustainable farming and forestry, healthy and safe food systems, climate resilience, biodiversity, zero pollution, vibrant rural areas and other Green Deal ambitions.

EU Missions acknowledge that societal challenges need a comprehensive approach that cuts across the boundaries of policies, programmes and different levels of governance. Such collaboration would not be possible without digital means of foresight, analysis, consultation and decision making.

The Missions are expected to enable joint learning and experimentation across Europe to reach their objectives, by mobilizing communities and governments. The active involvement of Member States, regions, local authorities, researchers, innovators, the private sector, citizens, civil society and investors is a critical success factor.

FURTHER READING:

Missions in Horizon Europe
bit.ly/EU_Missions

EU Missions factsheet
bit.ly/EUMissions_factsheet

Second ITEM workshop held in conjunction with ECML

Holger Fröining, Heidelberg University

In September, the second edition of the workshop on IoT, Edge, and Mobile for Embedded Machine Learning (ITEM) took place, again co-located with the European machine learning and data mining conference ECML-PKDD.

Due to the pandemic, the workshop had to take place virtually once again. Nonetheless, aided by outstanding keynote presentations by Sat Chatterjee and Jacques Pienaar, the audience and organizers enjoyed lively discussions and interactions. The keynotes showed that logic design can be helpful in interpreting machine learning models, and that the compiler approach MLIR (Multi-Level Intermediate Representation) is now usable for production systems, not just prototypical research.

The keynotes were complemented by paper presentations in the areas of machine learning (ML) applications, methods, and hardware, from institutions including the Karlsruhe Institute of Technology, the University of Duisburg-Essen, Paderborn University, Friedrich-Alexander University of Nurnberg, the Technical University of Darmstadt, Heidelberg University, and the University of Amsterdam.

Early takeaways from the talks include, on the hardware side, an improved understanding of analogue computations and advantages of accuracy-configurable processors. On the application side we saw work on atrial fibrillation, facial recognition for assistive robotics, and wind turbine condition monitoring. With regard to methods, sum-product networks as well as cell-based neural architectures and the architecture search for ultra-low power condition monitoring were addressed.

ITEM2020 was co-organized by Gregor Schiele (University of Duisburg-Essen), Franz Pernkopf (Graz University of Technology), Michaela Blott (XILINX Research, Dublin), and Holger Fröning (Heidelberg University). Günther Schindler (Heidelberg University) acted as chair of the technical programme committee. It is planned to continue ITEM in 2022, so if you are interested, stay tuned for updates or ask the organizers to add you to the mailing list.

FURTHER INFORMATION:
item-workshop.org

MAESTRO Core C library launched

Fani Garcia, Appentra

The MAESTRO project has released Maestro Core, a C library undertaking multi-threaded cross-application data transfer over high-speed interconnect using lightweight data wrappers that include user metadata.

Maestro Core is part of the Maestro middleware framework, a data- and memory-aware abstraction for workflow coupling, inter- and intra-application data exchange and redistribution. A central design goal was to enable modelling of memory and storage hierarchies to allow for reasoning about data movement and placement based on costs of moving data objects. Data objects can also carry user-defined metadata. Leveraging the Maestro middleware allows workflow management software to reason about the data without inspecting it, at scale.

The software is still at an early stage, but has been tested in various environments with success, in terms of both functionality and performance.

Maestro Core has been produced as part of the MAESTRO project, whose middleware framework addresses ubiquitous problems of data movement in complex memory hierarchies and at many levels of the HPC software stack. The MAESTRO consortium consists of seven partners (Jülich Supercomputing Centre, CEA, Appentra, ETH Zurich (CSCS), ECMWF, Seagate, HPE), each bringing specialist knowledge and expertise to the technical challenge.

FURTHER INFORMATION:
MAESTRO Project
maestro-data.eu
Maestro Core on the Jülich GitLab
gitlab.jsc.fz-juelich.de/maestro/maestro-core
Maestro Core user documentation
maestro-core.readthedocs.io/en/stable/

HiPEAC 2022 postponed until 20-22 June 2022

Given the situation with the COVID-19 pandemic in Europe, last autumn the HiPEAC 2022 organizing committee took the difficult decision not to hold the HiPEAC conference in January. Instead, the physical conference is due to take place on 20-22 June 2022 in Budapest. A number of virtual workshops are also taking place in early 2022, as part of the HiPEAC Winter Science Festival.

HiPEAC strongly feels that holding an in-person event offers the best experience for the community. A physical event would allow face-to-face networking, the industry exhibition, STEM (science, technology, engineering and mathematics) student day and all the other elements that make the conference so special.

Workshop organizers who are unable to postpone their event until June have been offered the opportunity to host a virtual session earlier in the year. Check out the programme on HiPEAC webinars.

FURTHER INFORMATION:
HiPEAC 2022
hipeac.net/2022/budapest

HiPEAC Winter Science Festival
hipeac.net/webinars

Machine learning, hardware security and more at Computing Systems Week Lyon

Hosted by the University of Lyon, Computing Systems Week Lyon took place as an in-person event on 25-27 October. The event brought the HiPEAC community together with GDR-SOC2, the systems-on-chip and embedded systems/connected objects research group at the French national scientific research centre CNRS.

Each day kicked off with a compelling keynote: Jürgen Teich (University of Erlangen-Nuremberg) on enforcing non-functional requirements on multi-processor systems-on-chip; Benoît Dupont de Dinechin (Kalray) on the co-design of the Kalray manycore accelerator for edge computing; and David Atienza (EPFL) on brain-inspired edge artificial intelligence computing architectures. These keynote talks were complemented by an exceptional technical programme with machine learning as a recurrent theme, along with sessions on emerging and quantum paradigms, hardware security and cloud computing. Sustainability was another key topic: in addition to a dedicated session with presentations from researchers, it was the topic of a panel session on the HiPEAC Vision, moderated by HiPEAC Vision Coordinator Marc Duranton (CEA).

Once again, activities aimed at students were organized, in order to support future generations of HiPEAC members. In addition to the HiPEAC Student Challenge, which this time focused on histogram equalization, there was an 'Inspiring Futures' careers session featuring long-term HiPEAC member Albert Cohen (Google), Alberto Bosio (École Centrale de Lyon) and Babis Chalios (Nubificus). See p.37 for more details.

On behalf of HiPEAC, many thanks to Ian O'Connor, Alberto Bosio and the team at École Centrale de Lyon for the excellent organization.

FURTHER INFORMATION:
hipeac.net/csw/2021/lyon
Subscribe to #HiPEACTV on YouTube for videos of talks from Computing Systems Week Lyon, in a dedicated playlist:
bit.ly/CSWLyon_HiPEACTV

Winners of the HiPEAC Tech Transfer Awards 2021 announced

Now in its seventh edition, this year's HiPEAC Tech Transfer Awards once again demonstrate how the HiPEAC community is sparking innovation in computer architecture and compilation. The 2021 award winners – which include solutions for battery-free sensing, hardware security, compute acceleration and compiler optimization – represent a range of intellectual property transfers, from patents to products to spin-offs.

For the purposes of the awards, technology transfer is defined as a contractually documented joint- or privately funded academia-industry project or technology licence agreement, with the goal of bringing a concrete research result into industrial practice. All applications are evaluated by an internal technology transfer committee, and first-time winners are awarded the sum of €1,000 for the team that developed the technology.

This year, six winners have been selected, as follows (in alphabetical order of the applicant):

"Focusing on real-life problems often paves the way to award-winning academic papers, and then to industrial uptake and innovation"

- **Yoav Etsion, Technion**: High-performance data analytics based on coarse-grain reconfigurable arrays

Technion developed a massively multithreaded coarse-grain reconfigurable processor known as Single-Graph Multiple-Flows (SGMF), which delivers an order of magnitude better performance/power than traditional von Neumann processors and even dramatically outperforms general-purpose graphics processing units (GPGPUs). Based on the promising results of the SGMF architecture, Yoav Etsion and Dani Voitsechov founded the spin-off Speedata.

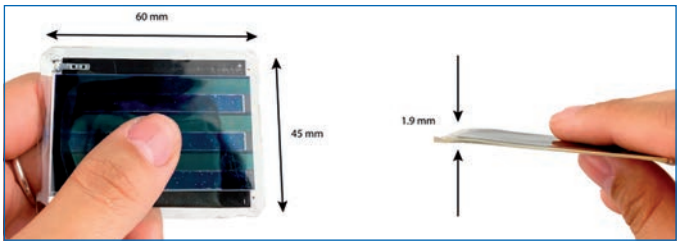
SPEEDATA

- **Tobias Grosser, University of Edinburgh**: Fast linear programming through transprecision computing on small and sparse data

This project designed a simplex solver targeted at compilers, based on a novel theory of transprecision computation, to reduce memory traffic, exploit wide vectors and use low-precision arithmetic units effectively. The solver has been shown to deliver up to an order of magnitude speedup on operations. It has been added to the LLVM infrastructure and has seeded a new polyhedral mathematics library, which can be used to optimize workloads in deep learning and high-performance computing, as well as improving hardware design flows such as Xilinx's Vivado.

- **Andres Gomez, University of St. Gallen**: miroCard, a reliable batteryless platform for wireless sensing

Developed as part of an open-source project at the University of St. Gallen, the miroCard is an innovative platform for reliable batteryless sensing, which is now being commercialized by the Swiss company Miromico AG.



- **Lukas Jünger, RWTH Aachen**: Acceleration technologies for automotive virtual platforms

As part of the 'Design for Simulation' joint research project between RWTH Aachen and the German auto manufacturer Audi, two technologies to enhance the performance of virtual automotive platforms were developed and transferred to Audi. They are as follows:

- A simulation acceleration technique that improves performance by over two and a half times.
- A framework allowing target software to be split into parts for host and parts for simulation execution, resulting in speedups of nearly eight times.

Audi has applied for patents to protect both technologies, with one patent granted already.

- **Leonidas Kosmidis, Barcelona Supercomputing Center (BSC)**: GPU4S Bench: An open GPU benchmarking suite for space on-board processing

The open-source benchmarking suite GPU4S Bench, developed in collaboration with Airbus Defence and Space, is an outcome of the GPU4S (GPU for Space) project funded by the European Space Agency (ESA) and coordinated by BSC. GPU4S Bench now forms part of the ESA suite OBPMark, a set of computational performance benchmarks developed specifically for spacecraft on-board data processing applications.

- **Davide Zoni, Politecnico di Milano**: InspectStudio: Security analysis and semi-automatic countermeasures at hardware level against side-channel attacks

InspectStudio is a configurable electronic design automation (EDA) tool that employs machine learning to analyse the register-level transfer (RTL) description of a generic computing platform to discover side-channel vulnerabilities. Using the information gathered, InspectStudio can fix the 'leaking' parts of the microarchitecture by suggesting microarchitectural alternatives that are resistant to side-channel attacks. This enables hardware designers and certification authorities to undertake vulnerability analysis in the hardware design flow. The tool and related technologies have been transferred from Politecnico di Milano to its spin-off Blue Signals Srl.

Blue Signals

'The outcome of this new round of tech transfer awards mirrors the broad spectrum of HiPEAC topics, from compilation over computing platforms to electronic design automation (EDA),' commented HiPEAC's Rainer Leupers, chair for Software for Systems on Silicon at RWTH Aachen. 'This year's winners prove once again that aiming for industrial impact in research does not imply compromising academic excellence. Indeed, there is a win-win: focusing on real-life problems often paves the way to award-winning academic papers, and then to industrial uptake and innovation. I would like to congratulate the winners, whose efforts are invaluable for our network and will inspire many more tech transfer instances.'

'Congratulations to the winners of the 2021 HiPEAC Tech Transfer Awards! It is always great to see HiPEAC members pursuing their research outcomes through technology transfer. Allowing market- or impact-driven organizations to apply research results in industrial applications can significantly amplify the economic, environmental or societal impact,' said Paul Pietrangelo, managing partner at Lira and entrepreneurial finance tutor at HiPEAC's ACACES summer school. 'With many organizations understanding the importance of technology as a competitive advantage, your research results may harness more value than you know. I therefore encourage all researchers and HiPEAC members to thoroughly investigate technology transfer: the first route may not always be the final or most impactful route.'

HiPEAC members' companies thrive in 2021

2021 saw a number of companies founded by HiPEAC members achieve new levels of success. In addition to Silexica's acquisition by Xilinx, as detailed in *HiPEACinfo* 64, there was good news for ZeroPoint Technologies, which raised € 2.5 million in seed funding. Created by HiPEAC founding partner Per Stenström (Chalmers University) and HiPEAC member Angelos Arelakis, ZeroPoint is the sole maker of a real-time memory compression intellectual property (IP) block for system-on-chips (SoC).

Meanwhile, MonetDB, founded by HiPEAC members Martin Kersten and Ying Zhang, received a second round of funding from digital workflow company ServiceNow. The investment will be used to support MonetDB's boutique serverless data warehouse solutions.

Hammering out neuromorphic security issues with NeuroHammer

Felix Staudigl and Dominik Sisejkovic, RWTH Aachen

While conventional computers – so-called von Neumann machines – are characterized by the fast and energy-efficient processing of numerical values, this widespread architecture is reaching its limits with the advent of machine learning (ML). Its fundamental advantage, the separation of the processing and memory units, has created the so-called von Neumann bottleneck. Modern ML applications require a high rate of data communication between the storage and processing units, which is now limited by the latency of the communication channel.

In contrast, the human brain with its biological neural network follows a 'computing-in-memory' paradigm. This minimizes communication by performing computation where the data is stored. Neuromorphic computing mimics this concept, enabling high data rates for ML applications while consuming a fraction of the energy of conventional systems. To implement the 'computing-in-memory' paradigm, emerging non-volatile memory (eNVM) technologies are required. eNVMs form dense crossbar structures that offer unique advantages in energy efficiency and latency.

In the past, hardware security attacks, like the famous RowHammer attack, have shown that modern computer systems are vulnerable to disturbance errors. Therefore, we asked a fundamental security question: do disturbance errors persist even in emerging neuromorphic technologies? If yes, can we exploit them? To answer this question, we introduced NeuroHammer, a security attack on eNVMs that deliberately causes bit-flips. In addition, we examined the effects of pulse lengths, ambient temperature, and electrode spacing on the effectiveness of the proposed attack. Finally, we discussed how NeuroHammer could be leveraged for a privilege-escalation attack scenario.

The detailed research results will be presented at the Design, Automation and Test in Europe (DATE) Conference in March 2022.

FURTHER INFORMATION:
date-conference.com

Czech supercomputer Karolina among the world's top energy-efficient supercomputers

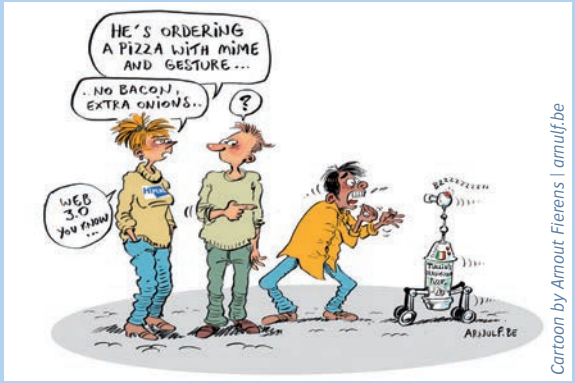
Karolina, located at IT4Innovations National Supercomputing Center in Ostrava (Czech Republic), has reached eighth place on the Green500 list, which evaluates the most energy-efficient supercomputers in the world. The new supercomputer, installed in summer 2021, reaches a theoretical peak performance of 15.7 PFlop/s and was acquired as part of the EuroHPC Joint Undertaking.

FURTHER INFORMATION:
bit.ly/Karolina_Green500

Keep up to date with the HiPEAC Vision

With the new, article-style format of the HiPEAC Vision, you can easily keep up to date on the latest technology trends and societal issues in computing. A number of articles are being updated for 2022 – check the HiPEAC website for further details.

FURTHER INFORMATION:
hipeac.net/vision/#/latest



Cartoon by Arnout Fierens | arnulf.be

Hacking mental health support with bitsxlamarató

The bitsxlamarató hackathon held on 10-12 December 2021 in Barcelona, Lleida and online aimed to improve quality of life for people affected by mental health issues.

The annual hackathon, which coincides with the 'Marató' fundraising initiative of Catalan broadcaster TV3, brought together around 190 participants on 10-12 December to support prevention and early diagnosis of mental health disorders. Organized by the Facultat d'Informàtica de Barcelona (FIB) at the Universitat Politècnica de Catalunya-Barcelona Tech (UPC), Hackers@UPC, LleidaHack, Barcelona Supercomputing Center (BSC) and l'Escola Superior d'Infermeria del Mar, it featured 25 projects in total.

Students, professors and health professionals all came together to develop IT solutions in this vital field, with teams aided by 10 mentors. In addition to talks and practical sessions, this year's edition also included meditation exercises and practical workshops on how to support people with mental health disorders. Once again, HiPEAC supported the hackathon, providing welcome packs and setting up a HiPEAC Jobs wall.

'After two years of pandemic, society is becoming overwhelmed by a tidal wave of mental health issues,' commented co-organizer Daniel Jiménez, UPC / BSC. 'I am delighted to see projects from this hackathon having an impact on major research and healthcare centres, with work continuing after the event. Among other impressive results, this particular edition gave rise to a super accelerated solution to a protein docking challenge and an operational app on equestrian therapy for children. We look forward to further work on the challenges to help those people who need support.'

FURTHER INFORMATION:
fib.upc.edu/ca/la-marato

Dates for your diary


Photo credit: BTK

HiPEAC 2022
20-21 June 2022, Budapest, Hungary
Calls for workshop papers currently ongoing
Sponsorship and exhibition opportunities available
hipeac.net/2022/budapest

ASAP 2022: IEEE International Conference on Application-specific Systems, Architectures and Processors
12-14 July 2022, Gothenburg, Sweden
Abstract submission: 14 February
asap2022.org

ASPLOS 2022: Architectural Support for Programming Languages and Operating Systems
28 February-4 March 2022, Lausanne, Switzerland
HiPEAC Paper Award conference
asplos-conference.org

DATE 2022: Design, Automation and Test in Europe
14-23 March 2022, Antwerp, Belgium and online
date-conference.com


Photo credit: Laura Vanzo / Visit Tampere

Computing Systems Week Tampere
26-28 April 2022, Tampere, Finland
hipeac.net/csw/2022/tampere

ISC High Performance 2022
29 May-2 June 2022, Hamburg, Germany
isc-hpc.com

Stars special

Shooting for the moon

Taking European technology to the next level

"Kennedy understood that many of the lasting benefits of innovation happen not just at the end of the process, but along the way, through dynamic spillovers. And in the case of America's moonshot, he turned out to be right. Much of the technology in our smartphones today can be traced back to the Apollo program and related missions."
Maria Mazzucato, Professor in the Economics of Innovation and Public Value, University College London. 'Mobilizing for a Climate Moonshot', Project Syndicate, 8 October 2019

'The race for the most advanced chips is a race about technological and industrial leadership.'
Thierry Breton, European Commissioner for Internal Market, September 2021

Traditionally, Europe has lagged behind the United States and China in terms of transforming basic research into cutting-edge, market-ready technology. This means that Europe has found itself dependent on imports from other parts of the world, making the continent vulnerable to fluctuating geopolitics. In the case of technologies as fundamental and widespread as computer chips, this can have an impact on whole industries.

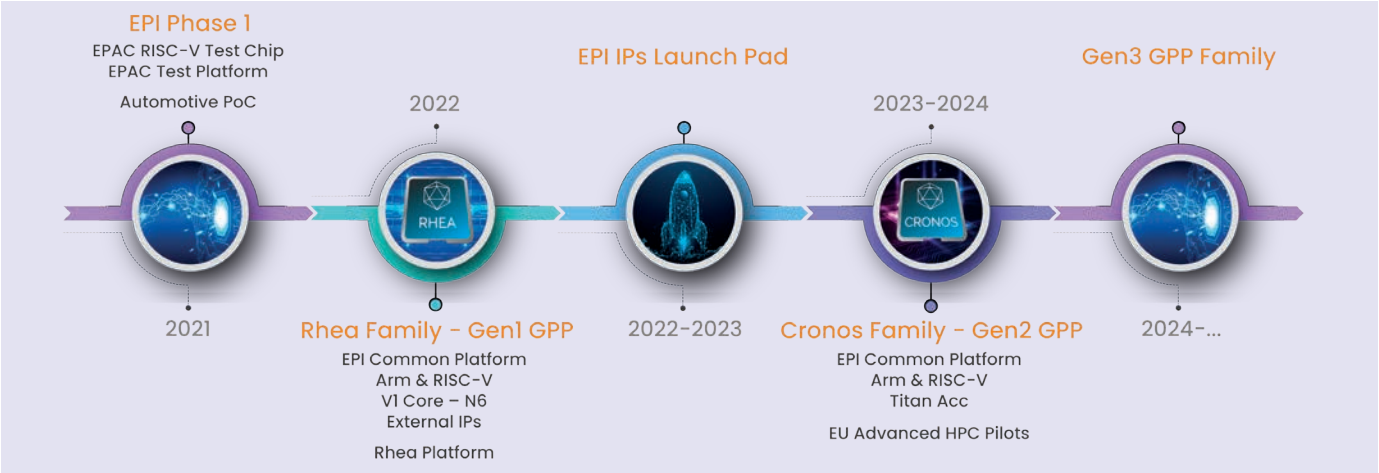
Little wonder, then, that figures such as HiPEAC co-founder Mateo Valero (Barcelona Supercomputing Center) began calling for an 'Airbus of high-performance computing (HPC)' to strengthen the European Union's chip-making capabilities some years ago. In 2018, a major programme was launched to deliver designed-in-Europe, low-power microprocessors for domestic supercomputers: the European Processor Initiative (EPI).

From the outset, it was clear that EPI would be a colossal undertaking: led by the French information and communication technology (ICT) giant Atos, by 2021 the consortium comprised a total of 28 partners from 10 European countries, among them many HiPEAC members. The total budget for the first three-year phase was €80 million. To bring the high-performance, low-power processor to market, a company was created, SiPearl, which has since opened subsidiaries in Germany and Spain in addition to its French headquarters. In a nod to the scale of the project's ambitions, the processor generations have been named after Titans in Greek mythology.

Now that the first phase, SGA1, has come to an end, EPI is celebrating achievements resulting from collaborative working and intense research, which no single partner could have achieved by itself. Notably, these were delivered in an impressively short timeframe, with a relatively modest budget in chip-design terms, and during a global pandemic, with all the associated challenges for working in collaboration.

'I'm proud of the outstanding results achieved by EPI teams after only three years of cooperation, paving the way towards Europe's technological sovereignty,' said Eric Monchalin, chairman of the EPI Board. 'This has created favourable conditions for the launch of the next phase to successfully deliver the European processors and accelerators for the EUPEX (EUropean Pilot for Exascale) and TEP (The European Pilot) projects, the precursors to European exascale systems.'

The following are some of the highlights from the first three years.



General Purpose Processor (GPP)

Led by Atos, with the contribution of SiPearl and other EPI partners, this stream defined architectural specifications for Rhea, the first generation of the EPI GPP implementation, using a co-design methodology. With 29 RISC-V cores, Rhea's Arm Neoverse V1 architecture will offer an effective, scalable and customizable solution for HPC applications.

The EPI GPP achieved register-transfer level (RTL) completion status, and the full Rhea design implementation is currently at the validation stage, using emulations. The Rhea processor will be instrumental in the launch of European exascale computers in 2023.

Rhea will integrate technologies by EPI partners and offers unique features in terms of memory architecture, memory bandwidth optimization, security, and power management.

Memory

To help evaluate architectural choices for the memory controller intellectual property (IP) – a critical factor in performance – CEA developed a complete simulation platform that allows efficient analysis of the memory device interface by decoding and tracking memory commands and data.

Security

Standalone intellectual property (IP) in the form of the Security Management System (SMS) developed by ProvenRun will provide advanced, sovereign security for HPC and edge processors. To further bolster security, the University of Pisa contributed a set of cryptography IPs named Crypto Tile, which were integrated into Rhea by SiPearl. This provides a hardware security module with full security services for high-end symmetric, asymmetric and hashing cryptography, all at substantially increased throughput and decreased energy use in comparison to a software solution.

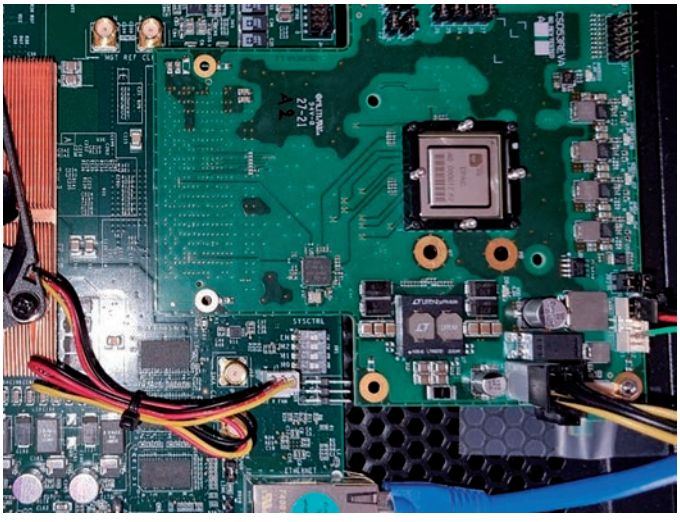
Other Crypto Tile features include secure key storage, secure IP configuration, side-channel attack protection, on-chip true random number generation (TRNG), support of Linux kernel drivers, extreme key lengths for maximum security levels, and high speed en(de)cryption throughput, as well as the basis for post-quantum cryptographic support.

Power

An open-source, RISC-V based power controller designed by the University of Bologna and ETH Zurich was integrated into Rhea, harnessing advanced control and artificial intelligence (AI) algorithms for the power management of large-scale systems-on-chip (SoCs).

In addition, Atos and E4 designed and manufactured the Voltage Regulator reference platform, emulated on a field-programmable gate array (FPGA) board, based on STMicroelectronics technology.

European Processor Accelerator (EPAC)



The EPAC stream set out to deliver energy-efficient acceleration for HPC and AI workloads. With the EPAC test chip proof of concept, EPI has demonstrated that it is possible to create an exclusively European design based on open-source instruction set architectures (ISAs), ensuring freedom from proprietary licences and export restrictions.

This stream fully embraced the open-source philosophy of give and take, contributing to the expansion of the RISC-V ecosystem and adding to the LLVM compiler database. The EPAC systems and FPGA software development vehicles make full use of the Linux operating system and contribute to the community with patches, device drivers, and additional functionality to popular open-source HPC software packages such as OpenMP and MPI. Furthermore, parts of the hardware such as the STX (stencil/

tensor accelerator) were developed using a permissively licensed open-source approach around the PULP platform.

The following suite of technologies was developed as part of this stream:

- The EPAC vector processing unit (VPU), designed by BSC and the University of Zagreb, shows the use of RISC-V long-vector architectures for high-performance computing is a viable approach, delivering high performance on a low energy budget, and that it can be scaled up in future.
- The vector unit is driven by Semidynamics' vector-specialized Avispado RISC-V core and Gazzillion Misses™ technology for energy-efficient processing.
- The dedicated and flexible RISC-V based many-core stencil and tensor accelerator (STX), designed by ETH Zurich and Fraunhofer, leverages stencil processing units to offer exceptional energy efficiency and programmability for machine-learning and stencil workloads.
- Meanwhile, the variable precision accelerator (VRP), designed by CEA, enhances efficiency and reliability for scientific high-performance computing applications such as multiphysics simulations.
- The EPAC test chip also includes multiple distributed banks of shared L2 cache and coherence home nodes (L2HN) designed by FORTH and CHALMERS and optimized for the high-bandwidth requirements of the vector processing units while offering a coherent view of the memory system that facilitates multicore programmability.
- All the processing units and the shared L2HN banks are connected via a high-speed network-on-chip (NoC) in a modular manner that permits the system to scale up. The test chip also includes advanced SERDES technology for very high-bandwidth off-chip and cross-chip communication. Both the NoC and SERDES have been designed by Extoll.
- The PCB (daughter board) to enable the testing of the EPAC test chip were designed and developed by E4.

Automotive

Coordinated by Infineon, the automotive stream paved the way for road-capable autonomous cars thanks to the delivery of an embedded high-performance computer (eHPC) platform and associated software development kit (SDK). This platform, in combination with a downsized, vehicle-tailored, general-purpose processor, meets future cars' increasing demands for computing power in a cost-efficient, economically viable and functionally safe way.

A road-approved BMW X5 demonstrated the proof of concept for the pioneering eHPC microcontroller unit (MCU), integrated into a specially designed flexible modular computing platform, together with several EPI technology IPs. Numerous test drives

were performed to collect data and evaluate test scenarios. Among other features, the platform includes AI-supported integrated cameras and Elektrobit radar imaging analysis software, with integrated preparation for the use of EPI accelerators in the system.

Infineon also expanded the architecture and performance ability of the automotive microcontroller in terms of architecture and performance so that it can act as a master, controlling and surveying a number of aspects. These include safety, security, fall back or redundancy for reduced application, with regard to the top Automotive Safety Integrity Level D (ASIL D) at system level, which is required for autonomous driving applications.

The platform created is scalable and open to further technologies: the modular computing platform has slots for other technologies developed as part of EPI, including:

- future automotive versions of the GPP
- the EPAC RISC-V based accelerator
- the Kalray Massively Parallel Processor Array (MPPA®) accelerator tile, demonstrated for object detection using Kalray's MPPA®-based Coolidge™ processor
- the Menta eFPGA

Tests now reveal that EPI has specific technologies suitable for autonomous driving up to at least level 4 – where the vehicle drives independently most of the time – thus paving the way for the future.

The automotive stream also included the development of a complete software ecosystem, to a large extent based on products

by automotive software specialist Elektrobit. This includes the automotive eHPC platform software stack, including the classic automotive open operating system architecture (AUTOSAR) development for auto eHPC MCUs, and the adaptive AUTOSAR development for HPC GPPs and the L4Re hypervisor (virtualization) that are crucial for automotive applications. With regard to safety, a specific concept was jointly created for a software lockstep.

Common activities

Acting as a provider for other technical streams, this stream delivered a number of results to support the work done elsewhere in the project. First, it established a co-design process to shape the design of European processors. Simulations and models were created to pin down the impact of design decisions on the performance of future applications. A benchmark suite of over 40 applications was used to support co-design and later evaluate the EPI processors. Applications were also prepared to run on future EPI systems, by adapting and testing them on comparable hardware platforms and emulators.

The specification of a 'common platform' architecture was also defined and used as a backbone for architecture exploration, as the starting point for the GPP implementation, and to define guidelines for security and heterogeneous integration.

Another important result was the integration of the power management design in the GPP specifications: power management firmware, off-chip integration consolidating power distribution board design, PLDA interface integration, and consolidation of the power management hardware integration.

Work was also done on multiple aspects of the support of system software development: general and hybrid programming environments, OpenMP and MPI runtimes, OpenMP extra threads support for dynamic load balancing and the introspection-based scheduling mechanism in the LLVM OpenMP runtime, offloading for both GCC and LLVM toolchains, testing of power and energy monitoring libraries on the available reference Arm Platform and the resource manager.

Another notable achievement was the development of three tools – gem5, MUSA and SESAM/VPSim – for a complete multi-level simulation environment that provides relevant virtual prototypes for a wide range of needs encountered in the EPI streams. These tools demonstrated broad capabilities, including detailed chiplet- and NoC-level simulation, system simulation for software design, and performance evaluation for design space exploration and hardware co-design activities.

- A gem5 cycle-accurate, computer architecture simulation package of Rhea, developed by Jülich and FORTH, with models for central processing unit (CPU) cores, memory devices, coherent caches and on-chip networks.
- MUSA, developed by BSC, which allows simulation of different communication neworks, numbers of cores per node, and relevant microarchitectural parameters.
- SESAM/VPSim, developed by CEA, which includes fast on-chip network and cache performance models, as well as decoupling the simulation of functional and extra-functional behaviours.

Future work

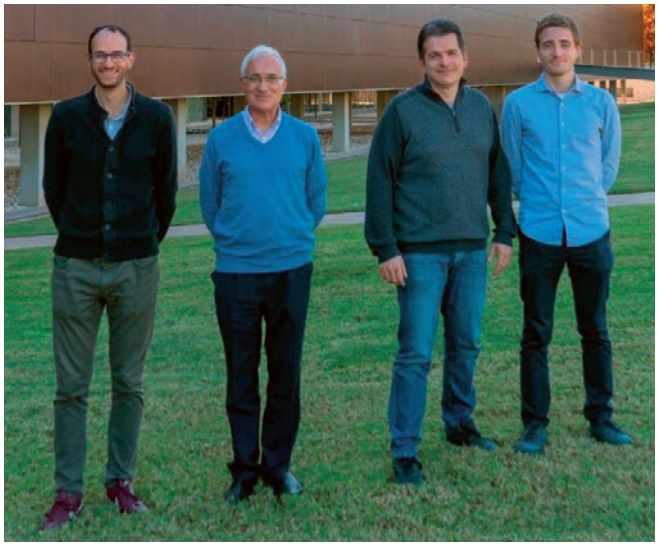
The second phase of the European Processor Initiative, SGA2, was recently launched. Building on the successes of SGA1, this phase will seek to design the second generation of the general purpose processor, named Cronos. HiPEAC members continue to be integral to this research effort, and to furthering Europe's ambitions for technological sovereignty.

The European Processor Initiative has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 826647.

The EPI booth and EPI tutorial at the 2020 HiPEAC conference. HiPEAC members are central to the project's work

RISC-ing it all: How De-RISC is creating the first fully European platform for space

Imagine a market-ready, made-in-Europe hardware-software platform for space and aeronautical applications – free from export restrictions, built on open-source technology, with all the performance of multicore processors and the fault tolerance necessary to withstand the harshest conditions. For the last two years, as reported in HiPEACinfo 59 and 61, the De-RISC consortium has been working to make this a reality. We caught up with the partners behind the research to get the latest on how De-RISC is crossing new frontiers in space technology.



The hypervisor experts: Vicente Nicolau, Paco Gómez, Miguel Masmano, and Antonio García-Vilanova (fentISS, De-RISC coordinator)

Why is De-RISC pioneering?

The De-RISC RISC-V platform responds to the space sector's need for higher performance than that provided by the monocore and basic multicore space-grade processors currently on the market. It will allow access to an increasingly rich software ecosystem as an alternative to the existing software based on SPARC, as well as freedom (or drastic reduction) from export restrictions imposed by commercial instruction set architectures (ISAs). De-RISC will also provide improved support for the design and validation of safety-related real-time applications, and the platform is developed in accordance with space industry development standards.

What makes the XtratuM hypervisor pivotal to the success of De-RISC?

Currently, fentISS' XtratuM is the most popular hypervisor for European space missions. XtratuM guarantees the spatial and temporal isolation required for space safety-related real-time systems, thus allowing the efficient deployment of mixed-criticality applications. XtratuM has already been deployed in 360 satellites of three different missions already in space and it will be used as part of a number of NewSpace and conventional missions to be launched over the next five years.

What is fentISS contributing to the project?

fentISS ported the hypervisor XtratuM to the RISC-V architecture hardware platform of the project, as well as the ARINC-653 compatible LithOS runtime. The supporting tools for developing XtratuM-based systems over the hardware platform have also been developed, and fentISS has led the definition of the software requirements of the platform. Currently, we are implementing support for full virtualization of the Xtratum-Next Generation (XNG) hypervisor in the scope of RISC-V, while next year we will focus on hypervisor validation.

How do you foresee De-RISC technology being used after the project finishes?

The De-RISC ambition is to become European leaders in the supply of an integrated modular avionics (IMA) hardware and software platform, offering users the opportunity to migrate their applications to the RISC-V open-source ISA, being suitable for the avionics domain in addition to the space domain. This technology will be increasingly adopted in space and aircraft systems over the next few years.

The applications dream team: Jérôme Quévremont, Laurent Corbin and Jimmy Le Rhun, Thales Research & Technology

What kind of applications will De-RISC technology be used for?

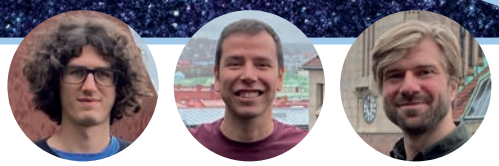
The De-RISC platform is primarily designed to fulfil the requirements of space systems. It features fault-tolerant mechanisms at every level of the platform, integrates dedicated space-grade input / output (I/O) interfaces, and will be radiation tested to representative orbital conditions in the project. However, the high flexibility of a field-programmable gate array (FPGA) implementation and the dependability features of the platform also make it well suited for all kinds of embedded, real-time, safety-critical and mission-critical systems.

What is Thales' contribution to the project?

Thales was responsible for the definition of requirements at the beginning of the project, and performs the end-user validation of the platform. To do so, we are working on several software use-cases, from number-crunching benchmarks to evaluate the raw performance up to a representative satellite payload software. For the latter we partnered with CNES, the French space agency, to use their generic payload flight software framework, LVCUGEN.

Why is De-RISC technology important for Thales?

An important aspect for Thales is the openness of the RISC-V architecture, which allows the implementation of dependability mechanisms not usually found in commercial off-the-shelf (COTS) multicore processors, such as the reduction of timing interference channels, better management of shared resources and advanced monitoring. Independence from export restrictions is another important property, ensured by the exclusively European development of the De-RISC platform.



The hardware pioneers: Stefano Ribes, Fabio Malatesta, and Jan Andersson, Cobham Gaisler

What is Cobham Gaisler's role in the project? Why is De-RISC important for the company?

We provide the hardware platform for De-RISC. The peripherals used in the system-on-chip design are reused from our library of intellectual property (IP) controllers and we apply our NOEL-V RISC-V processor in a RV64GCH configuration in the platform.

We already apply a hardware / software co-design methodology for our developments, but De-RISC has allowed us to include external parties in this process. The project provides a framework for us to work with a software vendor, an end user and an academic partner, and we are therefore able to adapt the hardware building blocks to the needs of project partners. The hypervisor software contribution from fentISS is also important for us and beneficial for the whole RISC-V community.

How would you like to see De-RISC technology being used?

The NOEL-V platform has already been adopted for an on-board computer product by an external company. We hope to see this trend continue and we are also applying the technology developed in our next-generation standard product as part of a European Space Agency project.

The multicore maestros: Jaume Abella, Sergi Alcaide and Guillem Cabo, Barcelona Supercomputing Center

What is BSC's contribution to this project?

Our contribution includes:

- An extended statistics unit, named SafeSU, which provides capabilities to diagnose and control multicore timing interference, and to implement safety measures on top.
- Performance validation stressing tests, mostly targeting multicore interference scenarios.
- Cross-domain analysis of the suitability of the De-RISC platform.

All three items build on previous technology and know-how developed by BSC in the scope of other H2020/FP7 projects and ESA-funded projects, whose target technology readiness level (TRL) was generally lower than that of De-RISC.

Why are multicore processing capabilities important for space applications?

Performance demands are constantly growing in response to the needs of increasing automation and system autonomy. This entails higher performance needs for individual functionalities, as well as a larger number of software functionalities needing to be consolidated onto a limited number of platforms.

Why is interference mitigation particularly important for space applications?

Some safety-critical and mission-critical applications, such as those related to the integrity of the spacecraft (safety-critical) and to the mission instruments (mission-critical) have real-time requirements. However, hardware components inside the processor and in the board cannot be replicated due to costs and reliability concerns. Hence, sharing is mandatory, but the impact of sharing on timing must be properly managed to ensure that safety- and mission-critical applications execute in a timely manner.

Detecting oil spills in real time, thanks to satellite AI

Gianluca Giuffrida, IngeniArs

Recently, campaigns regarding environmental issues have been intensifying, with the aim of raising awareness about the risks of pollution and global warming. IngeniArs, a small / medium enterprise (SME) located in Pisa, Italy, is playing its part in mitigating the negative impact of pollution on the environment through a set of smart applications for the detection of oil spills in real time.

Oil spills represent a major threat to marine ecosystems and migratory birds, killing flora and fauna and making the environment unhealthy for decades. Currently, the identification of these spills is carried out as post-process analysis, slowing intervention. Satellite images are first downloaded by ground stations and then analysed by artificial intelligence (AI) algorithms. The transmission delay added to the processing time worsens this problem.

If a spill is identified quickly, it should be possible to provide at least a partial resolution of the problem. To this end, IngeniArs has developed a deep convolutional neural network (D-CNN) model for on-board satellite detection of oil spills. This model takes synthetic-aperture radar (SAR) images, which provide high resolution images under all weather conditions day and night, as input.

The D-CNN algorithm semantically segments SAR images into multiple classes: 'sea', 'oil spill', 'lookalike (oil spill)', 'ship' and 'land'. The recognition of these categories provides the following benefits:

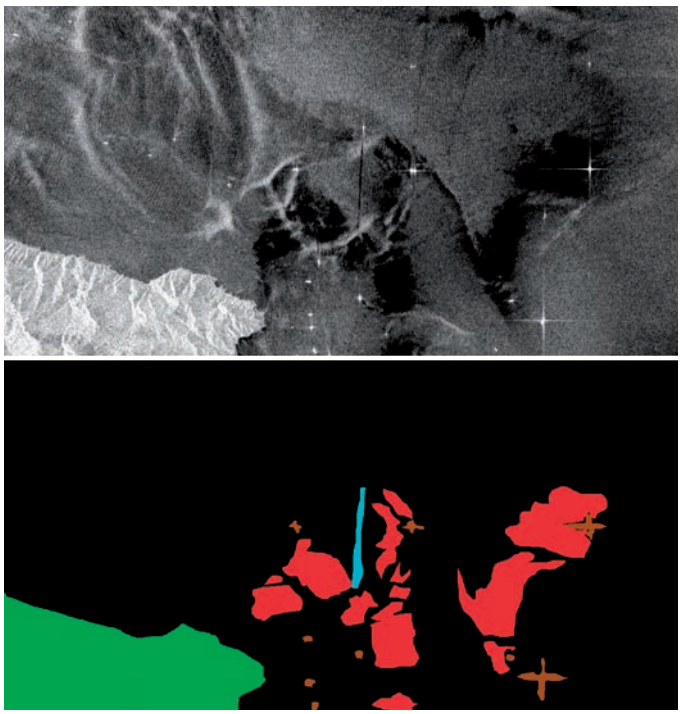
- Reduction in irrelevant data before transmission, reducing the bandwidth used and / or sending only a small map representing the segmentation maps.
- Early notification of important events such as oil spills. Thanks to the five classes, this also allows understanding of the causes of the spills.
- Fast intervention into the spill areas, thanks to swift revisiting time – the interval between two consecutive satellite passages at the same point – of small and nano satellites for Earth observation and the information extracted.

In addition, to enhance the reliability of the network, we tried to reduce false negatives in pixel classification. The 'lookalike' class was added to represent pixels which look like oil, but where the model cannot be sure.

To use the model directly on board the satellite, IngeniArs tailored it directly for commercial off-the-shelf (COTS) embedded hardware platforms, such as the Intel Movidius Myriad 2 or the NVIDIA Jetson Nano, which have been used for micro and nano satellite demonstrators. This hardware requires a dedicated development flow to deal with the strict constraints imposed by the harsh environment, as well as the limited amount of energy available on board.

This is just one example of a project in which IngeniArs is investing, in collaboration with the Information Engineering Department at the University of Pisa, in order to improve environmental monitoring via satellite and using proactive systems to facilitate the green transition.

request@ingeniars.com



How European collaboration powered the first edge AI satellite

'The European Space Agency is really taking Europe back to the top table, with the first edge artificial intelligence (AI) solution deployed in space,' comments David Moloney, chief technical officer at Irish start-up Ubotica. Launched in September 2020, Φ-sat-1 (phi-sat-1), an enhancement of the Federated Satellite Systems (FSSCat) mission, is the first experiment to demonstrate how artificial intelligence can be used for Earth observation.

Data from Φ-sat-1 will play a key role in contributing to Digital Twin Earth, a digital replica of the planet which accurately mimics Earth's behaviour. By visualizing, monitoring and forecasting natural and human activity, Digital Twin Earth aims to provide a real-time healthcheck of the planet and contribute to the objectives of the European Union's Green New Deal.

As described in *HiPEACInfo 62*, in addition to a hyperspectral sensor from Dutch company cosine, the tiny satellite features an AI engine by Ubotica based on the Intel Movidius Myriad 2 system-on-chip (SoC). This technology allows Φ-sat-1to filter out less than perfect images so that only usable data are returned to Earth.

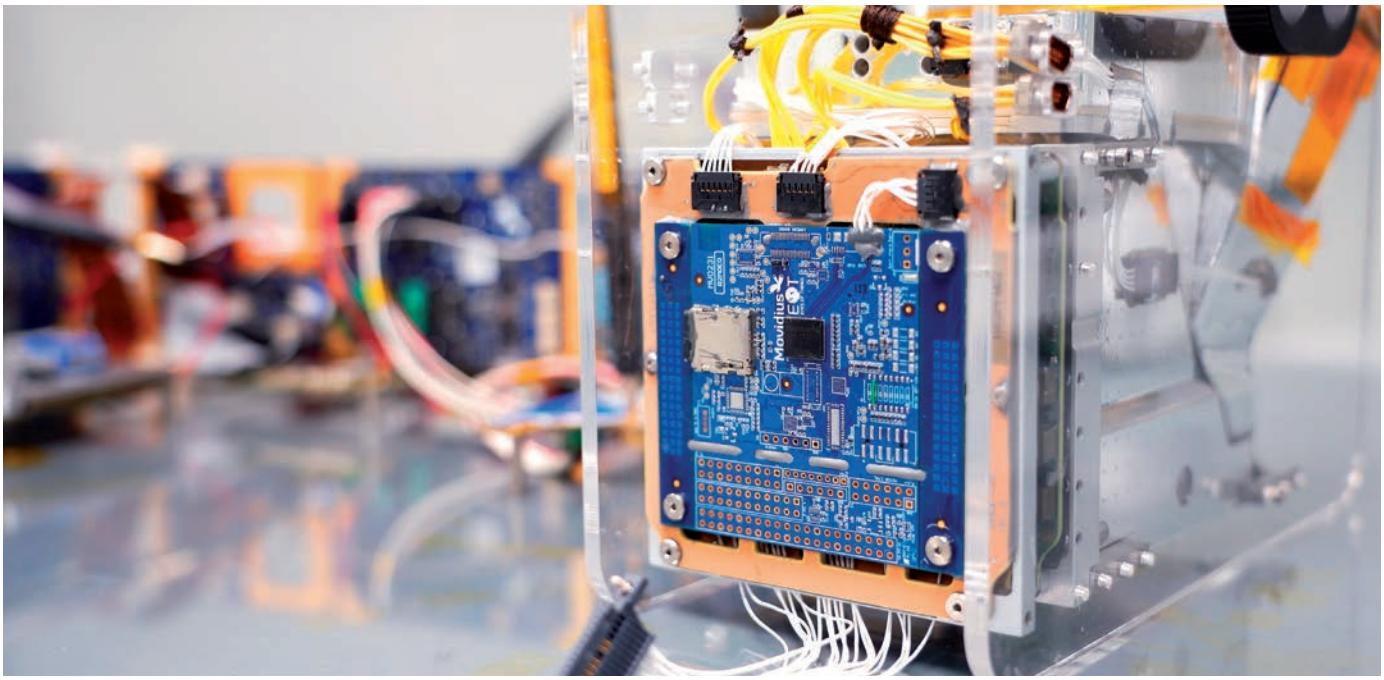
The satellite is a fine example of how European cooperation can lead to world-class technology. In addition to Ireland's Ubotica and the Netherlands' cosine, the satellite was made possible by Tyvak International (Italy), the Universitat Politècnica de Catalunya-Barcelona Tech (Spain), Sinergise (Slovenia) and Deimos Engenharia (Portugal). Moreover, as reported in *HiPEACInfo* 57, the Myriad 2 was irradiated by the H8 beamline of the Super Proton Synchrotron (SPS), a particle accelerator in the CERN complex in France, in order to prepare it for space.

David notes that, from his perspective, the HiPEAC network has been fundamental to the collaborations behind the technology. 'The networking opportunities provided by HiPEAC allowed the Irish start-up Movidius – a company that I co-founded, which was later bought out by Intel – to meet and cement relationships with partners such as the University of Pisa. We participated in several European-funded projects, and also found some great international interns through HiPEAC, some of whom we went on to hire afterwards.' Similarly, Ubotica has been collaborating with European partners and the European Space Agency for some years.

As for the future, work has already begun on the Φ-sat-2 mission, which will further demonstrate the capabilities of AI technology for Earth observation.

Suscribe to #HiPEACTV on YouTube and check out presentations by ESA's Gianluca Furano, Ubotica's Fintan Buckley and David Moloney, plus much more bit.ly/HiPEACTV_subscribe

Check out HiPEAC Jobs for the latest vacancies from EU companies such as Ubotica hipeac.net/jobs



Thank EU for the impact

Deep tech from lab to market, thanks to public European investment

What is the real impact of European Union (EU) funding? HiPEAC is undertaking an impact assessment study to find out. For this article, study lead Ken Scott (University of Edinburgh) spoke to three different companies in the HiPEAC ecosystem, all of whom had benefited in different ways from EU funding.

COMPLEXITY, MANAGED: EMMTRIX TECHNOLOGIES

Founded in 2016 as a spin-off from the Karlsruhe Institute of Technology (KIT), emmtrix Technologies operates in the fields of software parallelization and performance optimization, as reported in *HiPEACinfo* 46. The idea for the company originated in the cutting-edge research carried out during the European FP7 project ALMA, which focused on architecture-oriented parallelization for high-performance embedded multicore systems. Subsequent enhancements were implemented with support from the Horizon 2020 project ARGO (to find out more about both projects, see 'Further information', below).

emmtrix Technologies provides efficient software development tools targeting high-performance embedded systems, including multicore and vector processors as well as accelerators such as graphics processing units (GPUs) and digital signal processors (DSPs). 'As with many spin-off companies, emmtrix started with a talent pool of computer science experts who were highly motivated and comfortable dealing with complexity and confidently communicating complex technical topics,' explains managing director Timo Stripf. 'In the early stages, we focused on channels such as attending conferences and exhibitions and engaging with a network of potential customers.'

However, this approach did not deliver the anticipated results, and the emmtrix team came to realize that, although they were first-class engineers, business development was a specific skillset they needed to incorporate. It also became clear that, given the complexity of the high-performance embedded multicore systems market, they needed to better understand how the market(s) operated and what their potential customers' real requirements were. 'We came to the conclusion that we didn't have a sufficiently clear market fit. Our offer was too complex and we were trying to solve too many problems at once,' says Timo.

Having decided to focus on specific markets (automotive, avionics and industrial automation), the emmtrix team then realized that there were two further hurdles to negotiate. 'First, each sector is likely to have specific software certification requirements. Second, every customer wants a different blend of problems solved – or, conversely, opportunities exploited,' Timo explains.

For anyone else considering taking the step of commercializing an idea and optimizing opportunities provided by EU funding schemes, Timo has the following recommendations: 'Don't underestimate the value of investing time in preparing funding submissions,' he notes. 'Be honest with yourself: it is hard to plan start-ups. Writing a business plan is a particular skill; try to complement your technical expertise with the commercial planning skills of people who have experience in this area.'

The emmtrix team. Human resources are crucial to a start-up's success

Adaptability is key, as well: 'Don't assume all companies in a sector have the same needs and will be swayed by the same marketing strategy – be prepared to adapt and simplify your pitch,' he emphasizes. 'Also, make sure you align with appropriate certification standards before committing to a marketing campaign.'

As for recruitment, Timo suggests dedicating the same effort as you would to any technical development. 'Plan for onboarding and recognize that it will have an impact on productivity, at least in the short term. Treat skills transfer as an investment, not a task to be rushed through.'

"Try to complement your technical expertise with commercial planning skills"

Reflecting on the emmtrix journey to date, Timo stresses the importance of EU funding schemes, which provided a lifeline not only in the early product development phase but also when it came to recruiting staff, once it had become clear which specific skills and experiences were required to perform in such a complex and nebulous market. 'That said, while EU funding has been pivotal in allowing emmtrix to grow, it is essential to prepare for a time when you will need to operate independent of such support,' adds Timo.

FURTHER INFORMATION:

emmtrix Technologies: emmtrix.com

ALMA FP7 project: cordis.europa.eu/project/id/287733

ARGO Horizon 2020 project: cordis.europa.eu/project/id/688131

YOU ARE WHAT YOU EAT: A FRESH APPROACH TO FOOD MONITORING WITH EMBIO DIAGNOSTICS

EMBIO DIAGNOSTICS
Detect Everywhere

When is a tomato not a tomato? You've probably noticed that, despite looking very similar, fresh foods can taste very different – and have a different nutritional impact – depending on how far they have had to travel before reaching your plate. You may also have bought fruit and vegetables that don't go bad for weeks, yet are somewhat tasteless and, again, may have questionable nutritional content. The reason? The use of pesticides and preservatives that facilitate growth and prolong shelf life but have a negative impact on taste and quality.

EMBIO Diagnostics was founded in 2016 by Constantinos Loizou, who had investigated how variables such as pesticides and pathogens affect growth and food quality. 'Our vision was to provide customers with industry-leading monitoring solutions for portable diagnostics and data gathering, utilizing machine learning and artificial intelligence (AI),' notes Constantinos.

With food quality and food waste – with the associated impact on the cost of food and the carbon footprint across the supply chain – becoming increasingly important, the company's launch came at an opportune moment. However, when it came to scaling up, there were two main limiting factors. 'We lacked equipment and skilled people who could not only build our cutting-edge mobile technologies, but also provide the continuous development required to keep innovating and add value to customers' businesses,' explains Constantinos.

The impact of the 2013 economic crisis was still being felt in the Cyprus economy and public funds were in short supply. Fortunately, however, Constantinos swiftly realized that the EU-funded TETRAMAX Innovation Action – which aimed to support product and service innovation for European industries through application-specific digital technologies – was an ideal fit with the ambitions and needs of EMBIO Diagnostics.

After making a well-structured and compelling proposal to TETRAMAX, EMBIO Diagnostics was awarded an initial round of funding that allowed them to further develop digitization and online application development, as well as purchasing essential equipment. 'However, the real value-add was being able to recruit specialist developers and, in turn, build a strong team whose collective talent has developed a suite of products and services that goes far beyond our initial expectations,' says Constantinos.

Given the significant impact of European funding on the growth of EMBIO Diagnostics, what recommendations does Constantinos have for other small / medium enterprises SMEs seeking such support? 'Make sure you invest time in the preparation of the application: don't rush it. All SMEs are limited by lack of resources – and that goes particularly for the founder – but it is essential that careful thought goes into the application,' he cautions. 'It is all too easy to assume that the reader will understand the value of your products and services – they may not. You spend 24/7 on the business opportunity; they may only have a matter of minutes to reach an initial decision.'

While the business case for purchasing assets is important, the need to recruit, train and build a robust, talented team with the potential to bring collective intelligence is an often overlooked requirement, says Constantinos. 'You should be open to declaring that a proportion of any funds awarded will be earmarked for this purpose.'

EMBIO Diagnostics is now an evolving, and self-sustaining operation with international growth ambitions and in 2020 was a finalist in the highly competitive 'Foodtech 500' awards for food safety innovation. Examples of the increasing range and impact of their innovations can be seen on the company website.

FURTHER INFORMATION:
EMBIO Diagnostics website embiodiagnostics.eu

TETRAMAX funding was instrumental in taking EMBIO products, such as the PROMIoTOR system seen here, to the next level

EXAPSYS: SCALING UP WITH THE HIPEAC NETWORK

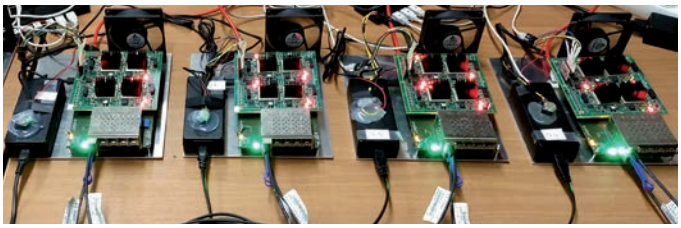
The main business of EXAPSYS (EXAscale Performance SYStems) plc, a spin-off of technologies developed by TSI, FORTH and Synelixis Solutions SA, is high-performance computing (HPC) and cyberphysical systems (CPS) applications. The EXAPSYS core research and development team develops novel highly parallel applications for homogeneous and heterogeneous systems, efficient accelerators based on field-programmable gate arrays (FPGAs), and innovative simulators for highly parallel heterogeneous systems.

'Initially, our products were at the lower end of the technology readiness level (TRL) curve,' explains Yannis Papaefstathiou, chief executive and co-founder of the company, which won a HiPEAC Technology Transfer Award in 2020. 'While they showed promise, it was evident that investment was required to progress to the next stage and allow the company to be commercially viable.'

EXAPSYS were successful in securing EU funding through the EuroHPC Joint Undertaking, as well as from the Greek Government's General Secretariat of Research and Innovation (GSRI), whose support was also a significant factor in the company's development. The funding initially allowed EXAPSYS to further develop their products and in addition allowed collaboration with another SME that, in turn, fortified the product suite.

However, as Yannis points out, in addition to the financial support there was another equally significant enabler, namely the network of contacts facilitated via the HiPEAC network and consequently access to decision makers that otherwise would have been a significant challenge for a spin-off / SME.

The EXAPSYS team in Zoom mode



Parallelizing for performance is EXAPSYS' main business

Yannis highlights that attendance at HiPEAC conferences was a significant benefit in helping catalyse the development of key business relationships. 'The style of HiPEAC conferences is a factor that should not be underplayed; it allows you to build strong relationships with senior executives that would be highly unlikely in most other events of this kind,' he says. 'It turns out that this is perfectly suited to EXAPSYS's business: our innovative, unique technologies require the time and space to raise awareness, attract interest, listen to specific high-tech challenges and eventually build the trust necessary to move to investment.'

It was in such conversations with industrial attendees that it became clear that, while EXAPSYS had tools and modules that were of interest and potential value, most of the target organizations needed something slightly different to what the EXAPSYS core technology was designed to support. This insight prompted the addition of specific new features to the products, thus giving EXAPSYS a competitive edge that would otherwise have been extremely hard to plan for.

'Another benefit from our engagement with HiPEAC is that we have been able to build consortia with complementary companies who can work on different aspects of a shared goal. Collectively, we build increasingly compelling offers for major industrial partners,' explains Yannis. 'Being able to leverage the HiPEAC network to this end is hugely important – just think how much time might otherwise have been wasted trying to do so on your own.'

In summary, Yannis suggests that other companies think beyond the financial aspects of European Union funding models. 'Don't overlook the benefits of being part of an ecosystem that facilitates collaborations, allowing you to nurture strong commercial relationships with major multinational organizations without the frantic style associated with many industry conferences and exhibitions.'

FURTHER INFORMATION:
EXAPSYS website exapsys.eu

The European Union research and innovation programme Horizon Europe started last year. One of the main differences between Horizon Europe and its predecessor, Horizon 2020, is a new emphasis on making sure that the project stays alive after its end, that knowledge and developments within the project are used, and that a positive impact on European society is achieved. This article explores initial lessons and recommendations from early experiences of Horizon Europe.

Getting ready for Horizon Europe

The growing importance of project exploitation

New directions for Horizon Europe

The exploitation of project results in Horizon 2020 was often found to be a tick-box exercise, with little integration from the start and a lack of specialist skills and / or interest. Exploitation take a much greater role in Horizon Europe: it now forms part of the Key Impact Pathways to demonstrate the impact on society. The proposal / reporting templates have been updated with specific references to exploitation – see the impact canvas, for example – and there is an emphasis on continuous reporting, even after the project's end. Third-party exploitation is encouraged, and incentives have been introduced for exploitation.

Some definitions

Within Horizon Europe, the following terms have been specified to aid common understanding:

- Pathway to impact: Logical steps towards the achievement of the expected impacts of the project over time, in particular beyond the duration of a project. A pathway begins with the project's results, to their dissemination, exploitation and communication, contributing to the expected outcomes in the work programme topic, and ultimately to the wider scientific, economic and societal impacts of the work programme destination.
- Results: any tangible or intangible effect of the action, including data, know-how or information, whatever its form or nature, whether or not it can be protected, as well as any rights attached to it, including intellectual property rights.
- Key Exploitable results: the outputs generated during the project which can be used and which create impact, either by the project partners or by other stakeholders. Key results can be reusable and exploitable as they are (e.g. inventions, prototypes, services), or elements (knowledge, technology, processes, networks) that have the potential to contribute towards further work in research and / or innovation.
- Exploitation. The use of results in further research and innovation activities, including:
  - commercial exploitation such as developing, creating, manufacturing and marketing a product or process
  - creating and providing a service
  - standardization and policy-making activities
- Outcomes: The expected effects, over the medium term, of projects supported under a given topic. The results of a project should contribute to these outcomes, supported by the dissemination and exploitation measures. This may include the uptake, diffusion, deployment, and / or use of the project's results by direct target groups. Outcomes generally occur during or shortly after the end of the project.
- Impact: Wider long term effects on society (including the environment), the economy and science. It refers to the specific contribution of the project to the work programme expected impacts described in the destination. Impacts generally occur some time after the end of the project.

Project consortia need to recognize exploitable results and their stakeholders, and identify the value added from their use. Each Horizon Europe beneficiary is expected to use their best efforts to exploit the results they own, or to have them exploited by another legal entity, in particular through the transfer and licensing of results. In this respect, beneficiaries are required to adequately protect their results – if possible and justified – taking into account possible prospects for commercial exploitation and any other legitimate interest.

FURTHER INFORMATION:

horizoneurope.com

bit.ly/Horizon_Europe_webinar_210421

bit.ly/Horizon_Europe_proposal_webinar_210324

bit.ly/HorizonEurope_webinar_Diss_Expl__210609

Accelerators for deep learning are proliferating, but as their complexity grows it is getting harder to accurately model them in design-space exploration. To solve this, Francisco Muñoz-Martinez, Manuel E. Acacio (both University of Murcia), José L. Abellán (Catholic University of Murcia) and Tushar Krishna (Georgia Institute of Technology) teamed up to create a cycle-level simulation tool.

STONNE: A cycle-level microarchitectural simulation framework

The design of specialized architectures for accelerating the inference phase of deep learning (DL) is currently a flourishing area of research. While first-generation systolic-based accelerator proposals used simple fixed dataflows tailored for dense deep neural network (DNNs) applications, more recent architectures such as MAERI or SIGMA make the case for flexibility to efficiently support a wide variety of layer types, dimensions, and sparsity. (See 'Further information', below, for more details on these architectures).

As the complexity of these accelerators grows, the analytical models currently being used for design-space exploration are unable to capture execution-time subtleties, leading to inexact results in many cases. As was the case in other well-established research areas, such as central processing unit (CPU) and graphic processing unit (GPU) design, this creates the need for cycle-level simulation tools to allow for fast and accurate design-space exploration of DL accelerators, and rapid quantification of the efficacy of architectural enhancements during the early stages of a design.

To this end, researchers Francisco Muñoz-Martinez and Manuel E. Acacio from the University of Murcia, José L. Abellán from the Catholic University of Murcia and Tushar Krishna from the Georgia Institute of Technology joined forces to develop STONNE, a cycle-level microarchitectural simulation framework that can plug into any high-level DL framework like PyTorch as an accelerator device and perform full-model evaluation of state-of-the-art systolic and flexible DNN accelerators, both with and without sparsity support. In this way, STONNE can fully simulate real DNN models like Resnets-50, ssd-Mobilenets and BERT from different domains such as computer vision, object segmentation and natural-language processing on state-of-the-art accelerators.

Francisco Muñoz-Martinez at Georgia Institute of Technology

To allow the simulation of different accelerators, STONNE builds on the observation that most current DL accelerator architectures can be logically organized as three configurable network fabrics: distribution network, multiplier network, and reduction networks. This design provides an easily expandable and configurable set of microarchitecture modules that, conveniently selected and combined, can model both systolic and flexible DL accelerators such as the tensor processing unit (TPU), MAERI or SIGMA.

The simulator was published at the prestigious 2021 IEEE International Symposium on Workload Characterization (IISWC21) and is fully available under the terms of the MIT License. Currently, the simulator is gaining attention by the research community and is already being used by a large number of institutions worldwide.

FURTHER INFORMATION:

Access STONNE via github github.com/stonne-simulator/stonne.

Eric Qin et al. SIGMA: 'A Sparse and Irregular GEMM Accelerator with Flexible Interconnects for DNN Training'
2020 IEEE International Symposium on High Performance Computer Architecture (HPCA) ieeexplore.ieee.org/document/9065523

Hyoukjun Kwon et al. 'MAERI: Enabling Flexible Dataflow Mapping over DNN Accelerators via Reconfigurable Interconnects'
ACM SIGPLAN Notices, Volume 53, Issue 2, February 2018 pp 461-475
dl.acm.org/doi/10.1145/3296957.3173176

SMART4ALL-powered technology transfer

EU-funded upgrades to the nautical tourism and apiculture sectors

Continuing our series on the SMART4ALL Innovation Action (see *HiPEACinfo 61* and *HiPEACinfo 64*), in this issue we learn about two 'focused technology transfer experiments', boosting economies in Southeastern Europe. Lasting nine months, these projects each received € 80,000 to transfer knowledge and technologies between the partners involved.

MOORS' LAW: SMARTY INTELLIGENT DOCKING PILLARS

Participants: SaMMY IKE (Greece) and Spark Works Ltd (United Kingdom)
Technology transferred: smart metering and artificial intelligence features for SaMMY IoT platform

The nautical tourism market is one of the most quickly advancing markets, especially in the Mediterranean Sea. However, due to outdated infrastructure, many marinas and ports still operate using traditional methods for booking, yacht hosting and service offerings. Based on legacy installations, these offer restricted interconnection, automation and monitoring capabilities, meaning that opportunities to maximize efficiency are lost.

This is where the SMartY solution, developed in partnership between SaMMY IKE and Spark Works Ltd, comes in. Computer engineer Ioannis Kostopoulos launched SaMMY Yacht in response to the need for smart platforms for the management of yachting marinas, an important part of the tourism market in his native Greece. A step towards fully automated digital smart marinas, SMartY was created by enhancing the SaMMY internet of things (IoT) platform with smart metering and artificial intelligence features provided by Spark Works.

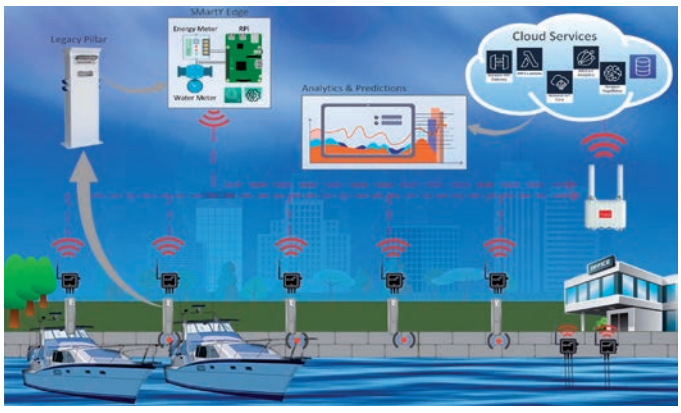
Smart mooring pillars including advanced features such as utility consumption monitoring and billing do exist, but are costly, averaging around €5,000 per pillar. What's more, they rely on the presence of existing connectivity with backbone networks to deliver the full range of services, driving up installation and operational costs.

In contrast, SMartY allows marina administrators to offer innovative services – a key attraction for high-end yachting customers – at a significantly lower cost, less than €1,000 per pillar. Thanks to the integration of low-cost, low-energy cyber-physical systems (CPS), these services allow the increased capabilities that edge data analytics provide to be delivered with minimal interventions to existing docking infrastructure, removing the need to replace pillars.

SMartY transforms mooring pillars into IoT hubs, including smart metering devices that allow users to interact with them via their mobile phones and measure their water and electricity consumption in real time, manage their expenses and save resources. The smart metering devices are integrated with the SMartY Amazon Web Services (AWS) cloud infrastructure through wireless interfaces. This entails minimal alterations to existing installations while providing a range of benefits, including:

- **Tolerance of failures in the backbone network**: the system remains operational regardless of the presence of internet connectivity. Data is stored locally and pushed to AWS when needed.
- **Optimum backbone utilization**: the smart data handling doesn't abuse bandwidth, which benefits the network's health and reduces operational costs. For example, connections are made over 4G or narrowband IoT (NB-IoT).
- **Analytics**: the edge nodes are low-cost devices equipped with all the required processing capabilities to generate analytics close to the data source and hence deliver the required intelligence closer to the field.

With this technological capability, SMartY enables fully automated processes in booking, charging and paying for utility consumption, as well as COVID-safe processes for contactless interactions between marina operators and yachters. In addition, it provides a powerful set of tools for marina administrators, via advanced analytics on utility usage, that allows detection of losses and operational malfunctions and thereby helps marinas become more efficient, eco-friendly and transparent.



BEELINE TO SMARTER APICULTURE WITH THE APIARY PROJECT

Participants: Terra Spatium SA (Greece), Beeing (Italy) and Bianor Services EOOD (Bulgaria)

Results: new features integrated into the miBeez toolbox; development and assessment of blockchain technology in the back end infrastructure of the system and use of historical data for traceability of the honey products; demonstration of the system in partner countries

Valued at approximately €1 billion a year, with the added value from crop pollination estimated at €22 billion a year, apiculture is big business in the European Union (EU). There are around approximately 600,000 beekeepers in the EU, holding a total of around 16 million hives, as well as numerous associated service providers (suppliers, shipping companies and so on). The most valuable beekeeping product is honey, but while around 250,000 tonnes of honey is produced annually in Europe, domestic production only caters to 60% of demand, with the rest provided by imports. Production costs in Europe also tend to be higher relative to markets outside of the continent.

Delivering increased quality and quantity of production at reduced costs, technology-powered precision agriculture could help reinvigorate the European apiculture sector. Apiculture is a particularly good candidate for precision techniques, given the extreme challenges facing it. Bee colony collapse, climate change, the spread of disease and pesticides have dramatically increased bee losses, with a negative impact on both the financial viability of apiculture and environmental sustainability. In addition to pushing up the price of apiculture products, bee losses also lead to ineffective crop pollination.

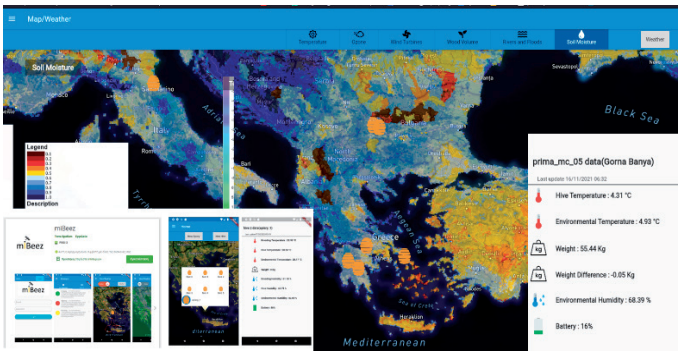
To bolster beekeeping and penetrate the EU apiculture market, Greek company Terra Spatium developed miBeez, a patented internet of things (IoT) apiculture system and service to promote sustainable beekeeping through the application of best practices. It includes a beekeeping management module and beehive monitoring module, using sensor networks and a decision support module that takes internal and external environmental conditions into account to provide guidance and warnings to the producer.

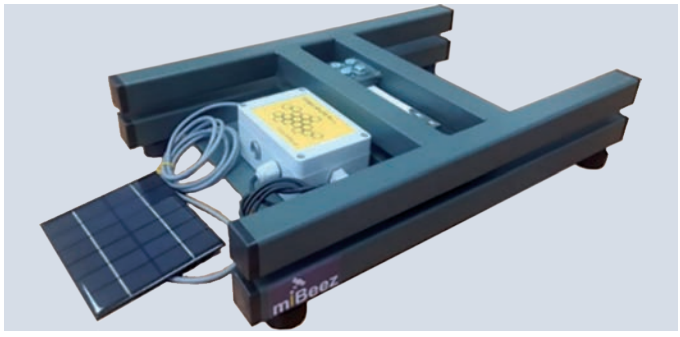
An integrated toolbox, miBeez is designed for use with all kinds of beekeeping activity, helping beekeepers manage production and monitor colonies, while providing input for optimal decision making. This results in improved quality and quantity of apiculture products, as well as reducing time and cost requirements.

The service aims to make beekeeping more efficient and help European producers compete against low-priced imports. As a knock-on effect, it can help create both beekeeping and technology jobs and attract new producers into the sector, thus making rural areas more sustainable. It will also help reverse the current bee population decline, known as colony collapse disorder, and contribute to the vital crop pollination upon which the production of many agricultural products depends.

As part of the APIARY (Advanced PrecIsion ApiculturRe sYstem) project funded by SMART4ALL, Terra Spatium was able to further develop miBeez, adding new features such as space-related layers, as well as investigating the feasibility of integrating blockchain technology in the back end. Thanks to the participation of Bulgarian company Bianor, which contributed to the definition of users and areas and to training users, the project was able to demonstrate the system for a period of three months in real-life conditions in Greece, Bulgaria and Italy.

Following the completion of the pilot phase and improvements to system definition and integration, miBeez is now a market-ready system and service comprising IoT devices accessed by a user-friendly mobile application and desktop portal.





SMART4ALL has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 872614

Parallel universe

How fentISS delivers virtualization solutions for space and other real-time applications

In another example of successful technology transfer within the HiPEAC community, we learn about fentISS, a spin-off of the Universitat Politècnica de València focusing on virtualization solutions.

COMPANY: fentISS (Fent Innovative Software Solutions)
MAIN BUSINESS: virtualization solutions
LOCATION: Valencia, Spain
WEBSITE: fentiss.com

fentISS was born in 2010 as a spin-off of the Universitat Politècnica de València (Spain). The company offers software solutions enabling critical and non-critical applications to share a common hardware platform without interfering with one another, a process known as virtualization.

XtratuM, fentISS' core product, is a bare-metal, space-qualified hypervisor aimed at providing virtualization for safe and efficient embedded real-time systems. This solution guarantees that applications running over the same computer do not conflict with one another in terms of time, memory and other resources. A qualified hypervisor such as XtratuM not only allows the allocation of different criticality levels in the same embedded system, but it also minimizes the recertification cost of already existing applications running as XtratuM partitions. Different guest operating systems can be used by partitions running over this solution, allowing a system to cover a wide range of requirements.

"XtratuM is a bare-metal, space-qualified hypervisor aimed at providing virtualization for safe and efficient embedded real-time systems"

Today, fentISS is a leading company specializing in embedded software development for high-quality products. Thanks to its extensive experience, it has been chosen to collaborate with space agencies like the French National Centre for Space Studies (CNES; French: Centre national d'études spatiales) and the European Space Agency (ESA), as well as in projects under the Horizon 2020 research and innovation programme, including De-RISC, XANDAR and HERMES. The company is also notable for its roster of internationally recognized clients in the space domain, among them Thales Alenia, Airbus Defence and Space, and OneWeb.

fentISS currently has 396 satellites orbiting with the XtratuM hypervisor, the vast majority proceeding from OneWeb, a constellation which will provide global satellite internet broadband services to individual consumers with an expected fleet of 648 satellites. A further nine missions are planned for the following three years, including deep space missions such as JUICE and MMX, which will take fentISS' virtualization solutions to Jupiter and the two moons of Mars.

Photo credit: Alessandroguerriero | Dreamstime.com

Modelling and simulation of COPD patients and clinical staff in emergency departments

With hospitals resources under strain, solutions are urgently needed to help emergency departments treat patients quickly and efficiently. In this article, Mohsen Hallaj-Asghar, Alex Vicente-Villalba, Montserrat Antonin, Alvaro Wong, Dolores Rexachs and Emilio Luque describe how the Computer Architecture and Operating Systems Department and the Escuelas Universitarias Gimbernat Nursing School at the Autonomous University of Barcelona teamed up to create a simulator to train healthcare staff.

"Our proposal consists of implementing a simulator for chronic obstructive pulmonary disease (COPD), which is aimed at helping students and new healthcare staff to take decisions"

In hospitals, the emergency department (ED) is the main route of access for people who need immediate attention from healthcare staff. Recently, the demand for these services from people with chronic illnesses has been increasing. Due to funding deficits, waiting times and integrated attention are suffering. Consequently, this is one of the main problems which public hospitals are facing. The issues caused by a rapidly ageing population are accompanied by a significant increase in chronic illnesses, as well as the number of patients with multiple pathologies.

In terms of chronic conditions, chronic obstructive pulmonary disease (COPD) is one of the main problems facing healthcare systems due to its high prevalence, growing incidence, high mortality, social impact and economic cost.

Given the great diversity and complexity of situations which can arise in emergency departments, specific training is needed for the clinical staff who work there as they have to deal with saturation and be able to make complex decisions rapidly.

Owing to the demand for specific training for healthcare staff, it has been possible to demonstrate that the use of clinical simulations shortens the time needed to learn these necessary skills. However, these simulations involve a high cost as well as large infrastructure, which many healthcare centres are unable to provide. It is here where computer-generated simulations can play a key role in creating virtual worlds resembling reality. This favours the development of training software for use in healthcare environments. ▶

Our proposal consists of implementing a simulator for COPD, which is aimed at helping students and new healthcare staff to take decisions (training), as well as generating knowledge thanks to the evaluation of the simulator's feedback.

The simulator consists of two models:

- Patient model: This component models the evolution of COPD behaviours in relation to inputs from healthcare staff which produce changes in variables related to the patient's state (example, satO2, breathing regularity, etc), offering a new state (output).
- Clinical staff model: This component models how healthcare staff take decisions, that is to say employing clinical reasoning by using inputs, which in this case is information on the patient's state and some outputs. The latter refers to the actions taken by the healthcare staff, such as the treatment.

Both models converge in simulations for chronic obstructive pulmonary disease patients. This means that both models execute the same simulation parameters in order to be able to make a comparison between the two, as shown in the figure. As previously mentioned, the purpose of the simulator is to train healthcare staff, as well as to generate knowledge by comparing the simulation of new staff with the simulation carried out by the simulator. By providing such knowledge feedback for staff, the simulator should help contribute to the positive evolution of the disease.

FURTHER READING:

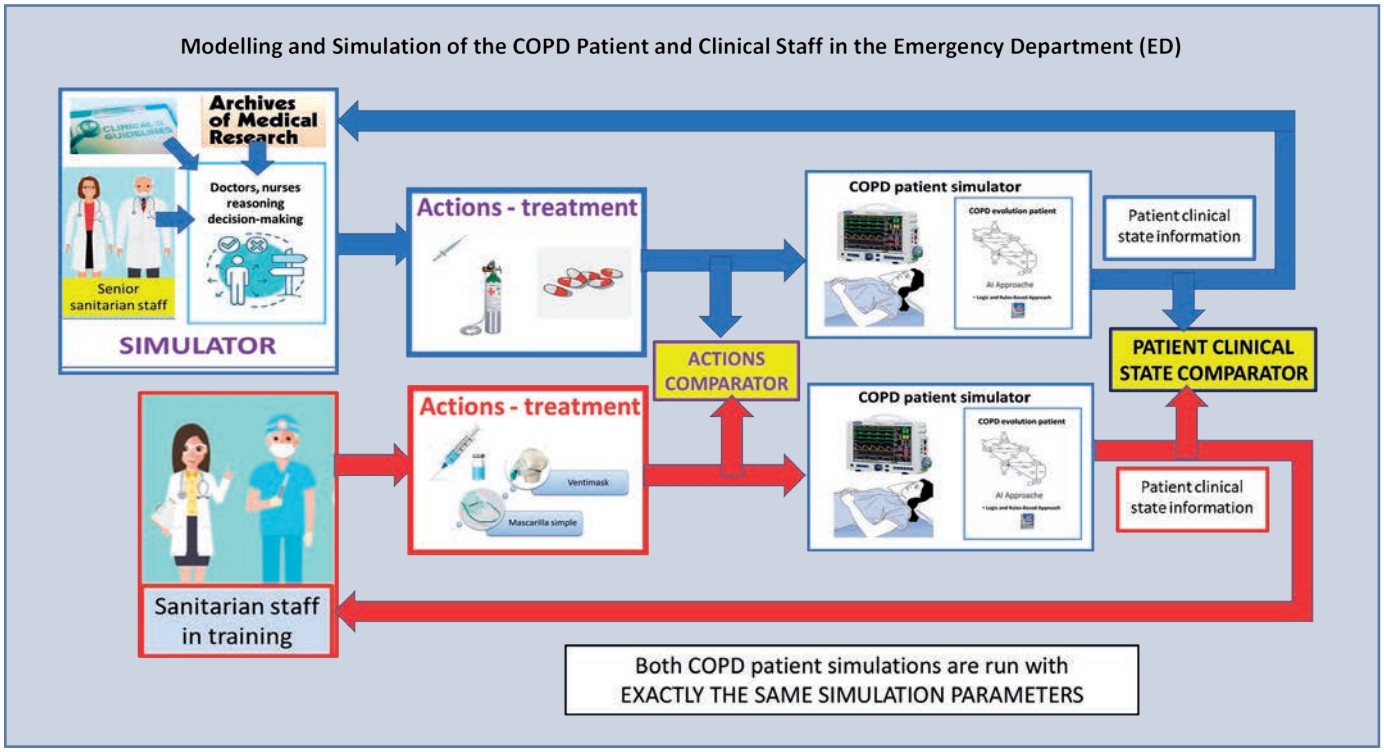
Almoaber, B; Amyot, D. 'A Review on the Contribution of Emergency Department Simulation Studies in Reducing Wait Time'. International Journal of E-Health and Medical Communications (IJEHMC), 2017, vol. 8, no 3, p. 1-21
bit.ly/ED_simulation-studies_review

Picton G, Braniff C, Bossy M. 'Transferring skills from simulation to clinical practice: an in- situ report'. BMJ Simulation and Technology Enhanced Learning 2020; 6:186-187
bit.ly/BMJ_transferring_skills_simulation

Gul, M; Guneri, A F. 'A comprehensive review of emergency department simulation applications for normal and disaster conditions'. Computers & Industrial Engineering, 2015, vol. 83, p. 327-344
bit.ly/Review_ED_simulations

Villalba, A V, Antonin M; Rexachs, D; Luque, E. 'Computer simulation as a methodology for theoretical learning of clinical skills in nursing'. International Journal of Integrated Care (IJIC). 2018 Supplement2, Vol. 18, p1-2. 2p.
bit.ly/Simulation_clinical_skills

A. V. Villalba, M. Antonin, D. Rexachs, E. Luque. 2019. 'A Reactive "In Silico" Simulation For Theoretical learning Clinical Skills And Decision-Making'. The eleventh international conference on advance in system simulation. pp.3-7. SIMUL 2019



COPD patient and staff simulation / modelling

Welcome to the latest in our series on disruptive European-funded research. In this issue, we find out about three-dimensional (3D) stacked hardware layers for machine translation, discover how artificial intelligence (AI) is making cyber-physical systems (CPS) more efficient, and get up to date on the latest technologies powering data processing and analytics.

Innovation Europe

THE FVLLMONTI: TAKING TRANSISTORS TO THE NEXT DIMENSION, WITH CRISTELL MANEUX

FVLLMONTI is a FET ProActive project that seeks to provide an innovative solution for applications such as machine translation without relying on sending information to the cloud. Specifically, the project is creating ferroelectric vertical nanowire field-effect transistors (VNWFETs) to produce 3D stacked hardware layers of neural networks. We caught up with FVLLMONTI Coordinator Cristell Maneux to find out more.

How is FVLLMONTI disrupting nanoelectronics?

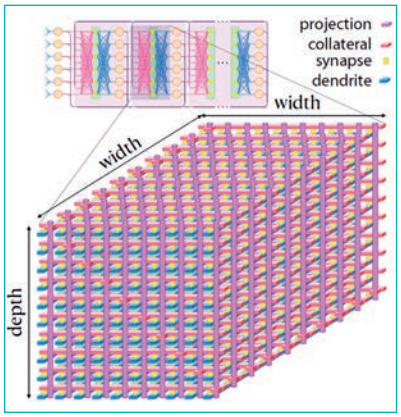
Currently, in the field of nanoelectronics, the method for producing computer processing units is to cut a wafer to create a die. Wafers are very thin, and the transistors on the hardware are actually two dimensional (2D), although the interconnects are 3D. In this project, we intend to fabricate and simulate model transistors that are really, naturally, 3D. This is important because it gives us access to a denser processing unit.

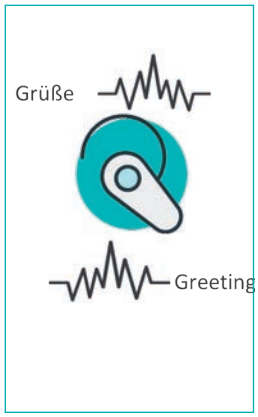
Why is it important to get more density on processing units?

More density allows us to have more processing power, which is a requirement for efficient neural networks. By creating 3D transistors, using nanowires, we have access to naturally 3D crossbars, which fit nicely to the neural network architecture, delivering superior performance for machine learning applications. Another thing which is really mandatory is to have memory port points embedded into this technology, and to do this we intend to use ferroelectric gate transistors.

What's the main application of this technology?

The one we're targeting is translation, but basically speaking it could be used in any application that needs to stay on the device, so we don't have to access the internet, for example, and so can save energy. This would mean that you could have one person speaking in French, for example, and the other person speaking in English, and a lightweight in-ear device could be used to translate in real time. This would avoid the somewhat ridiculous situation we have at the moment, where you have to send data to a translation machine in the cloud and retrieve it again, all while two people are standing in front of one another.





FURTHER INFORMATION:
fvllmonti.eu

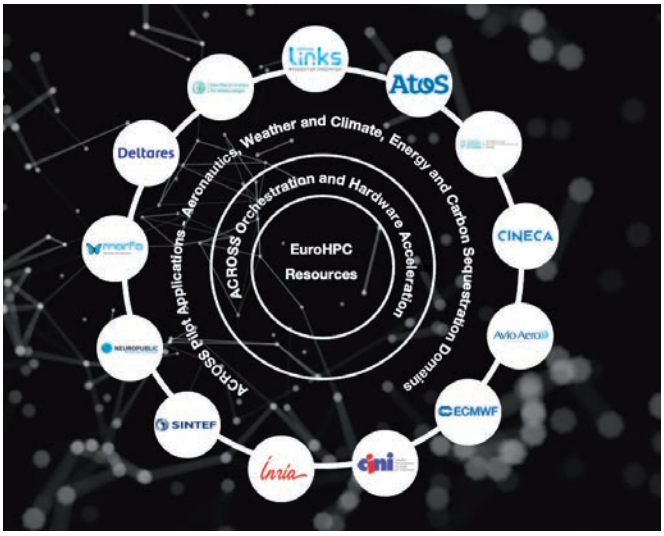
FVLLMONTI (Ferroelectric Vertical Low energy Low latency low volume Modules fOr Neural network Transformers In 3D) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101016776

HARNESSING ACCELERATION FOR HIGHER PERFORMANCE DATA ANALYTICS AND SIMULATIONS WITH ACROSS

Artificial intelligence and big data are poised to revolutionize the high-performance computing (HPC) domain, thanks to the new tools they offer for analysing the immense amount of data generated by scientific experiments and modern simulation tools. The EuroHPC project ACROSS aims to develop an innovative execution platform to serve emerging workflows that combine large and complex numerical simulations with high-performance data analytics (HPDA), machine learning and deep learning techniques. This platform will be co-designed using a large set of advanced hardware and software tecnologies, with a view to delivering near-exascale performance.

As such, hardware acceleration technologies will be of primary interest for ACROSS in order to deliver maximum performance with minimal energy consumption. Specifically, in addition to established graphics processing units (GPUs), ACROSS will also integrate and exploit field-programmable gate arrays (FPGAs), neural network processors (NNPs) and neuromorphic hardware. Looking forward, ACROSS also envisions the adoption of future outcomes of the European Processor Initiative (EPI), thereby simultaneously promoting first-class European technology.

Energy efficiency, efficient usage of computing resources and application performance will be achieved through the design and implementation of an innovative workflow and resource management system. A multi-level approach will be used to define the workflows, manage the resource allocation beyond traditional queuing systems using a machine learning based approach, and efficiently execute workload tasks on the allocated heterogeneous resources. Pilots have been selected to demonstrate the capabilities of the ACROSS platform in three relevant industrial and scientific domains: (1) aeronautics; (2) weather, climate, hydrological and farming; and (3) energy and carbon sequestration.



The ACROSS project will be tested via pilots in aeronautics, climate and energy

The ACROSS platform will be backed by supercomputing and cloud resources provided by European HPC centers (CINECA and IT4Innovations), including a new generation of supercomputers.

PROJECT NAME: ACROSS: HPC Big Data ArtifiCial intelligence cross Stack PlatfoRm TOwardS ExaScale
START/END DATE: 01/03/2021 – 29/02/2024
EU Call: H2020-JTI-EuroHPC-2019-1
KEY THEMES: high-performance computing (HPC), exascale, artificial intelligence (AI), high-performance data analytics (HDPA), accelerators (GPUs, FPGAs, neuromorphic), HPC job scheduling and workflows management
PROJECT COORDINATOR: Olivier Terzo, Fondazione LINKS (Italy)
olivier.terzo@linksfoundation.com
DISSEMINATION MANAGER: Alberto Scionti, Fondazione LINKS (Italy)
alberto.scionti@linksfoundation.com
PARTNERS: Italy: Fondazione LINKS (coordinator), CINECA, GE Avio Aero, Consorzio Interuniversitario Nazionale per l'Informatica (CINI), Morfo Design SRL; France: Atos (Bull SAS), Institut National de Recherche en Informatique et en Automatique (INRIA); Czech Republic: IT4Innovations, VSB – Technical University of Ostrava; UK: European Centre For Medium-Range Weather Forecasts (ECMWF); Norway: SINTEF AS; Greece: Neuropublic AE Pliroforikis & Epikoinonion; Netherlands: Stichting DELTARES; Germany: Max-Planck-Gesellschaft Zur Forderung Der Wissenschaften EV
BUDGET: €8.8 million (€4 million EU contribution)
acrossproject.eu
@across_project
Across Project
ACROSSProjectEU

The ACROSS project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement no. 955648. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Italy, France, the Czech Republic, the United Kingdom, Greece, the Netherlands, Germany and Norway.

AI-AUGMENTED AUTOMATION FOR MORE EFFICIENT CPS, THANKS TO AIDOART

Cyber-physical systems (CPS) are growing increasingly complex. This poses challenges not only in software development and analysis but also for the use and maintenance of CPS once they are deployed. By automating aspects of CPS development and use, artificial intelligence (AI) can help manage the complexity and enhance the efficiency of CPS.

As Gartner reported in 2019, by 2023 40% of infrastructure and operations teams will use AI-augmented automation in large enterprises, resulting in higher information technology (IT) productivity. Supporting continuous development in complex systems, AI-augmented automation promises considerable, long-term added value in ultra-large system development.

Hence the focus of AIDOaRT, a three-year Horzion 2020-ECSEL project involving 32 organizations from seven different countries, grouped into clusters. The project's overarching goal is to support requirement definition, monitoring, modelling, coding, and testing as part of continuous system engineering (CSE) in CPS and CPS systems of systems (CPSSoS) via AI-augmented automation. It aims to enhance the engineering process using AI-augmented methods, integrating DevOps and model-driven engineering principles, to observe and analyse data collected from both runtime and design-time artefacts in rapid CSE cycles.

The project's objectives are as follows:

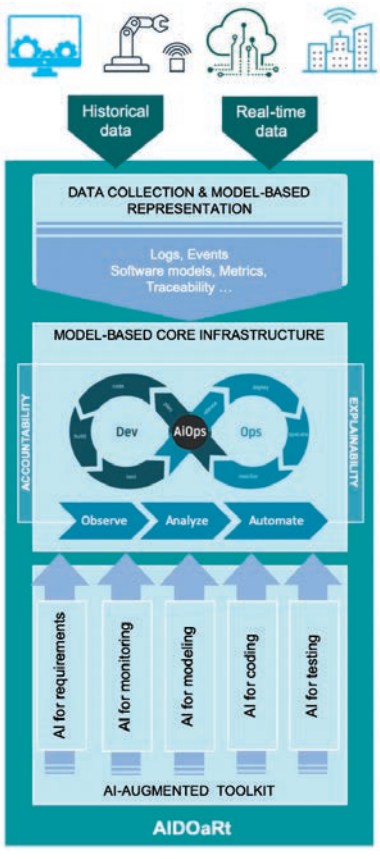
- Provide a model-based framework to support the CPS development process by introducing AI-augmented automation.
- Enhance the DevOps toolchain by using AI and machine learning (ML) techniques in multiple aspects of the system development process, including modelling, coding, testing and monitoring.
- Support the collection, representation and traceability of runtime data and software models; assist in the analysis of historical and real-time data in combination with design information; and support the automation of tasks in the DevOps pipeline in accordance with results from previous analyses.

The consortium expects industrial uptake of AIDOaRT technologies for the development of complex systems that will scale up to real systems' demands, and that will thereby benefit a number of critical applications.

PROJECT NAME: AIDOaRT: AI-augmented automation for efficient DevOps, a model-based framework for continuous development At RunTime in cyber-physical systems
START/END DATE: 01/04/2021 – 31/03/2024
EU Call: H2020-EU.2.1.1.7. – ECSEL, H2020-EU.2.1.1
KEY THEMES: artificial intelligence-augmented automation, modelling, coding, testing, monitoring, continuous development, cyber-physical systems (CPS)
PARTNERS: Sweden: Mälardalen University (coordinator), Bombardier Transportation Sweden AB, Volvo Construction Equipment AB, RISE Research Institutes of Sweden AB, Westermo Network Technologies AB; Austria: AIT Austrian Institute of Technology GmbH, Automated Software Testing GmbH, AVL List GmbH, Dynatrace Austria GmbH, Graz University of Technology, University of Linz; Czech Republic: CAMEA spol. s.r.o, Brno University of Technology, Finland: Åbo Akademi University, Anders Innovations OY, Qentinel Quality Intelligence OY; France: Clearsy SAS, Institut Mines-Télécom (IMT), Prevision.io, Softeam; Italy: Abinsula srl, Intecs Solutions spa, Ro Technology srl, Tekne srl, University of L'Aquila, University of Sassari; Spain: ACORDE Technologies SA, Foundation for the Open University of Catalonia, HI Iberia Ingeniería y Proyectos SL, Instituto Tecnologico de Informatica, Prodevelop SL, University of Cantabria
BUDGET: € 22,968,348, with requested funding of € 6,898,684

aidoart.eu
@aidoart
AIDOaRt

AIDOaRT has received funding from the European Union's Horizon 2020 Electronic Component Systems for European Leadership Joint Undertaking (ECSEL-JU) research and innovation programme under grant agreement no. 01007350.



LSST image credit: Rubin Observatory/NSF/AURA Cell/DNA/Chromosome image credit: Free Exam Academy

PLATON TO DELIVER SCALABILITY AND PERFORMANCE FOR LARGE-SCALE DATA SERIES PROCESSING

Funded by the European Commission as a Marie Skłodowska-Curie Individual Fellowship, and lasting a year, PLATON, led by Professor Panagiota Fatourou, focuses on solutions for large-scale data series processing.

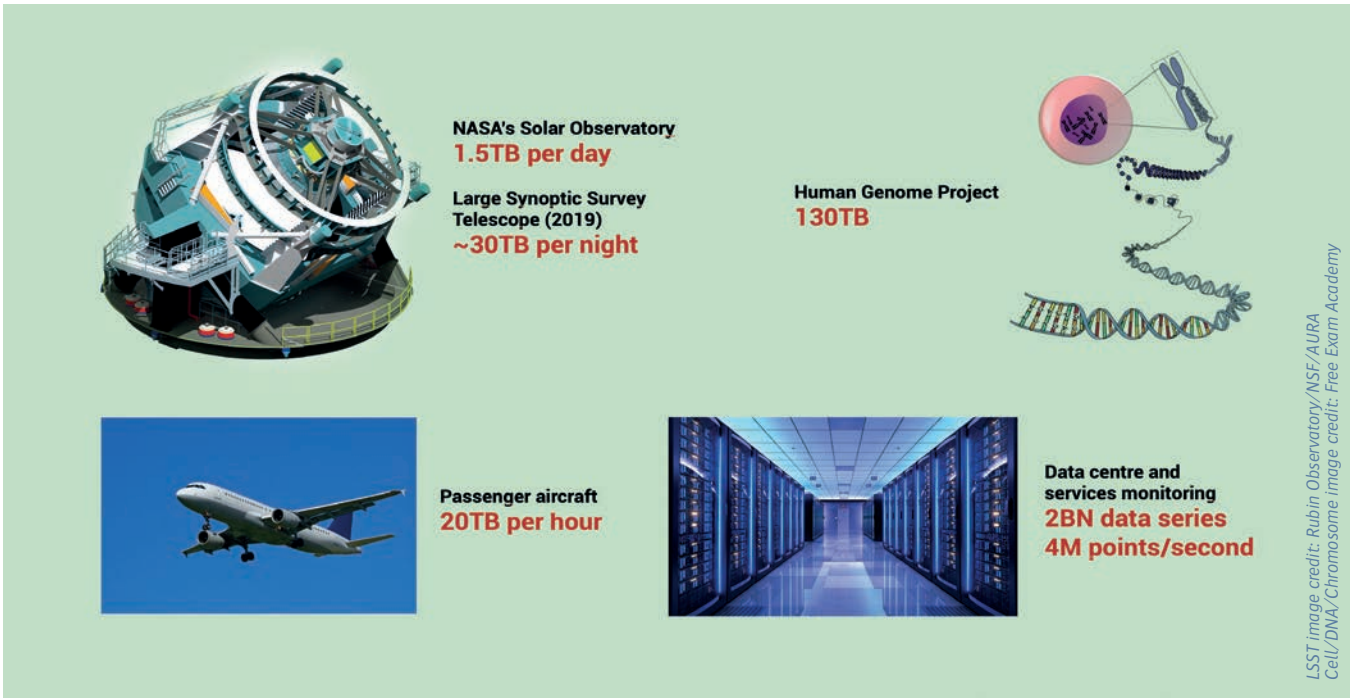
Today, processing large collections of real-world data series is a major challenge for a wide range of application domains, including finance, seismology and other earth sciences, astrophysics, neuroscience, engineering, and so on. Due to recent advances in the development of modern scientific instruments in earth sciences and the growing importance of the internet of things (IoT), data series collections are experiencing an unprecedented growth in size. One of the most pressing issues in data series processing, therefore, is achieving enhanced performance and high scalability.

PLATON aims to deliver solutions for large-scale data series processing that meet such performance and scalability goals. Specifically, PLATON aspires to build, for the first time, the necessary methods, algorithms and tools for highly efficient, scalable, and fault-tolerant processing of huge collections of data series. Scalability will be achieved by exploiting the full computational capacity (multiple nodes, multiple cores, accelerators) of modern computing platforms. Meanwhile, PLATON will demonstrate its value proposition using real datasets from different domains.

The proposed research project has the potential for significant economic and social impact in Europe, given that multiple scientific and industrial fields are currently in need of the right tools to handle their massive collections of data series – a fact acknowledged by the European data strategy.

PROJECT NAME: PLATON: Platform-aware LArge-scale Time-Series prOcessiNg
START/END DATE: 01/10/2021 – 30/09/2022
EU Call: H2020-EU.1.3. | H2020-EU.1.3.2.
KEY THEMES: data series processing
Coordinated by: Université de Paris, France
CONTACT INFORMATION:
Panagiota Fatourou, LIPADE - University of Paris
faturu@csd.uoc.gr
Themis Palpanas, LIPADE - University of Paris
themis@mi.parisdescartes.fr
PLATON factsheet on CORDIS
bit.ly/PLATON-website
@PlatonProject
PLATON Project - Marie Skłodowska Curie Actions
Platon Project - Marie Sklodowska-Curie Action

This project receives funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 101031688.



Examples of modern applications' data greed

DAPHNE: INTEGRATED DATA ANALYSIS PIPELINES FOR LARGE-SCALE DATA MANAGEMENT, HPC, AND MACHINE LEARNING

Modern data-driven applications deal with increasingly large, multimodal datasets, a variety of machine learning (ML) models, simulation and optimization. This growing complexity has led to a trend towards integrated data analysis (IDA) pipelines that combine systems for query processing and data management (DM), high-performance computing (HPC), and ML training and scoring. Examples include ML-assisted manufacturing, biomedical engineering, natural sciences, remote sensing, transportation, recycling, healthcare, and finance.

The DAPHNE project aims to build an open and extensible system infrastructure for such IDA pipelines. Interestingly, DM, HPC, and ML systems share many compilation and runtime techniques, and the underlying hardware infrastructure converges as well. Yet, the programming paradigms, resource management, data representations, and execution plans differ substantially. Dedicated system infrastructure has the potential to improve the productivity of developing and deploying IDA pipelines, while effectively utilizing heterogeneous hardware and eliminating unnecessary overheads.

Besides existing benchmarks, DAPHNE's real-world use cases include earth observation for classifying local climate zones; success rate prediction of ion beam tuning in semiconductor manufacturing; testing, simulation, and estimation of material degradation of semiconductor devices; and gas ejector optimization for fuel cells. This diverse set of real-world use cases will be generalized to a benchmark for IDA pipelines.

DAPHNE system infrastructure

Many ML algorithms, query processing, and numerical computation can be expressed with frame and matrix (tensor) operations. DAPHNE provides a domain-specific language (DSL) called DaphneDSL, with the design principle of abstract data types for data independence (data representations, device placement, and value types), while also allowing extensibility. A Python application programming interface (API) called DaphneLib, embedded SQL query processing, and a hierarchy of DSL-based library functions ensure seamless integration in typical workflows and environments.

DAPHNE utilizes MLIR as a multi-level intermediate representation and compiler library. DaphneDSL scripts are parsed into DaphneIR, an MLIR dialect with additional optimization passes such as shape inference, simplification rewrites, and lowering to calls of concrete kernel implementations for different data representations (e.g. dense, sparse, compressed), value types (e.g. FP32, UINT8) and devices (e.g. central processing units (CPUs), graphics processing units (GPUs), field-programmable gate arrays (FPGA), and computational storage).

DAPHNE use cases include earth observation

Distributed or on-device data is accessible from metadata of passed variables and managed via distribution primitives. A tiled (vectorized) execution engine allows processing fused operator pipelines on splits of input frames or matrices, task-based multi-device scheduling, and device-specific code generation. For extensibility, we plan to provide an extension catalogue for new data types, kernels, I/O primitives, and scheduling algorithms, as well as script-level, sideways entry into the multi-level compilation chain.

PROJECT NAME: DAPHNE: Integrated Data Analysis Pipelines for Large-Scale Data Management, HPC, and Machine Learning
START/END DATE: 01/12/2020 – 30/11/2024
EU CALL: H2020-ICT-2020-1
KEY THEMES: machine learning systems, large-scale data management, high performance computing, artificial intelligence, optimizing compiler, extensibility, heterogeneous hardware
PARTNERS: Austria: Know-Center GmbH (coordinator), AVL List GmbH, Infineon Technologies Austria AG, Kompetenzzentrum Automobil- und Industrieelektronik GmbH; Germany: Deutsches Zentrum fuer Luft- und Raumfahrt e.V., Hasso-Plattner-Institut for Digital Engineering gGmbH, Technische Universität Dresden; Switzerland: Eidgenoessische Technische Hochschule Zuerich; Universitaet Basel; Greece: Institute of Communication and Computer Systems; Poland: Intel Technology Poland sp. z o.o.; Denmark: IT-Universitetet i København; Slovenia: Univerza v Mariboru
BUDGET: €6,609,665
daphne-eu.eu
@daphne_eu

FURTHER READING:

Patrick Damme, et al. 'DAPHNE: An Open and Extensible System Infrastructure for Integrated Data Analysis Pipelines', CIDR 2022
bit.ly/CIDR_DAPHNE_paper

DAPHNE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 957407.

RISC-V is having something of a moment. As popularity grows for the open-source instruction-set architecture (ISA), Codasip Chief Marketing Officer Rupert Baines, a member of the HiPEAC Industrial Board, sets out why it is the perfect time to be involved in RISC-V research.

Why it's time to join the RISC-V revolution

For some time now, I have been arguing that RISC-V offers us a tremendous platform for innovation and collaboration; indeed, it has the potential to change the business model of the entire semiconductor industry. Over the last few years, industry momentum in RISC-V has continued to grow, unabated… in fact, with more impetus. Why is this?

As highlighted by Codasip Chief Executive Officer Ron Black at the 2021 RISC-V Summit, and as well documented by the HiPEAC community, scaling is failing. After 50 years of driving semiconductor economics, the underlying semiconductor laws (Moore's, Dennard's, Amdahl's) are failing, while advanced manufacturing is becoming prohibitively expensive. This has prompted a near crisis in the industry. While new paradigms such as optical computing or new materials such as carbon nanotubes represent intriguing opportunities for future development, in the short term the best bet for increased performance is customizing the hardware to the workload.

Since its inception, RISC-V has represented a fundamental shift in processor architecture. Its flexibility and modularity open up new possibilities, allowing system-on-chip designers to tune not only the microarchitecture but the ISA itself, in order to co-develop hardware and software and provide truly optimized performance, power and area. There's also no denying that its open source nature makes it very attractive to anyone wishing to build something new without having to foot the bill for licensing fees. Thanks to these advantages, RISC-V is powering a wave of novel processors and domain-specific accelerators capable of delivering the performance gains demanded by modern applications.

> ***"Over the last few years, industry momentum in RISC-V has continued to grow, unabated"***

One of the first companies to commercialize RISC-V intellectual property (IP), Codasip has long been using RISC-V to offer customers the opportunity to customize and differentiate designs, through a hardware-software co-design process tailored to individual needs, in addition to ready-made designs. We cover everything from domain-specific acceleration to creating new processors from scratch. So far, we've enabled a whopping two billion RISC-V cores with our Codasip Studio design automation tool and CodAL processor description language, while helping customers use architecture licences, customization, and domain specific compute.

This is a timely opportunity for RISC-V, and growth opportunities are calling. For computer architects, such as those in the HiPEAC community, it's a hugely exciting moment, and at Codasip we are thrilled to be at the centre. We're proudly European, and very keen to cooperate on European procedures or initiatives, whether commercial, grants related or universities. And if you're looking for a fulfilling career in RISC-V, we're hiring!

The Codasip team at the RISC-V Summit 2021

FURTHER INFORMATION:

Keynote talk: 'Scaling is Failing' by Ron Black, CEO, Codasip

RISC-V Summit 2021 bit.ly/Codasip_RB_RISC-V21

Codasip website codasip.com

Subscribe to #HiPEACTV on YouTube for exclusive videos on RISC-V featuring internationally recognized champions of the ISA such as Calista Redmond and Luca Benini.

youtube.com/c/HiPEAC

Finishing your PhD or thinking about changing careers in 2022? Having trouble deciding whether industry, academia or setting up your own company would be the best fit? To help you decide, HiPEAC Jobs invited senior figures from each sector to share their experiences, as the HiPEAC Jobs team explains.

Large multinational, academia or your own company?

Tips on planning your next career move with HiPEAC Jobs

On 26 October 2021, HiPEAC Jobs hosted an in-person 'Inspiring Futures' careers session at Computing Systems Week Lyon. During the session, we heard from Albert Cohen (Google), Alberto Bosio (École Centrale de Lyon) and Babis Chalios (Nubificus), who shared their perspectives on careers with an audience of local and international students. The discussions centred around the challenges in and differences between careers in academia, major international companies and small / medium enterprises. Below are the main findings from the session.

The industry asset: Albert Cohen, Google

Albert is a longstanding HiPEAC member who spent a number of years at a research centre (Inria) before transferring to Google. Joining industry later in his career, after his research had attracted the attention of major companies, gave him the freedom to undertake more research projects on his own terms, he explains.

Albert is keen to stress that academia and industry can happily co-exist in the same career. 'You can keep doing research if you work in industry, whereas you can also collaborate with industry while working in academia,' he points out. 'Deep tech can be done in both, and you can also switch between the two; you just have to know what the conditions are and leverage what you learn from each.'

> *"You can keep doing research if you work in industry, whereas you can also collaborate with industry while working in academia"*

Open source has been pivotal to Albert's career, and it has opened up a wealth of opportunities. It was years of work contributing to open-source projects such as LLVM that drew the attention of big tech companies, who then headhunted him. 'If what you do is good, it will attract the attention of industry,' he notes, citing his own experience in fiendishly complicated compilers, which solve important industry pain points, as an example.

Whether you choose academia or industry, Albert recommends doing a PhD first, as this opens the door to more impactful work even in an industry setting.

The mobile professor: Alberto Bosio, École Centrale de Lyon / Institut des Nanotechnologies de Lyon

A researcher who has so far dedicated his career to academia, Alberto Bosio has hopped around research centres in Italy and France. He stresses the importance of mobility in academic careers, both during and after your PhD. 'Every time I've moved, I've had the opportunity to collaborate with new colleagues. That means new ideas and new ways to tackle problems, in both directions,' he says.

Another piece of advice is not to stick in your research field. 'Try to avoid staying in the same topic; try to be as open as you can,' he says. 'Conferences such as HiPEAC are the best way to really make your network and to become aware of new issues.

▶

You can also have discussions with your peers or with senior people in the field.' As for funding, he points out that projects are the major source of income for research, both publicly funded projects and projects with private companies.

> *"Every time I've moved, I've had the opportunity to collaborate with new colleagues. That means new ideas and new ways to tackle problems"*

The agile entrepreneur: Babis Chalios, Nubificus

Originally from the Greek island of Lesbos, Babis' career has taken him all over Europe in a variety of work settings: having graduated from the National Technical University of Athens, he did his PhD at Queen's University Belfast before undertaking post-doctoral work at Barcelona Supercomputing Center. His next move was to Volta, a small company in Barcelona, where he appreciated the nine-to-five culture, but found that he missed the research aspect.

Babis got together with a group of friends he'd met in Athens, who had kept in touch mainly thanks to HiPEAC events such as the ACACES summer school. In their free time, they started working on research ideas, which led to the foundation of Nubificus, a young company distributed over the UK, Spain and Greece and which successfully participated in bids for European funding.

Throughout his career, Babis has been driven by the desire to keep learning. However, in research, he notes that your ideas often don't reach the end of line: 'You do stuff to publish a paper, or you don't have a tool that actually works. We wanted to take research ideas and implement them,' he explains. 'The flipside is that in industry you do actually have to deliver a working product.'

Working at an SME can be very flexible in terms of where and how you work, Babis notes. Perhaps the most challenging aspect of launching your own company is discipline: 'When you're your own boss, you have to manage yourself. You are ultimately responsible for what is delivered,' he points out.

> *"When you're your own boss, you have to manage yourself. You are ultimately responsible for what is delivered"*

What if none of these options appeals to you? Don't forget less conventional career paths into management, innovation management, marketing or business positions. These allow you to make a difference in the field by combining your technical knowledge with organizational, sales, interpersonal and other 'soft skills'.

Check out the talks from the Inspiring Futures session at CSW Lyon in our dedicated CSW playlist on #HiPEACTV:
bit.ly/CSWLyon_HiPEACTV

HiPEAC Jobs organizes regular events for anyone looking for advice with their career. For more information, along with careers resources such as articles and videos, check out the HiPEAC Jobs career centre:
hipeac.net/jobs/#/career-center

Looking for your next opportunity? HiPEAC Jobs is your go-to resource for the most interesting advanced computing jobs in Europe. Find your next career move here: hipeac.net/jobs


Photo credit: Ümit Yıldırım on Unsplash

The HiPEAC network includes almost 1,000 PhD students who are researching the key topics of tomorrow's computing systems. In this issue, Ali Mohammed explains how his thesis provides solutions for complex scheduling problems in high-performance computing.

Three-minute thesis

Featured research: Simulation-assisted scheduling for better balanced HPC loads

NAME: Ali Mohammed
RESEARCH CENTRE: University of Basel
SUPERVISOR: Florina M. Ciorba
THESIS TITLE: Design of Robust Scheduling Methodologies for High Performance Computing

This thesis came about in response to the need for load-balancing solutions in computationally intensive scientific applications on high-performance computing (HPC) systems with unpredictable application and system characteristics. Large, irregular loops represent the main source of parallelism in such scientific applications. The irregularities in loop iterations – along with system heterogeneity and the unpredictability of communications among parallel threads and processes – result in load imbalance, which has a negative impact on application performance.

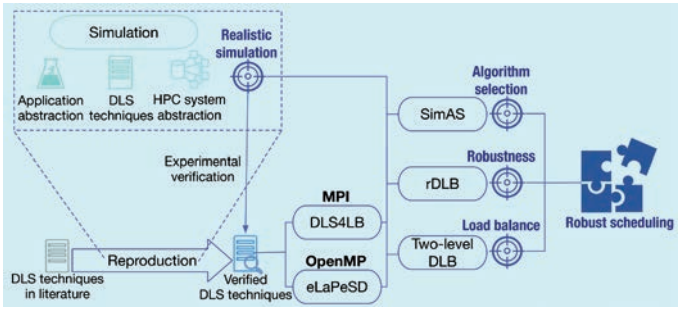
As applications and HPC systems become ever larger and more complex, robust scheduling solutions are required to solve this problem. A challenging topic, robust scheduling involves the study of the effects of scheduling on performance in isolation while application and system characteristics change unpredictably during execution. To add to the complexity, applications employ multiple levels of software parallelism to exploit the multiple layers of hardware parallelism.

As an analogy, imagine planning a schedule to transport a group of people between two points. Subgroups of people (a computation process) are assigned to different bike stations (compute nodes). People in a subgroup gather together (synchronize) to take a bus to their bike station. Each person (a thread) rides a bike (a central processing unit core) to the destination. People ride bikes (execute) at different paces, take different routes (application irregularity), and the bikes are not identical (hardware heterogeneity). Factors such as traffic lights, pedestrians, cars, and accidents are uncontrollable (unpredictable conditions).

Simulation is instrumental in solving these problems. We devised an approach for realistic simulations of parallel applications executing on HPC systems. These simulations enabled us to verify the implementation of dynamic loop scheduling (DLS) techniques, designed to schedule irregular application loops under system variability. DLS techniques were employed for the first time (in simulation and direct experiments) simultaneously at thread level and at process level for scheduling applications from the mathematics, computer vision and astrophysics domains.

We showed the existence of significant interplay between thread-level and process-level load balancing. In the presence of perturbations, we devised a simulation-assisted approach that switches to the most efficient DLS technique during execution. We also introduced a robust dynamic load-balancing approach (rDLB), which acts proactively, without needing perturbation / failure detection, to enhance performance by a factor of 30.

As part of this work, we have provided OpenMP and MPI scheduling libraries, as well as giving insights into the interplay between thread- and process-level scheduling. The methods devised improve the performance of scientific applications by up to 21% via two-level dynamic load balancing. This thesis enables a deeper understanding of scheduling at various levels and lays the groundwork for robust multilevel scheduling.



Ali's supervisor Florina M. Ciorba commented: 'Dynamic load balancing remains a key issue given the ever-growing needs of computationally demanding scientific applications. Indeed, load balancing is an issue of increasing importance given the convergence of software stacks and applications from high-performance computing, big data, and artificial intelligence / machine learning. By enabling clever scheduling of parallel applications, this work will ultimately contribute towards the breakthrough scientific discoveries of tomorrow, realized on high-performance parallel and distributed computing systems.'

Photo credits: Laura Vanzo / Visit Tampere

Computing Systems Week

Focus on smart IoT

26-28 April 2022 - Tampere, Finland

The latest disruptive computing technologies

Special session on European Digital Innovation Hubs

Ample networking opportunities

bit.ly/CSWSpring22_Tampere

#CSWSpring22