Toward an international alignment of Post-Exascale roadmaps

Raymond Namyst

Inria

About yesterday's roadmapping session

- Japan
  - The FugakuNEXT project: an architecture to support HPC and AI workloads, Masaaki Kondo
- Latin America
  - Overview of HPC, AI and Quantum activities in LAC, Carlos Hernandez
- Chile
  - The Inria Chile driving force for technological innovation in AI, Nayat Sanchez-Pi
- DOE
  - Impact of co-scientist agentic systems on scientific production (and future AI facilities), Franck Cappello
  - Challenges in meeting energy needs of upcoming HPC/IA systems, Jeff Vetter
- France
  - Objectives of the SPE project, Raymond Namyst

« Strategy for Post-Exascale » : a European Coordination Support Action

- Coordinated by Inria (Jean-Yves Berthou)
- April 2026 – March 2029
- Main objective: help

Objective 1

High-quality HPC/AI roadmap identifying research and innovation gaps, disruptive technologies, and emerging needs in Europe. Policy Paper Factory will produce short, targeted policy notes on identified topics.

Objective 2

Build and animate a Community of Practice (CoP) bringing together experts from academia, research infrastructures, industry, civil society and European organizations and in particular AI Factories and future Giga Factories.

Objective 3

Represent Europe in the International post-Exascale (InPEx) initiative, positioning Europe as a leader and key contributor to global discussions and workshops. Beyond InPEx, collaborations will be established with major European and international initiatives/events

SPE Partners

- Inria (FR)
  - University of Bordeaux
  - GENCI
- BSC (ES)
- CSC (FI)
- FZJ (DE)
- CEA (FR)
- CERN (CH)
- TUM (DE)
- BDVA (BE)
- QuIC (DE)
- ETP4HPC (NL)
  - Fraunhofer
  - Partec
  - SiPearl
- GRNET (GR)
- ICSC (IT)
  - CINECA
  - INFN
  - UNITO
- UEDIN (UK)
- Neovia (FR)
- PCSS (PL)

Roadmap refresh

- Template sent to partners to collect preliminary inputs
  - Context (landscape, key trends)
    - What global changes matter most in the next 5-10 years?
  - Existing academic and industrial ecosystem (stakeholders)
    - Who are the core stakeholders to involve in a post-exascale roadmap?
  - Strengths and Weaknesses analysis (positioning)
    - What weaknesses could become critical within 3-5 years?
  - Scientific and Technical bottlenecks, opportunities and disruptions
    - Most important scientific bottlenecks
    - Possible disruptions
  - Societal, scientific and economical impact

Several reports were collected so far

- From
  - EuroHPC advisory groups
  - European computing centers
  - Research institutions
  - National Strategy Agencies

- Still under consolidation

Scientific and Technical directions

- HPC/AI Hybridization
  - AI-accelerated simulation, AI-driven digital twins
  - Controlled integration of AI
    - Hybrid algorithms
    - New models
    - Robustness
  - Interoperability and composability HPC + AI (+ Cloud)
    - Workflows, APIs
  - Bottlenecks
    - Reliable, curated training data sets
    - I/O bottlenecks under AI+simulation workloads
    - Skills and talents: lack of hybrid HPC-AI-ops profiles

Scientific and Technical directions

- AI-assisted numerical simulation
  - Generalization of AI in scientific workflows
    - Coding assistants, Orchestration, Oracles
    - Exploit European expertise to improve agent orchestration
    - Need for science-oriented genAI
  - The Software Heritage goldmine
  - Trust/IP/confidentiality concerns
  - Scale of investment not reachable by public actors
    - Public/private continuum of infrastructures and services
- Energy and sustainability considerations
  - Reduce the inference cost
    - Futher develop inference-specific accelerators
    - Use specialized agents

Scientific and Technical directions

- Quantum computing integration
  - Develop a full HPC-QC software stack
  - Drive progress using applications
    - Develop and scale converged HPC/AI/QC applications
    - E.g. Fluid dynamics, Drug discovery, Satellite image analysis, Medical imaging

- Emerging, new computing paradigms
  - Neuromorphic, analog computing

Scientific and Technical directions

- Shift from monolithic systems to modular/disaggregated architectures with faster upgrade cycles
  - Chiplets, photonics interconnect
  - Software stacks for strongly heterogeneous, fragmented hardware
  - HPC, AI and Quantum
    - Are different paradigms
    - Use different software stacks
  - Need to rethink algorithms

General challenges

- Energy Efficiency & Sustainability
  - Constant increase of power per ASIC, per rack
  - Cooling, power availability
- Evolution of HPC supercomputers
  - Many of them serve as AI Factories as well
  - Shift from traditional batch-oriented HPC systems toward AI-driven, interactive, API-based platforms
- Need for stronger collaborations and alignment
  - EU has developed a strong Middleware/application ecosystem
    - But we should avoid dispersion of efforts / multiplication of software
  - Agree on a common software basis to be long-term supported?

Building an international roadmap

- Objective
  - Collectively establish an InPEx post-exascale roadmap
    - Actionable in each region
  - Synchronization point every 6 months
  - With a first milestone in sept. 2026
    - initial vision
- Process
  - Let us organize how we will work
    - Need for an agile and small-size team
    - 1 representative person per continent (+ 1 backup) ~ 8 persons
      - Monthly meetings

Let’s move on!