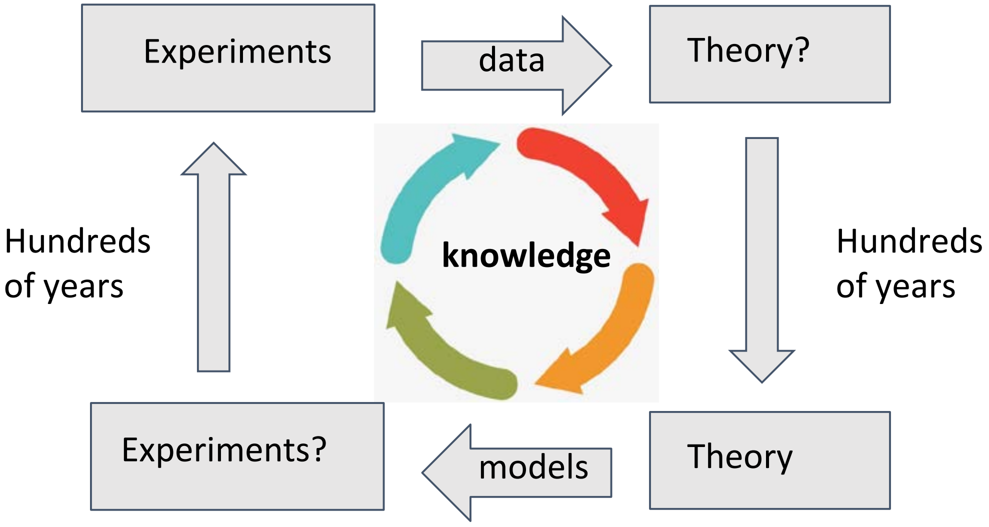
Grand challenge applications: technical requirements for the exascale era

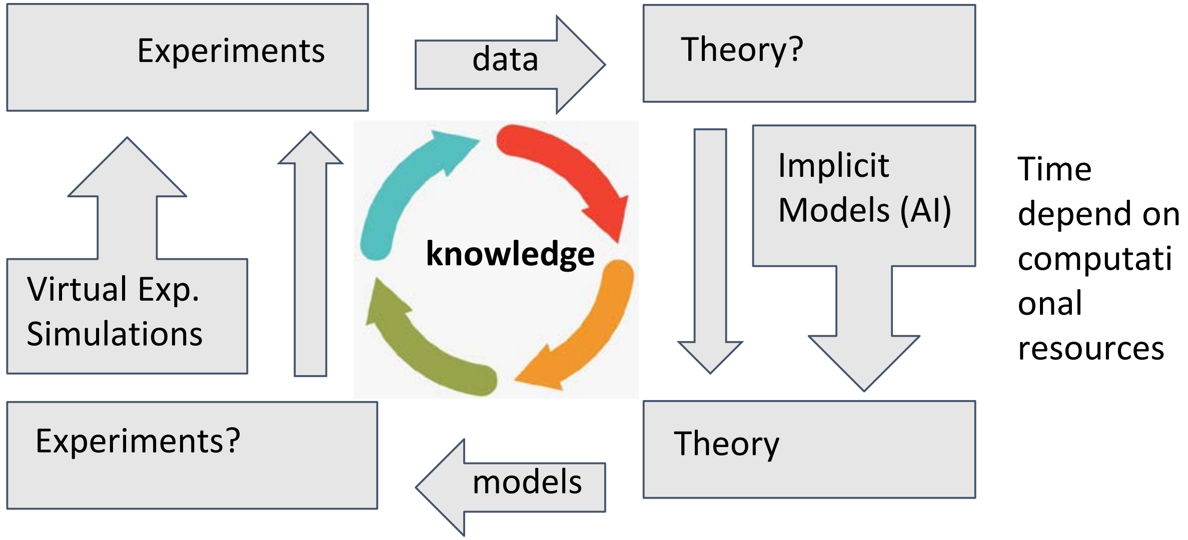
Carlo Cavazzoni
CINECA
EuroHPC RIAG advisor

The persistence need for more powerful computational resources

Scientific Progress Without Computers
(two paradigms)



Scientific Progress With Computers
(four paradigms)



HPC & Data Processing are a knowledge accelerator: the more the better!

The progress of modern scientific research itself leads to an ever-increasing need for computational capacity as an inevitable consequence.

“Incomplete” list Scientific cases (EuroHPC UG)

	Material Science	Life Science & Bio	Meteo and Climate	Earth Science	Astrophysics & Planetary	Fundamental Physics	Engineering	Energy	Social Sci, Fintech
Batch Simulation (numerical integration)	Quantum Chemistry, atomistic simulations	Molecular Modeling, Brain simulation	Climate and Weather	Volcano, earthquake, sea dynamic, rivers, etc...	Cosmo, galaxy, stars, etc.. formation and dynamic	LQCD	CFD, DNS, LB, FE	Fusion (ITER)	Training DNN
Data Driven (data analysis)	Synchrotron and Neutron sources	NGS, HBP	Radar	Seismic	SKA, GWaves, Satellites	LHC	IoT, sensors	RTM (Oil&Gas)	Fraud detection, Sentiment. AI for complexity
Data Assimilation (simulation + data)		Personalized medicine	Weather	Shallow water sim. Oceanograph y	Space missions (e.g. GAIA)	Solar wind forecast	Autonomous Vehicles	FWFI (Oil&Gas)	Risk Management, personalized finance (pensions)
Parameter Space (ensemble simulation)	New engineered Materials, designed from scratch	In silico Drug Design	uncertainty quantification	Natural hazard assessment	Star Life Cycle, Black Hole, Neutron Star		Aerodynamics, wings, propellers, turbines	Batteries	
Simulation steering (interactive and visualization)	Apply forcing. Tribology Friction	Molecular Mechanic Simulation					Digital Twins	Reservoir Simulation	
Coupled or chained simulation (workflows)	Property Calculators, Multi-scale simulations		Include soil and ocean effects		Cosmology & Galaxies		Multi-physics simulations	wind turbines	

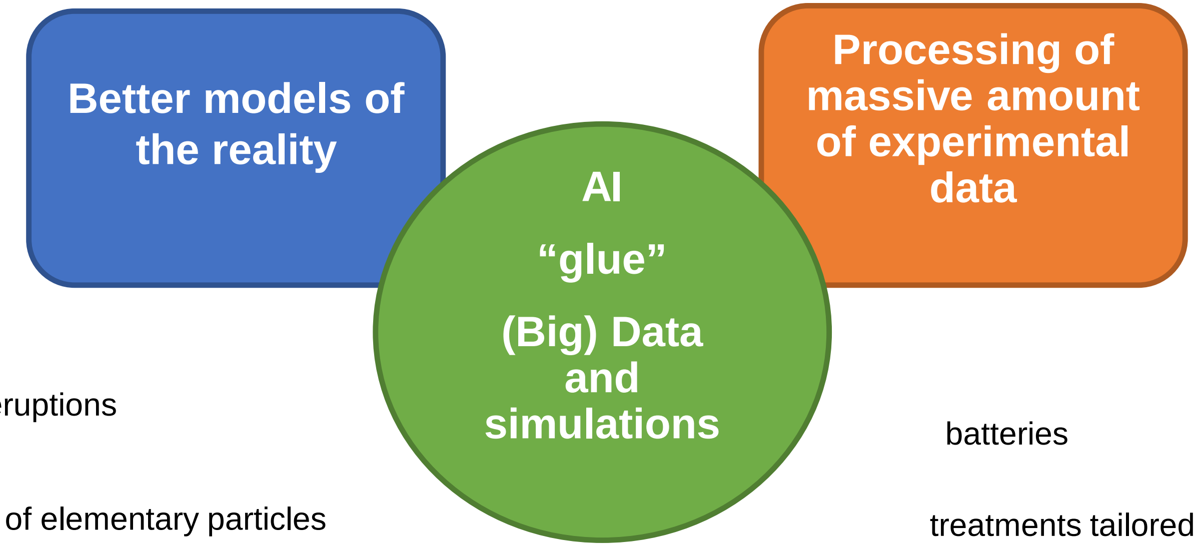
European exascale system will enable

new material from scratch

New algorithms

use information from sensors deployed in an internet-of-things

Simulate formation of galaxies



direct numerical simulations of the Navier-Stokes equations

source of gamma-ray bursts

computing to be applied in a whole range of non-traditional areas

predict how solar eruptions

batteries

properties of elementary particles

Next-generation weather forecasting

treatments tailored both to the patient

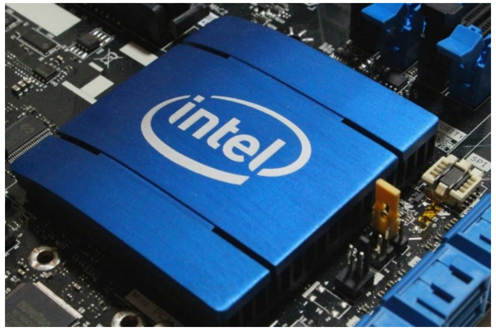
uncertainty quantification in predictions

understanding of general relativity

Improve the efficiency of nuclear power, hydropower, wind turbines

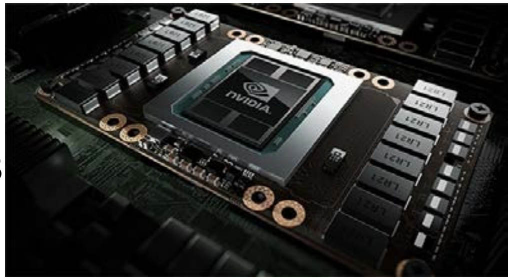
advance quantum chromodynamics

Floating-point throughput



Scientific computing is still dominated by floating-point usage

- increased resolution of grids
- larger system sizes,
- more accurate computational methods



Will require at least a factor 50-100x greater compute performance over the next few years

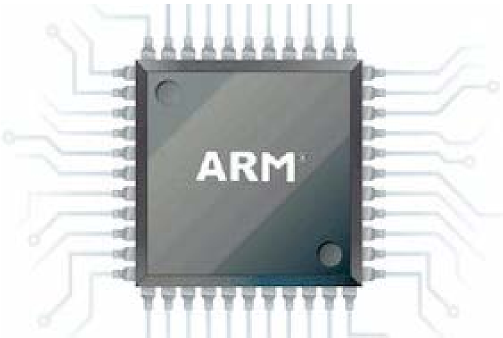
- This corresponds to exascale
- many application areas have a roadmap of the needs at least one order of magnitude further

Accelerators as well as vector units on traditional CPUs

- significant rise in the fraction of codes that have been rewritten
- higher throughput of single precision arithmetics,
- algorithms still require double precision.

***From the user point-of-view, it is important that the EuroHPC systems*:**

- **deliver very high double precision throughput as well as**
- **making use of the extra performance that comes from single precision.**



RISC-V

European Processor Initiative

Memory size, throughput and latency

Many applications are limited by memory bandwidth or communication latency

maintain, and ideally increase, the amount of memory per core *(what memory?)*

communication latency between chips and nodes has to be reduced

by at least an order of magnitude

to enable algorithms based on fast iterations to achieve significantly improved performance.

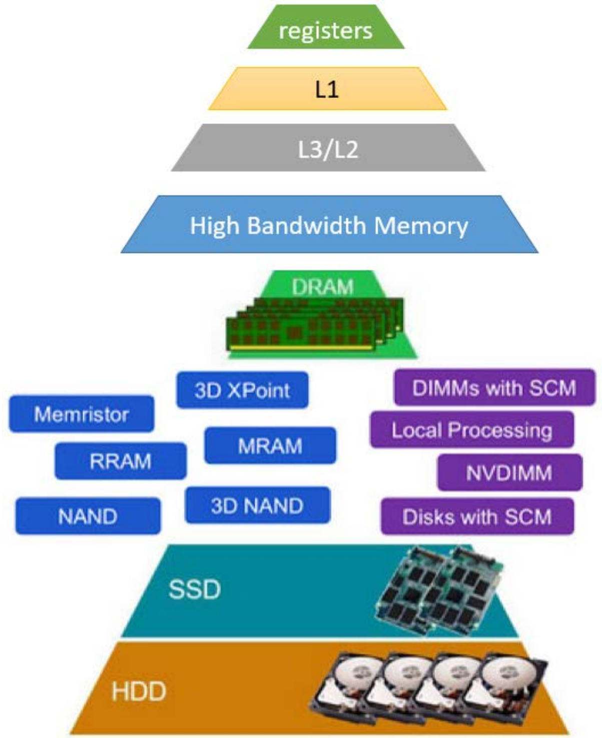
Several scientific problems are also limited by the memory bandwidth inside each node

some of them have been able to benefit from accelerators this is far from true for all.

It is important to maintain at least one "classical" system with CPUs and large memory bandwidth.

Co-design opportunity:

applications re-factored/rewritten to cope with many memory tiers (HBM,NVRAM)



Interconnect bandwidth & latency

data movement is, or is becoming, extremely important in many applications

operation counts are not necessarily a good indicator of time to solution
(when) computational efficiency is dominated by the cost of moving data.

In some applications interconnects cannot saturate the compute capability

their scalability is limited by network latency.

System interconnects should therefore

provide good all-to-all communications,
high bi-sectional bandwidth and low inter-node latency and
network topologies that scale to many thousands of nodes
maintaining low latencies and high bisection bandwidth will be important

Mellanox (NVIDIA)

OPA (Intel)

Slingshot (Cray)

Tofu (Fujitsu)

BXI (Bull)

Storage & I/O

Storage and I/O requirements are expected to grow even faster than compute needs,
with 100x greater needs than today, within the next 5-10 years,
in particular with much larger data sets being used e.g. for data-driven research and deep learning.

large-scale end-to-end data e-infrastructure to collect, handle, analyse, visualize, and disseminate petabytes to exabytes of data,
for instance by providing high-bandwidth gateways to future EOSC-provided dissemination.

increased reliance on databases, object storage, and many other new technologies.
While scientists are aware of these technologies, they have not been implemented in the vast majority of codes,
and at least the first phase of EuroHPC will thus have to rely on traditional filesystems for data handling.

Operations & Environments

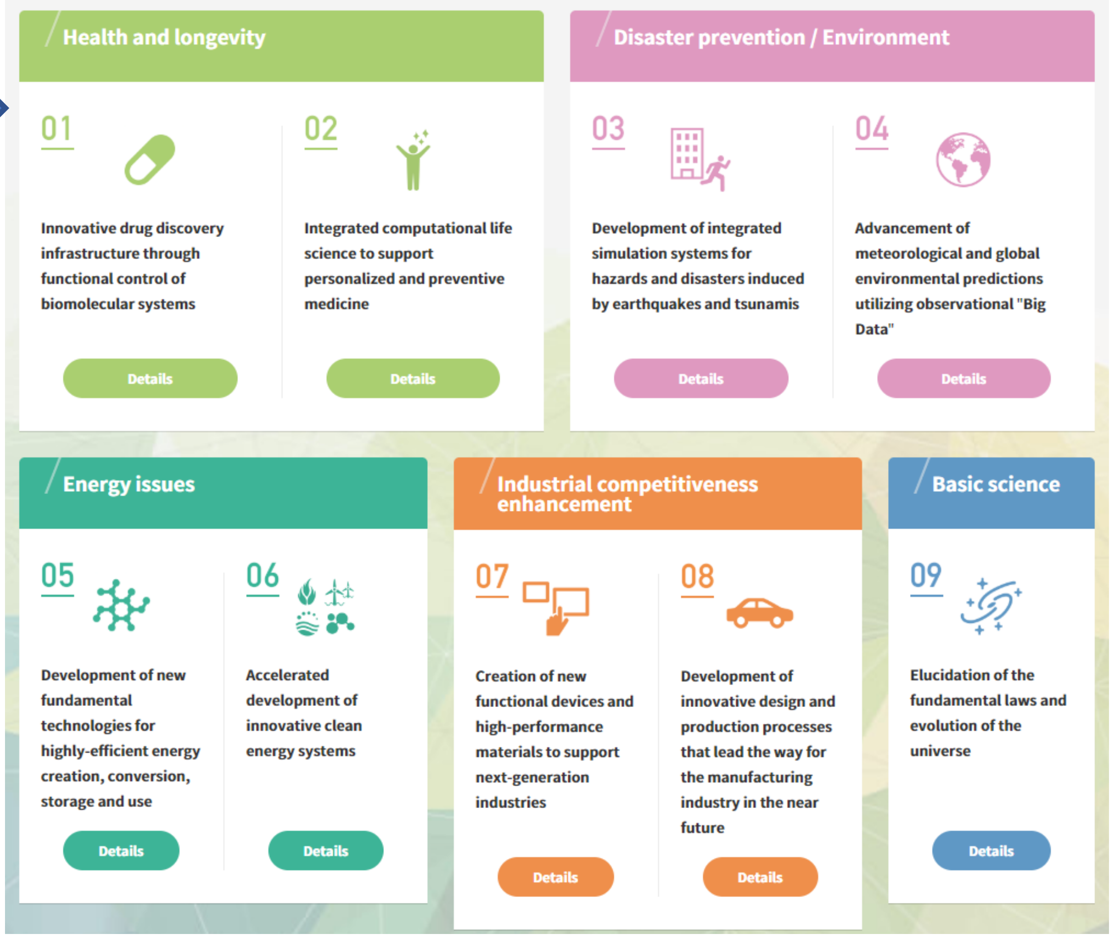
- Full Linux environment
- POSIX filesystem
- Fortran, C and C++
- GNU and LLVM compiler suites
- Parallel Debugger and Profiler
- Support for the latest available MPI and OpenMP standards
- HPC basic libraries (BLAS, LAPACK, SCALAPACK, FFT, etc…)
- Perl and Python fully supported on all nodes
- A standard API for accelerator (OpenACC, OpenCL, OpenMP)
- Queue systems able to handle ensembles simulation
- Containers with MPI
- Competent staff support ensured for longer period of time
- Comprehensive training program from novices to experts

Scientific Use Cases

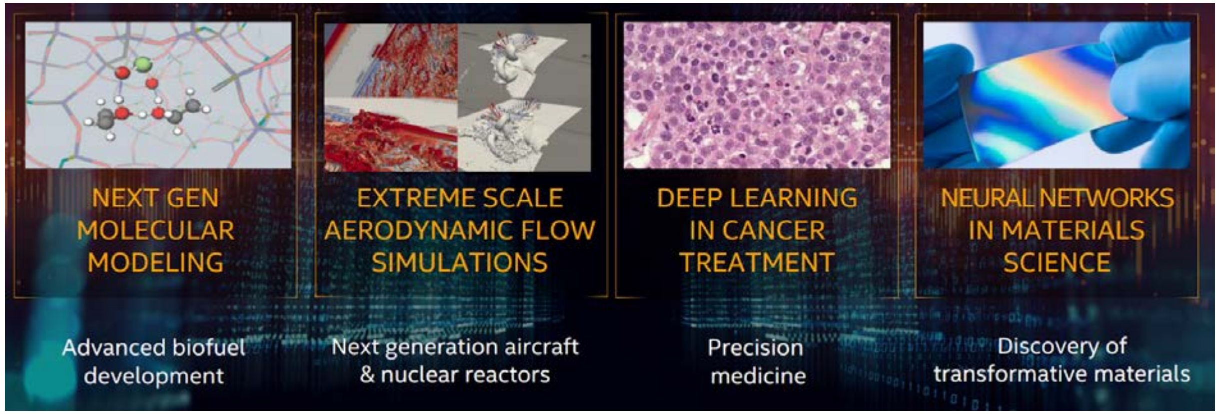
- Europe computational science is healthy
- In many area/domains Europe leads
- We have all sort of codes
  - CPU bound
  - Memory bound
  - Network bound
  - I/O bound
- But new challenges
  - DATA are coming to HPC (or the other way around…)
  - Complex often virtualized workflows
  - Interactive, on the fly processing, steering
  - Artificial Intelligence

Focus area for exascale co-design

JAPAN
Post-k



US (Aurora21)

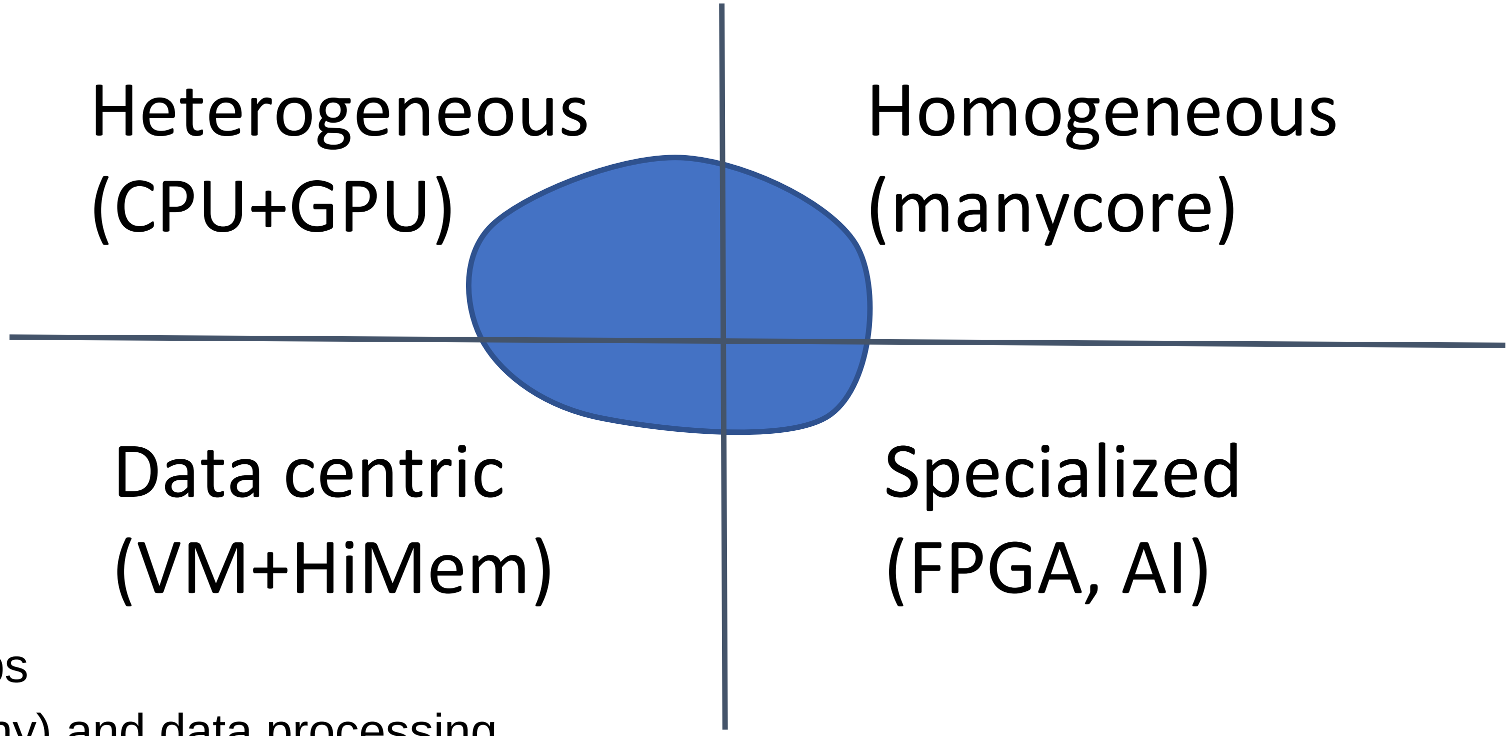


The Conundrum

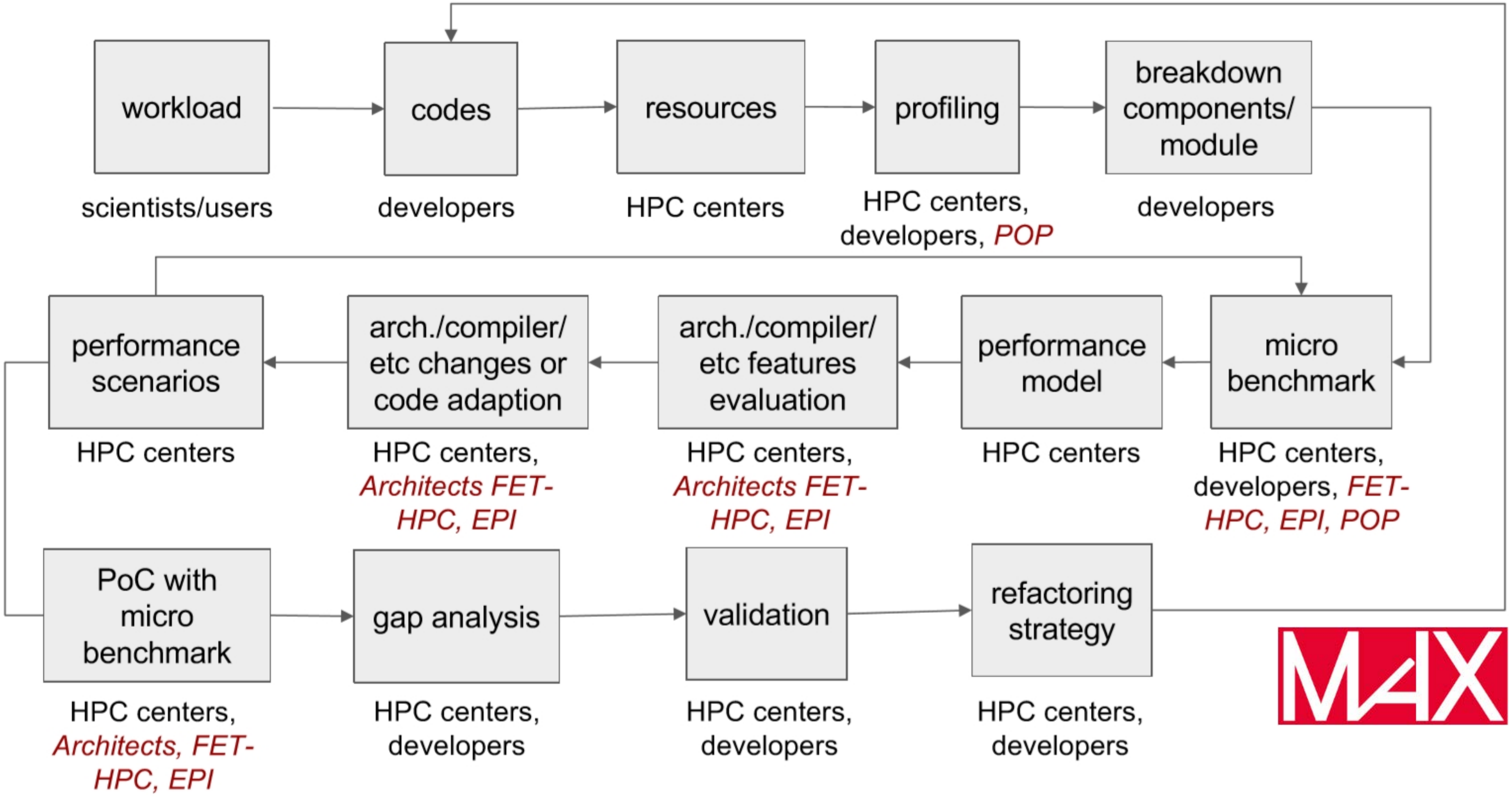
- The Scientific cases drive toward convergence
  - HPC + HPDA + AI
- Moore, Dennard, Amdahl and power constraints drive toward specialization
  - Tiny power efficient special execution units
  - Acceleration: GPU/FPGA
  - Little High Bandwidth Memory
  - Huge Non volatile memory

The opportunity Balancing hardware needs with cost-efficiency & usability

- HW
  - Heterogenous HW
  - Modularity
  - Network
  - Architecture
  - Balance Flops & Bytes
- SW
  - Less monolithic applications
  - Byte (Data) are as important as Flops
  - Workflow combining simulation (many) and data processing



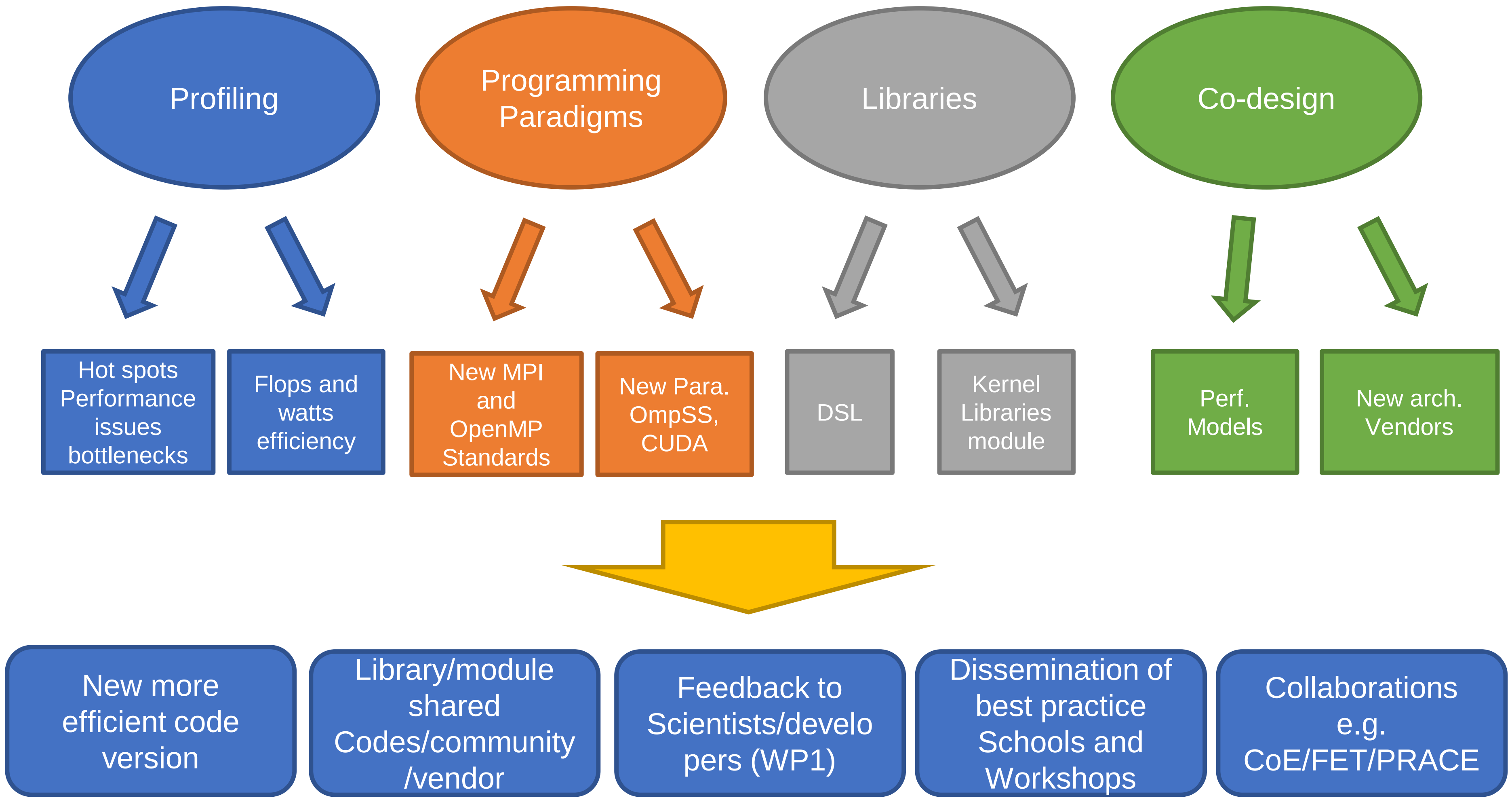
Co-design cycle



Conclusions

- Applications/Areas to focus on:
  - CoEs + Community efforts
  - Convergence between HPC, HPDA, AI
- Architectures requirement/constraints:
  - Network (switch/nic/topology/protocol) o primary importance
    - 400Gbit/node, Dragonfly, low latency protocol + eth
  - Memory -> less HBM/core ; more NVM/core
    - 100s MB/core, 10s GB/core
  - Integrate Accelerators of different kinds
    - EPI, GPUs, FPGA, Specialized devices
- Pilots:
  - At least one to integrate and validate EPI silicon
  - At least one to validate new accelerating technologies
  - At least one to validate network technologies and convergent solutions

Backup


Profiling
Programming Paradigms
Libraries
Co-design
Hot spots Performance issues bottlenecks
Flops and watts efficiency
New MPI and OpenMP Standards
New Para. OmpSS, CUDA
DSL
Kernel Libraries module
Perf. Models
New arch. Vendors
New more efficient code version
Library/module shared Codes/community /vendor
Feedback to Scientists/developers (WP1)
Dissemination of best practice Schools and Workshops
Collaborations e.g. CoE/FET/PRACE

Top 10 Challenges to Exascale

3 Hardware, 4 Software, 3 Algorithms/Math Related

- **Energy efficiency:**
  - Creating more energy efficient circuit, power, and cooling technologies.
  - > 50 Gflops per Watt
- **Interconnect technology:**
  - Increasing the performance and energy efficiency of data movement.
- **Memory Technology:**
  - Integrating advanced memory technologies to improve both capacity and bandwidth.
- **Scalable System Software:**
  - Developing scalable system software that is power and resilience aware.
- **Programming systems:**
  - Inventing new programming environments that express massive parallelism, data locality, and resilience
- **Data management:**
  - Creating data management software that can handle the volume, velocity and diversity of data that is anticipated.
- **Scientific productivity:**
  - Increasing the productivity of computational scientists with new software engineering tools and environments.
- **Exascale Algorithms:**
  - Reformulating science problems and refactoring their solution algorithms for exascale systems.
- **Algorithms for discovery, design, and decision:**
  - Facilitating mathematical optimization and uncertainty quantification for exascale discovery, design, and decision making.
- **Resilience and correctness:**
  - Ensuring correct scientific computation in face of faults, reproducibility, and algorithm verification challenges.

Common challenges

… beside performance

- Higher accuracy
- Larger systems
- Ensemble
- Data driven
- Validation and Verification
- Reproducibility
- Numerical stability
- Fault tolerance
- Provenance

What could be enabled (by exascale)

- Simulate formation of galaxies, neutron stars and black holes, predict how solar eruptions influence electronics, and model properties of elementary particles. This will explain the source of gamma-ray bursts in the universe, advance our understanding of general relativity, and help us advance quantum chromodynamics.
- Next-generation weather forecasting will provide local prognoses and longer lead times, with substantial financial impact
- Identify diseases that are caused by combinations of variants, with treatments tailored both to the patient and state of the disease.
- Improve the efficiency of nuclear power, hydropower, wind turbines, and not least batteries and high-voltage cables to enable transmission and storage.
- Enable direct numerical simulations of the Navier-Stokes equations for better accuracy and significant fuel savings.
- Integrate all aspects of design in models, use information from sensors deployed in an internet-of-things, include uncertainty quantification in predictions, and consider the entire life cycle.
- Fulfil the grand challenge of designing and manufacturing all aspects of a new material from scratch.
- Will enable computing to be applied in a whole range of non-traditional areas such as the humanities, social sciences, epidemiology, finance, promoting healthy living, determining return-on-investments for infrastructure.
- Fundamental research advances the state-of-the-art of scientific computing and helps attract new generations to Science, Technology, Engineering, and Mathematics.
- New algorithms that scale better and completely new approaches to solve mathematical problems in more energy efficient manners.