European Processor Initiative

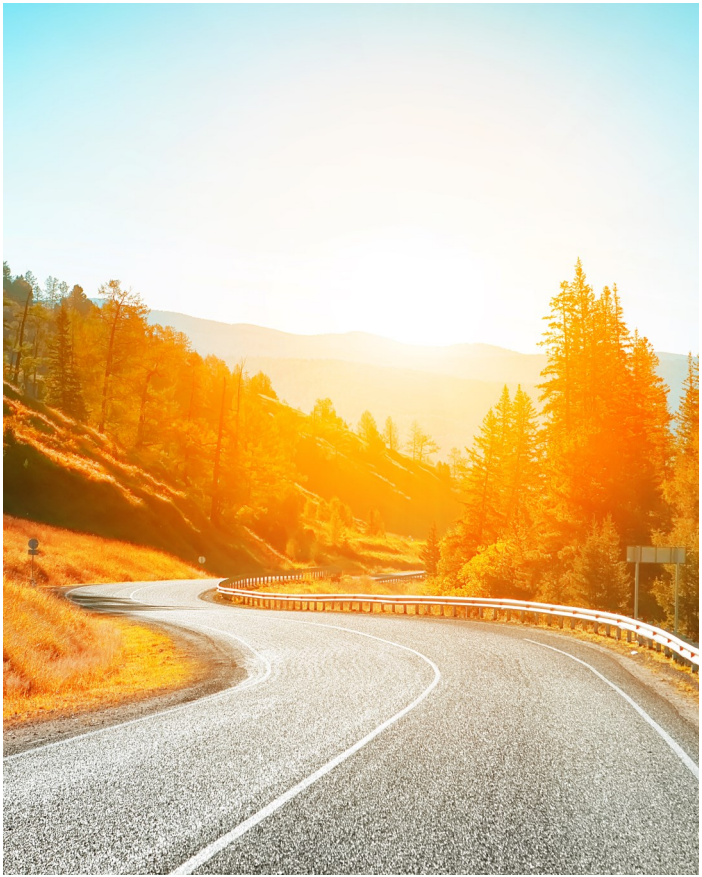
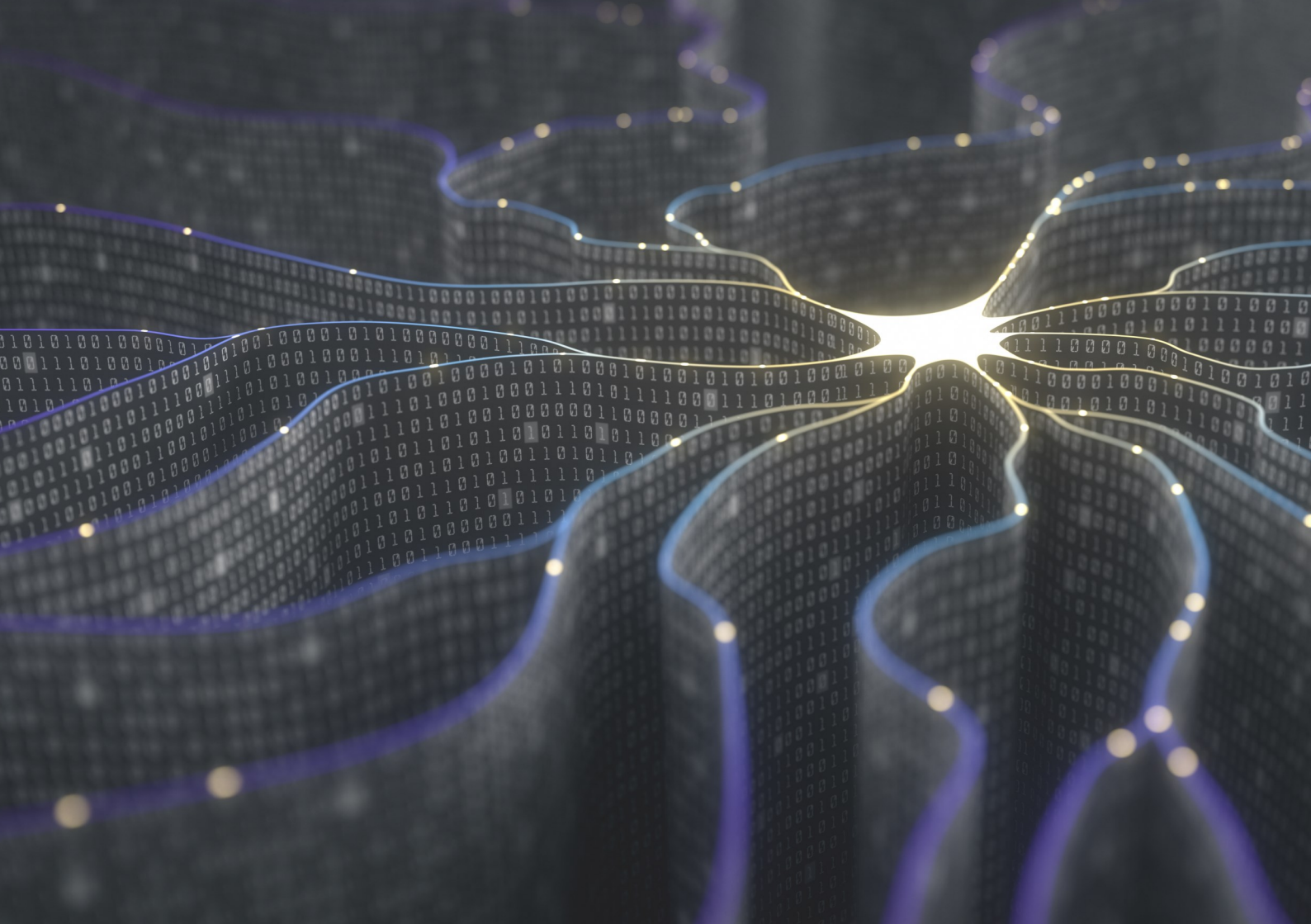


Table of Contents

- About EPI 2
- Forum Agenda 10
- Forum Sponsors..... 12
- Keynotes and Talks 16

Forum hashtag: #EPIForum2024



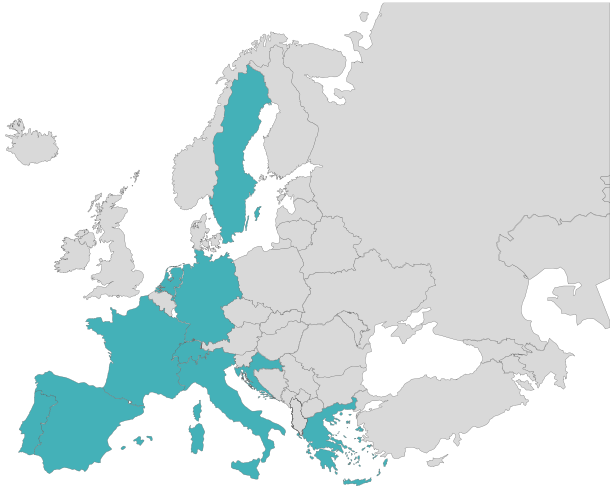
The Central Goals of the Initiative are:

- Contributing to the development of **European supercomputing technologies** that can compete on the global HPC market
- Developing **key components** for the European Union to equip itself with a world-class supercomputing infrastructure
- Strengthening the competitiveness and leadership of **European industry** and science
- Developing European microprocessors and accelerator technologies with drastically **better performance and power ratios**
- Tackling important segments of broader and/or emerging **HPC and Big-Data** markets

The European Processor Initiative (EPI) is a project currently implemented under the second stage of the Framework Partnership Agreement signed by the Consortium with the European Commission (FPA: 101036168), whose aim is to design and implement a roadmap for a new family of low-power European processors for extreme scale computing, high-performance Big -Data and a range of emerging applications.

The project aims to deliver a high-performance, low-power processor, implementing vector instructions and specific accelerators with high bandwidth memory access. The processor will also meet high security and safety requirements. This will be achieved through intensive use of simulation, development of a complete software stack and tape-out in an advanced semiconductor process node. The project will provide a competitive chip that can effectively address the requirements of the HPC, AI, and trusted IT infrastructure markets.

The European Union (EU) faces a critical need for processor supply and high-performance computing (HPC), necessitating the securing supply chains and systems to ensure sovereignty. Part of this effort is through EPI, which is currently in Phase 2 (2022-2025), and involves a consortium of 27 strategically selected European academic and industrial partners.



EPI Timeline:

EPI 1

2019-2021

2022

EPAC V1.5 sent for manufacturing

KVX RISC-V based accelerator architecture definition

EPAC - Accelerator

v1.5 Platform & SDV

Menta eFPGA IP definition

Kalray RISC-V SDK release

2023

2024

Rhea1 Family - Gen1 GPP

Arm & RISC-V (STX, VRP, ...)
Arm Neoverse V1 Core – N6
External IPs

RISC-V KVX FPGA emulator

EPAC - Accelerator

EPAC IPs forwarded to EUPILOT VEC and MLS tapeouts

Rhea2 Family - Gen2 GPP

Dual chiplet implementation

First EU exascale system with Rhea1

EPAC - Accelerator

EPAC IPs taped out in EUPILOT project

EPAC IPs integrated into multi-core FPGA emulation platform

2025

2026

EU exascale systems with Rhea2

Two EPI Processors:

RHEA (GPP)

EPAC (Accelerator)

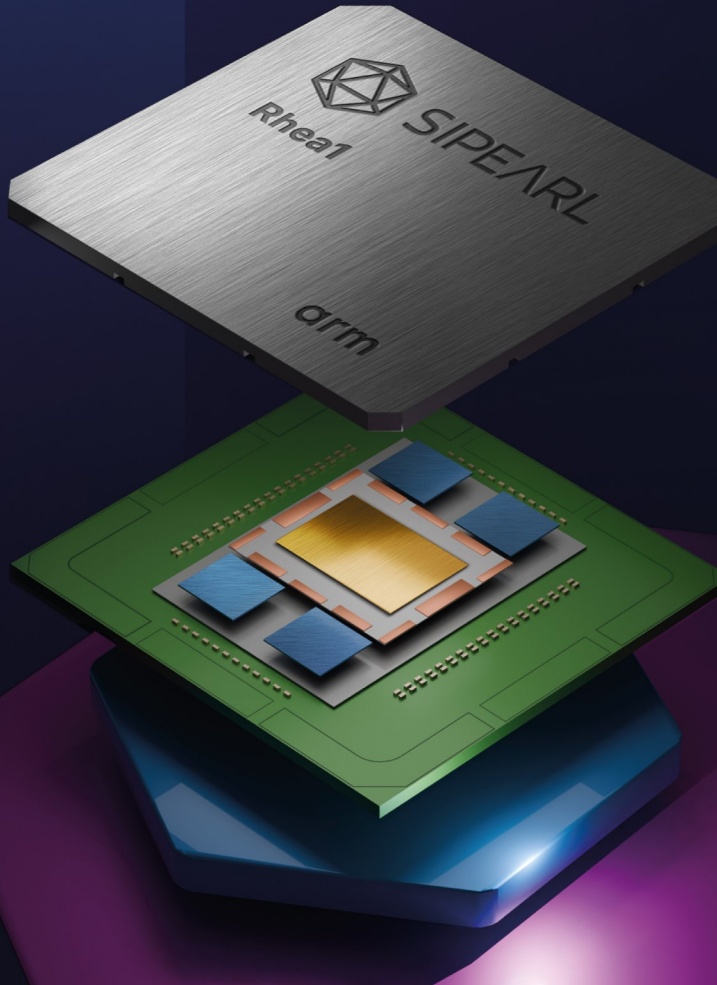
RHEA 1

HPC microprocessor

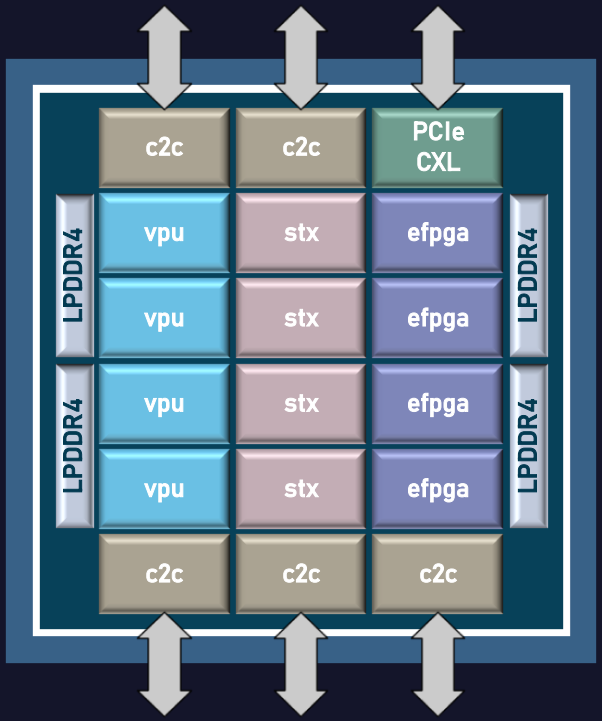
80 arm® Neoverse V1 cores
with 2 x 256 SVE each

4 x HBM

4 x interfaces DDR5



EPAC
Accelerator



Based on RISC-V ISA

VEC - Self-hosted RISC-V CPU + wide VPU (256 double elements) supporting RVV 0.7.1 / 1.0

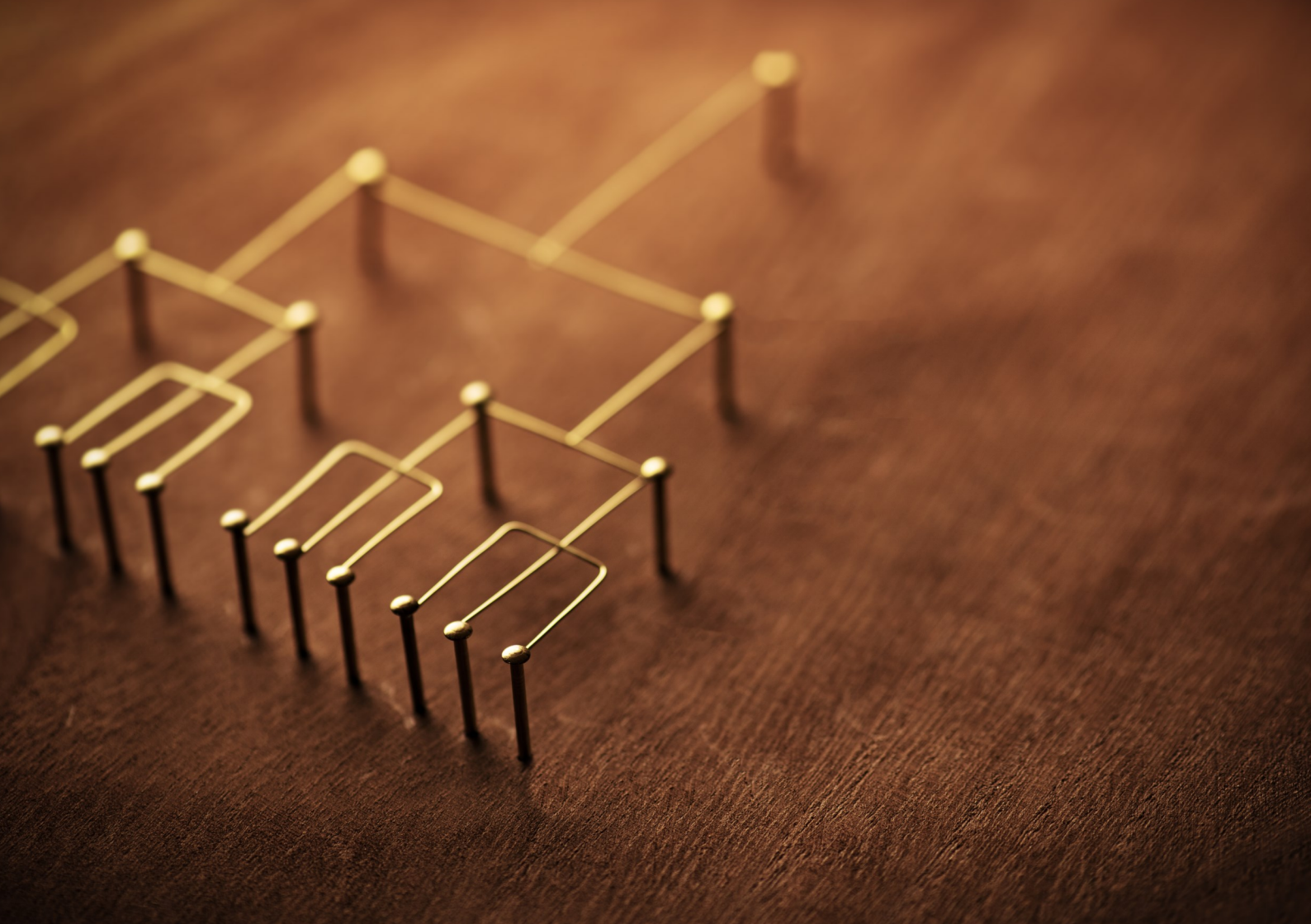
STX - RISC-V CPU + specific cores for stencil and neural network computation

VRP - RISC-V CPU with support for variable precision arithmetic (data size up to 512 bit)

eFPGA - On-chip reconfigurable logic

Ziptillion - IP compressing/decompressing data to/from the main memory

KVX - FPGA demonstrator of the Kalray RISC-V CPU targeting HPC and ML



DAY 1—WEDNESDAY, OCTOBER 9

Intro and EPI 15:00-15:30

- **EPI Forum Introduction** (Eric Monchalin, Eviden, an Atos Business & EPI Chair)
- **EPI Overview** (Etienne Walter, Eviden, an Atos Business & EPI General Manager)

Keynotes 15:30-16:30

- **AI Keynote** (Stephane Requena, GENCI)
- **European Supercomputers Keynote: Buy versus Build** (Mateo Valero, BSC)

Global Technology—Platinum Sponsors 16:30-17:10

- **High Performance Computing in the AI Era** (Jean-Pierre Panziera, Eviden, an Atos Business)
- **Enabling European Innovation with Arm Technology** (Eric Lalardie, Arm)

Coffee break 17:10-17:40

Global Technology—Gold Sponsors 17:40-18:10

- **AMD Solutions for HPC and AI** (Daniele Piccarozzi, AMD)
- **Overview of SiPearl's Seine Platform** (Philippe Notton, SiPearl)
- **Semidynamics All-in-One Solution for Next-Generation RISC-V AI** (Roger Espasa, SMD)

ARM/Rhea in High-End HPC 18:10-18:40

- **Rhea GPP Technology, Key Features of Rhea1, Targeting HPC and AI Inference Workloads with Best-in-Class Energy-Efficiency, Exascale System Level EPI View** (Craig Prunty, SiPearl)

Panel 18:40-19:25

- **Round Table: AI Is Shaking 64bit Computing, What About HPC; Codesign View; Challenges in SoC Processor Design**

Wrap-Up Day 1 19:25-19:30

EPI Forum Dinner 20:30-22:00

DAY 2—THURSDAY, OCTOBER 10

Keynote 9:00-9:30

- **EuroHPC Chips Initiatives: The Road Towards European Technological Sovereignty** (Alexandra Kourfali, EuroHPC JU)

Global Technology—Platinum Sponsors 9:30-9:50

- **Enabling AI Nations** (Rod Evans, NVIDIA)

RISC-V Accelerators 9:50-10:50

- **Pushing RISC-V to HPC: the story of the EPI Accelerator (EPAC)** (Filippo Mantovani, BSC)
- **VEC, VRP, STX** (Filippo Mantovani, BSC, Andrea Bocco, CEA, Tim Fischer, ETZH)

Coffee break 10:50-11:20

11:20-11:35

- **Fujitsu-Monaka: Data Center Ready Arm Processor** (David Snelling, Fujitsu)

Panel 11:35-12:25

- **The View on RISC-V Future**

Forum Closing 12:25-12:35

- Eric Monchalin, Eviden, an Atos Business, EPI Chair

EPAC Demo

Mario Kovač, Forum Moderator, EPI Chief Communications Officer

Platinum Sponsor – Arm

Born over 30 years ago with the goal of designing a computer intended to run on a battery, Arm's DNA is built around power-efficient CPUs. Today, Arm compute platforms are the most power-efficient on the planet and continue to push the thresholds of performance to enable the next generation of smart, AI-capable experiences on everything from the smartphone to the automobile and to the datacenter. This unique combination of performance and efficiency enabled Arm to change the world by fueling the smartphone revolution - and it's why Arm will power every technology revolution moving forward. With Arm Neoverse, we are empowering market leaders to design for the AI future. With Neoverse CSS and our Arm Total Design network of partners, Arm is reducing friction and lowering the cost of silicon development, letting our partners set the pace for leading-edge silicon innovation. The future of silicon is built on Arm.

Platinum Sponsor – Eviden

Eviden is a next-gen technology leader in data-driven, trusted and sustainable digital transformation with a strong portfolio of patented technologies. With worldwide leading positions in advanced computing, security, AI, cloud and digital platforms, it provides deep expertise for all industries in more than 47 countries. Bringing together 47,000 world-class talents, Eviden expands the possibilities of data and technology across the digital continuum, now and for generations to come. Eviden is an Atos Group company with an annual revenue of c. € 5 billion.

Platinum Sponsor – NVIDIA

NVIDIA (NASDAQ: NVDA) is the world leader in accelerated computing.

NVIDIA pioneered accelerated computing to overcome challenges that others have not been able to solve. Their innovations in AI and digital twins are reshaping the world's largest industries and having a profound impact on society.

Gold Sponsors - AMD, Semidynamics, and SiPearl

For more than 50 years AMD has driven innovation in high-performance computing, graphics and visualization technologies. Billions of people, leading Fortune 500 businesses and cutting-edge scientific research institutions around the world rely on AMD technology daily to improve how they live, work and play. AMD employees are focused on building leadership high-performance and adaptive products that push the boundaries of what is possible. For more information about how AMD is enabling today and inspiring tomorrow, visit the AMD (NASDAQ: AMD) website, blog, LinkedIn, Facebook and X pages.

Semidynamics is a European supplier of RISC-V IP cores, specializing in high-bandwidth high-performance cores with vector units targeted at machine learning and artificial intelligence applications. With an extraordinary team with fifteen years of experience in the semiconductor industry, we can help you succeed in your next project, from specification to verification. We can also partner with your in-house team and provide staff augmentation services, both on-site and remotely.

SiPearl is building the European high-performance low-power microprocessor dedicated to supercomputing and artificial intelligence. This new generation of microprocessors will first target EuroHPC Joint Undertaking ecosystem, which is deploying world-class supercomputing infrastructures in Europe for solving major challenges in medical research, security, energy management and climate with a reduced environmental footprint.

EPI Overview (Etienne Walter, Eviden, an Atos Business & EPI General Manager)

EPI Overview will cover the second phase of EPI, including developments in SGA2, Software Development Vehicle, GPP, EPAC 1.5, and current results.

AI Keynote (Stephane Requena, GENCI)

After few epochs of deep winters and thanks for the joint availability of new models (ANN, CNN, Transformers, Mamba…), huge volume of data across Internet and worldwide data sources and the performance of GPU-based clusters, artificial intelligence and more precisely deep learning has become one of the leading technology for ingesting, inferring, analyzing data toward new insights across a growing number of disciplines. The training and the fine-tuning (specialization) of AI models are by nature requiring the use of large computing resources which make happening nowadays a triple convergence between HPC and AI (infrastructures, services and usages). This keynote will address how GENCI together with the 3 national centers are implementing since 2019 this convergence serving scientific and industrial end users communities and how this vision and its associated stakes are embedded into the design of the 2nd Exascale EuroHPC system called Alice Recoque in the field of the newly announced European AI strategy called « AI Factories ».

European Supercomputers Keynote: Buy versus Build (Mateo Valero, BSC)

In 2018, Europe launched the EuroHPC initiative and created the EuroHPC Joint Undertaking funding structure with two main objectives. The first is to acquire, build, and deploy a world-class High-Performance Computing (HPC) infrastructure across Europe. The second objective is to develop Made-in-Europe HPC hardware and the applications (software) needed for future European supercomputers. The talk will cover both objectives and introduce the future developments in software and hardware for the MareNostrum 6 supercomputer, expected by 2027-2028.

High Performance Computing in the AI Era (Jean-Pierre Panziera, Eviden, an Atos Business)

The AI wave is drastically changing the whole IT ecosystem. Domains which have traditionally used HPC for Science and Engineering are now being completely transformed. The introduction of new computational models and the use of the new hardware accelerators brings new challenges to the user and the technology provider communities.

Enabling European Innovation with Arm Technology (Eric Lalardie, Arm)

Europe's semiconductor R&D projects and startups are developing prototypes and new products alongside growing the skills and capabilities base. Within Europe, companies and institutions are exploring technologies to address from edge to the data centre, and at the front of research in areas such as quantum computing. In this talk we show how Arm is enabling its partners with CPU technology and preconfigured compute subsystems that enables these solutions to be delivered faster and with a software ecosystem that reduces barriers to entry.

AMD Solutions for HPC and AI (Daniele Piccarozzi, AMD)

The talk will introduce AMD perspective on HPC and AI infrastructure and its open ecosystem approach.

Overview of SiPearl's Seine Platform (Philippe Notton, SiPearl)

As part of the development of high-performance microprocessors, server board design as a reference is essential for the server design of our direct customers. To address their various needs, SiPearl has developed a single modular reference server solution. In this keynote we provide an overview of the platform, the functional requirements, and the resulting design that will allow us to successfully bring up test and disseminate our solution among our partners.

Semidynamics All-in-One Solution for Next-Generation RISC-V AI (Roger Espasa, SMD)

In this talk we will describe Semidynamic's solution for future-proof AI compute, based on the combination in a single element of Semidynamics RISC-V core, vector and tensor unit. We will cover the new tensor instructions implemented by Semidynamics, how these can be used in AI convolutions and matrix multiplication. We will also cover the need for the vector unit in modern AI models, such as LLMs, to properly run activations.

Rhea GPP Technology, Key Features of Rhea1, Targeting HPC and AI Inference Workloads with Best-in-Class Energy-Efficiency, Exascale System Level EPI View (Craig Prunty, SiPearl)

Designed using high-performance energy-efficient Arm® Neoverse V1 cores, the European microprocessor Rhea1 was conceived by the EPI consortium to be brought to life by SiPearl. Rhea1 will be perfectly suited to traditional HPC workloads—its initial target market – and AI inference workloads. Thanks to both a generous memory capacity and high bandwidth in-package HBM, it will deliver extraordinary performance and energy-efficiency with an unrivalled byte-per-flop ratio. This talk describes the key features of Rhea1 and how they address the requirements for HPC and AI inference HPC at Exascale level.

EuroHPC Chips Initiatives: The Road Towards European Technological Sovereignty (Alexandra Kourfali, EuroHPC JU)

The EuroHPC JU is cultivating a robust ecosystem that accelerates chip advancements. It has recently expanded its mandate to include the creation of AI Factories designed to support the development, testing, and validation of large-scale AI models. Advancements in high-performance computing (HPC) are influenced by the design and development of specialized chips tailored for AI applications. The design of cutting-edge chips for HPC aligns with the EU's broader goal of technological sovereignty. Looking ahead, chip initiatives are evolving with a strong emphasis on open-source technology, particularly RISC-V architectures that will enhance Europe's strategic autonomy in processor and accelerator design and create a robust ecosystem of research institutions, startups, and SMEs that design and produce RISC-V-based processors and accelerators. In this talk, we will revisit how the emergence of AI factories and open-source technologies are facilitating a future robust ecosystem for HPC as key enablers for innovation that will contribute to Europe securing its position as a leader in semiconductor technology and long-term technological sovereignty.

Enabling AI Nations (Rod Evans, NVIDIA)

There is a global policy imperative to harness AI as an engine for economic growth, as a protector of public health, and as a tool for enhancing and expanding citizen services. By combining our GPU-accelerated computing platform with programs that support domestic AI ecosystems and capacity building, NVIDIA helps governments implement national AI strategies. This presentation will outline how NVIDIA is enabling countries to protect their sovereign data investing in sovereign AI to develop and harness such benefits on their own. Sovereign AI refers to a nation's capabilities to produce artificial intelligence using its own infrastructure, data, workforce and business networks.

Pushing RISC-V to HPC: the story of the EPI Accelerator (EPAC) (Filippo Mantovani, BSC)

The European Processor Initiative (EPI) is a project dedicated to developing a general-purpose processor and an accelerator, alongside the necessary software layers for their integration into the High Performance Computing (HPC) ecosystem. The EPI Accelerator (EPAC) is a collection of RISC-V-based IPs that demonstrated the viability of RISC-V technologies on High-Performance Computing (HPC). The Barcelona Supercomputing Center is contributing to the development of a RISC-V-based accelerator targeted at HPC applications, leveraging the RISC-V vector extension. This talk aims to provide an introduction to RISC-V, a description of the EPAC accelerator as well as insights and findings of the VEC accelerator developed for EPAC. Special emphasis will be placed on the RISC-V vector extensions (RVV), with a particular focus on implementations utilizing large vectors.

VEC, VRP, STX (Filippo Mantovani, BSC; Andrea Bocco, CEA; Tim Fischer, ETHZ)

VRP: The growth in scale of numerical linear systems kernels, has led researchers to move from direct solvers to Krylov-subspace iterative solvers. However, the latter suffer from an important instability due to accumulated round-off errors that may prevent convergence. Extended precision reduces relative error during computation by increasing the size of floating-point mantissas. State-of-the-art solutions exploiting extended precision are software-based. These solutions suffer from an important performance penalty. Our solution is the VaRiable Precision RISC-V hardware accelerator (VRP). It enables arithmetic operations and memory loads/stores, both with extended-precision floating-point numbers, both with up to 512-bit of mantissa precision, at bit granularity. We show how our VRP processor, implemented in both silicon in GF22FDX technology and FPGA, improves the convergence and the efficiency on multiple variants of the conjugate gradient kernel. We observe up to an 8X improvements on kernel iteration count, and up to a 40 % improvement on latency. In addition, by augmenting stability, the VRP enables to solve linear ill-conditioned systems without costly compensating techniques.

STX: Accelerating Sparse & Dense Linear Algebra: STX is a stencil accelerator based on a cluster of RISC -V Snitch cores, each tightly coupled with a high-performance FPU. It achieves high FPU utilization with two ISA extensions: 1) Sparse Stream Semantic Registers (SSSR) to accelerate stencil access patterns, and 2) Floating Point Repetition (FREP) to reduce control overhead.

Fujitsu-Monaka: Data Center Ready Arm Processor (David Snelling, Fujitsu)

Based on the same architectural principals as Fujitsu's A64FX (the processor at the heart of Fugaku), MONAKA is the next iteration of Fujitsu's ARM architecture. This iteration focuses on energy efficiency and deployment flexibility through novel technology features and architectural strategy. The talk will briefly summarise the key technology advances and how they contribute to the energy efficiency and suitability of MONAKA to data centre and edge applications. We will also highlight the importance of the ARM software ecosystem.

European Processor Initiative

TÉCNICO LISBOA

EVIDEN

ETH zürich

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

KERNKONZEPT

SURF

JÜLICH Forschungszentrum

semidynamics

KALRAY

Fraunhofer

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

CHALMERS

FER
UNIVERSITY OF ZAGREB
Faculty of Electrical Engineering and Computing

E4 COMPUTER ENGINEERING

CINECA

cea

UNIVERSITÀ DI PISA

FORTH
FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS

life.augmented

EXTOLL
latency matters.

KIT
Karlsruher Institut für Technologie

PROVE & RUN

GENCI

menta

SIPEARL

LEONARDO

ZeroPoint