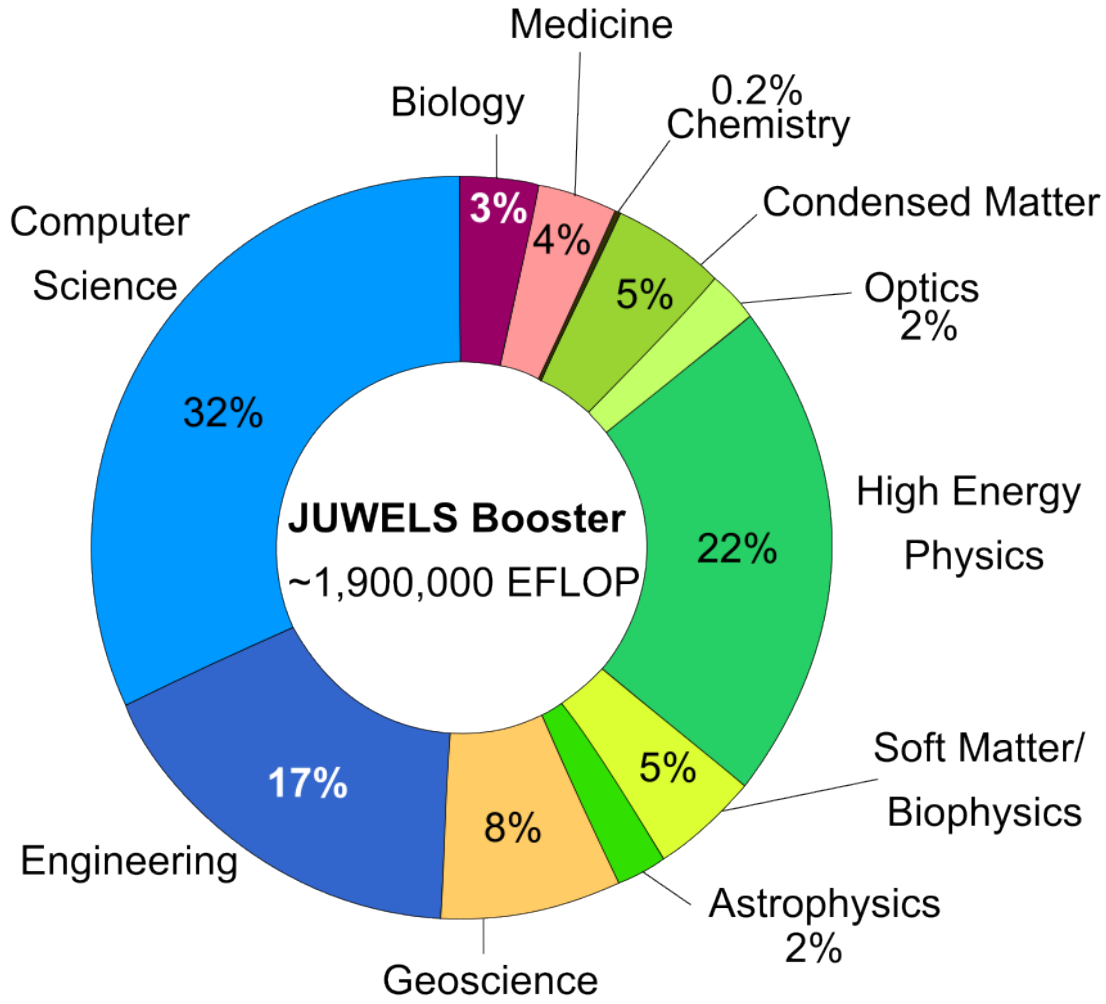
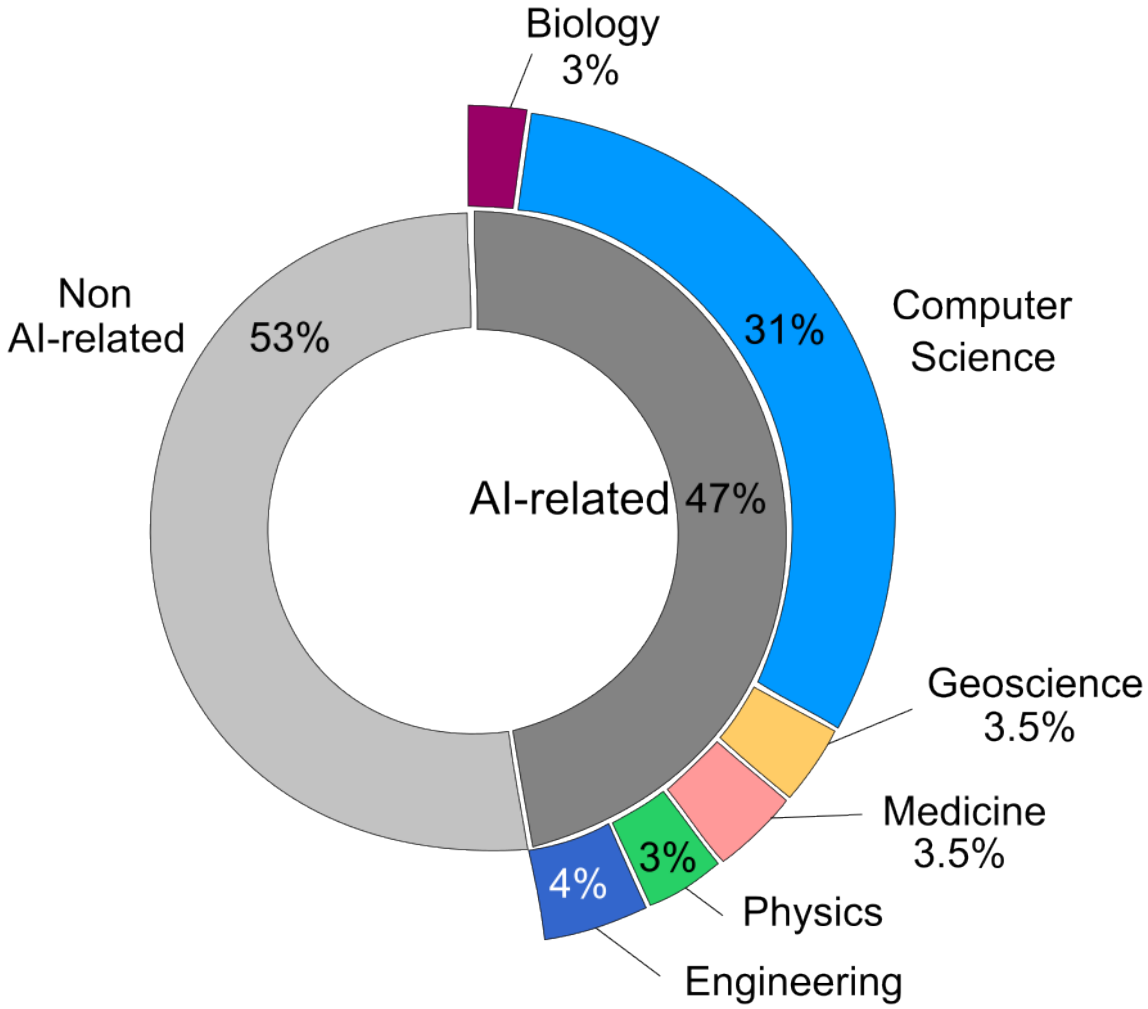
SIPEARL

Sustainable Computing Facilities through Industry Collaboration

Estela Suarez

Typical HPC user portfolio (example JSC)

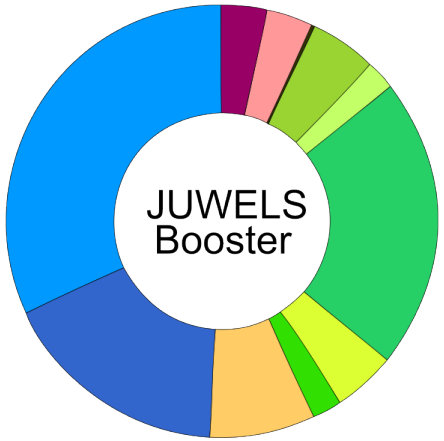




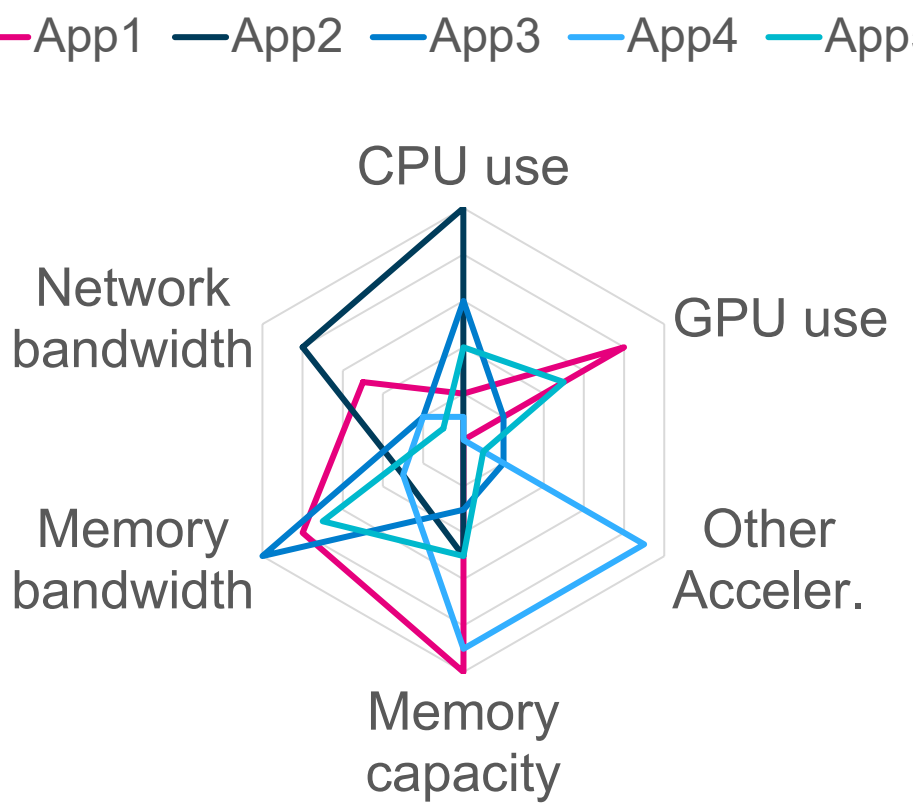
Data Source GCS/JSC -- Statistics: Nov 2024- Apr 2025

Data Source: F.Janetzko (FZJ-JSC)

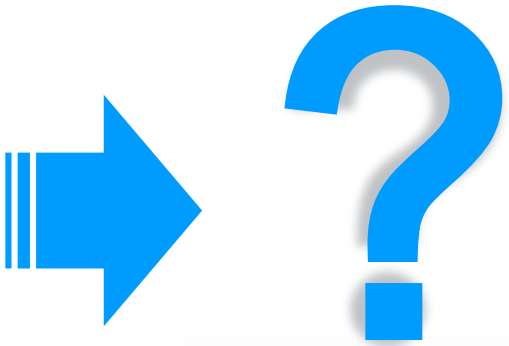
How to serve diverse requirements with one system?



Diverse Requirements

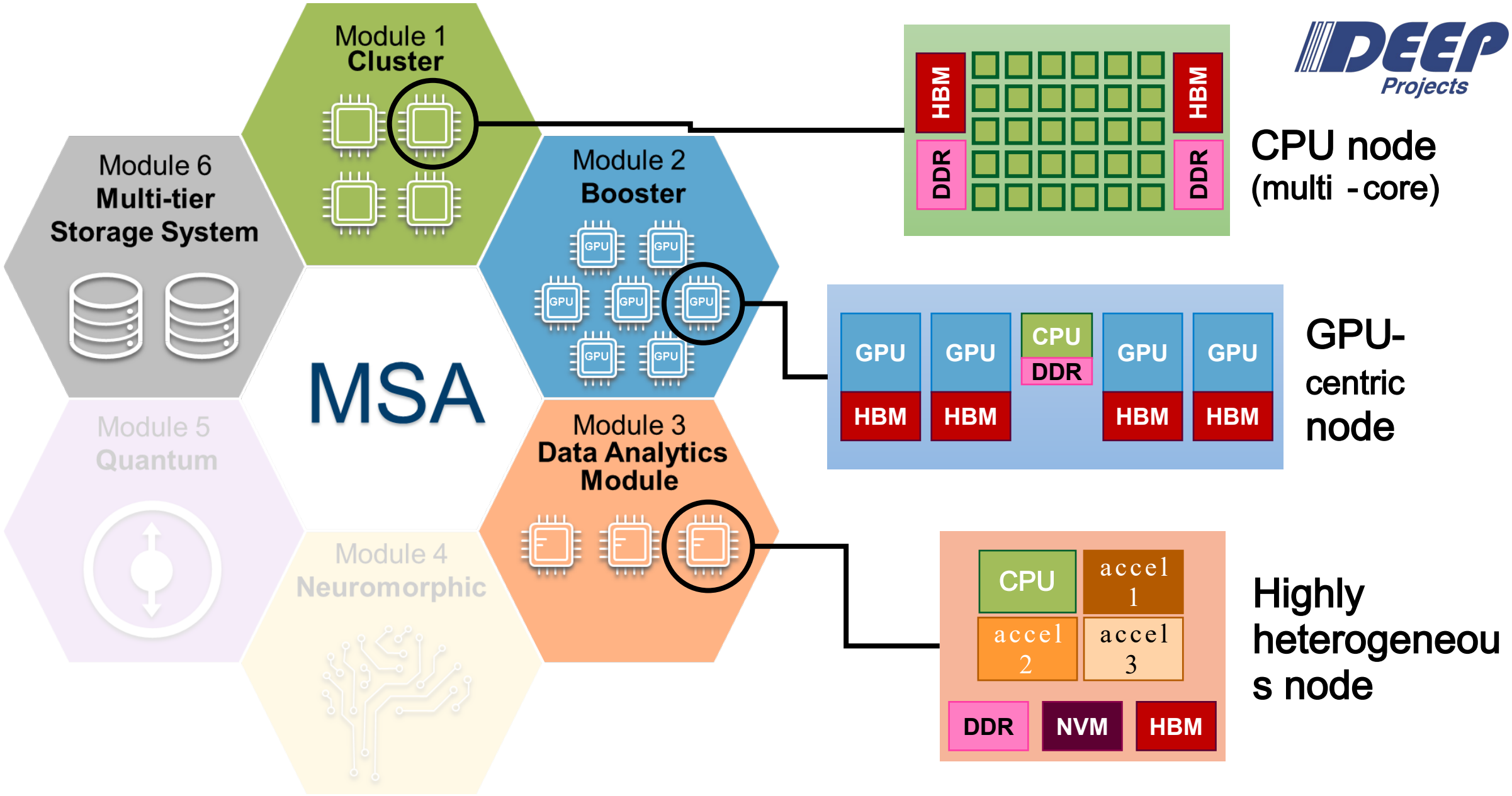


System design



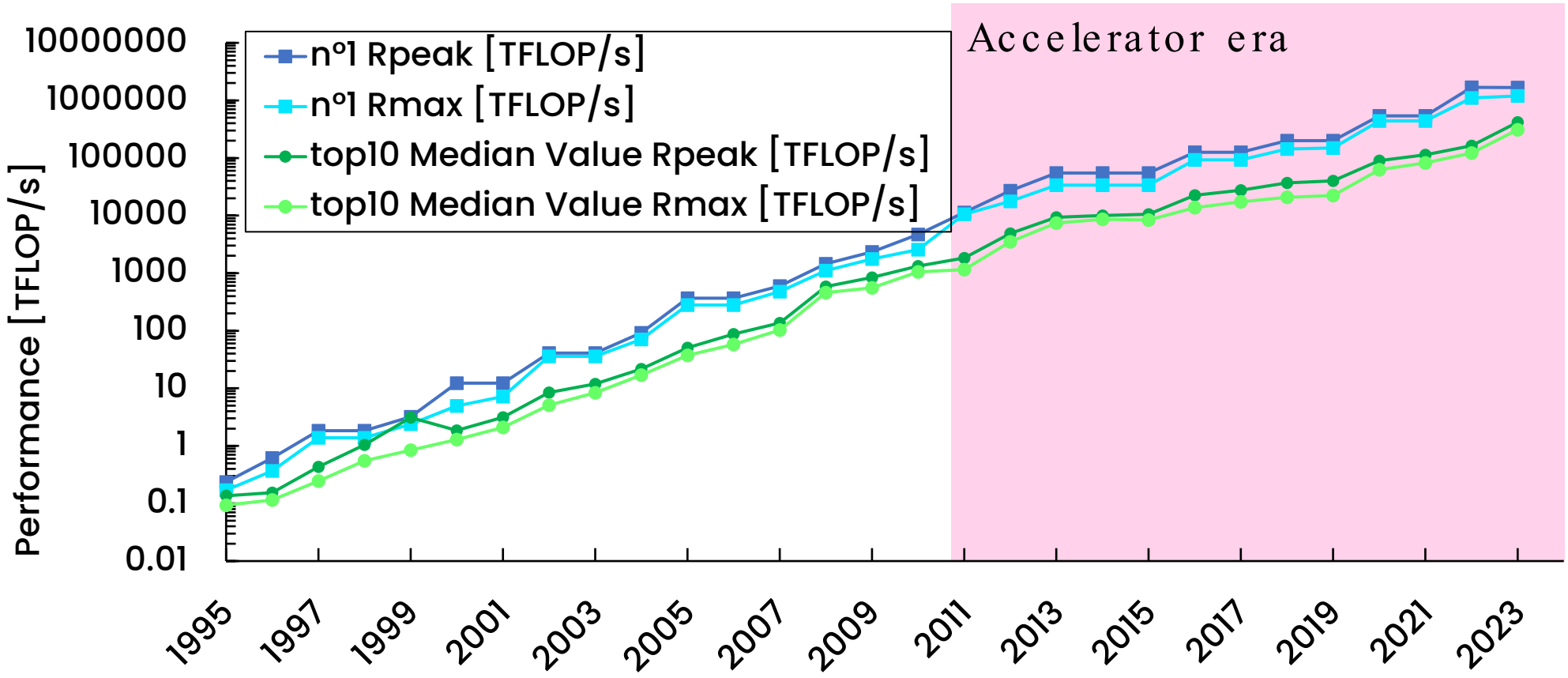
Using representative application benchmarks is key!

- M. Sato et al. 2022, *Co-Design and System for the Supercomputer "Fugaku"*, https://doi.org/10.1109/MM.2021.3136882
- A. Herten et al. 2024, *Application-Driven Exascale: The JUPITER Benchmark Suite* https://doi.org/10.1109/SC41406.2024.00038



- E. Suarez et al. *Modular Supercomputing Architecture – A Success Story of European R&D,* ETP4HPC White Paper. (2022) Available at https://doi.org/10.5281/ZENODO.6508394 .

High Performance Linpack (HPL) Performance



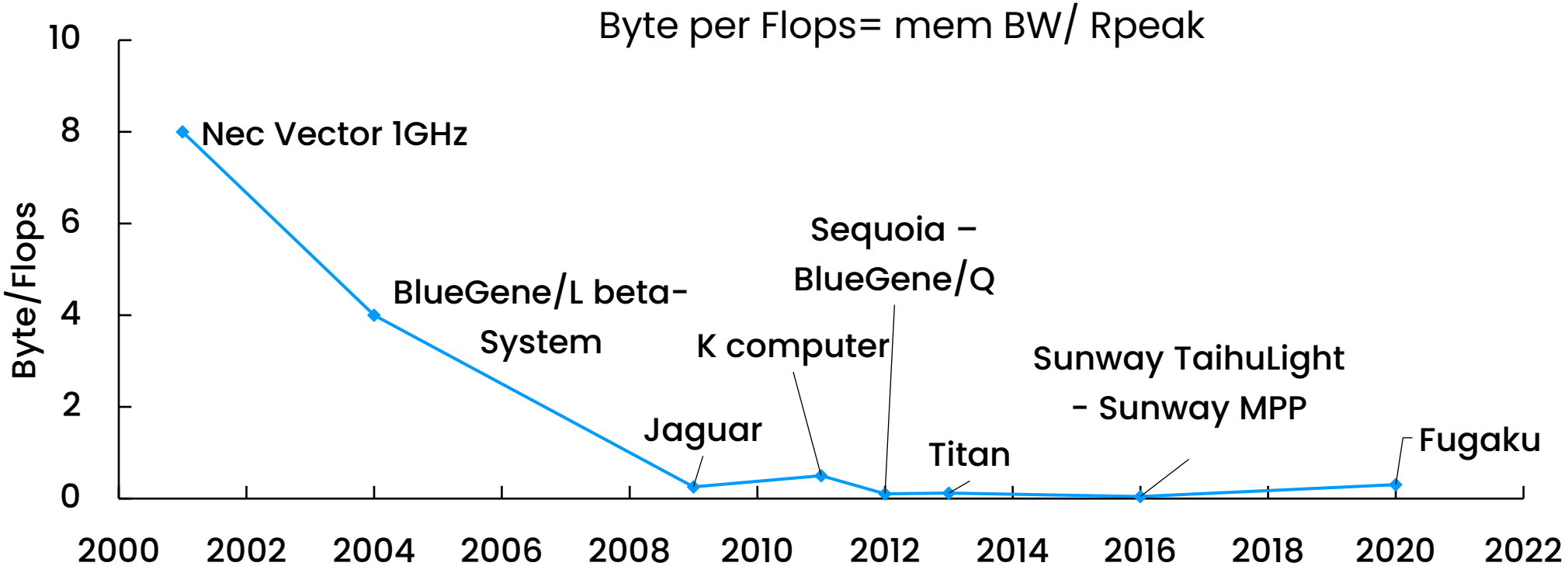
Data source: top500.org

GPUs brought significant performance gains

5

Byte/FLOP ratio

Consider compute core and memory bandwidth, key for application efficiency

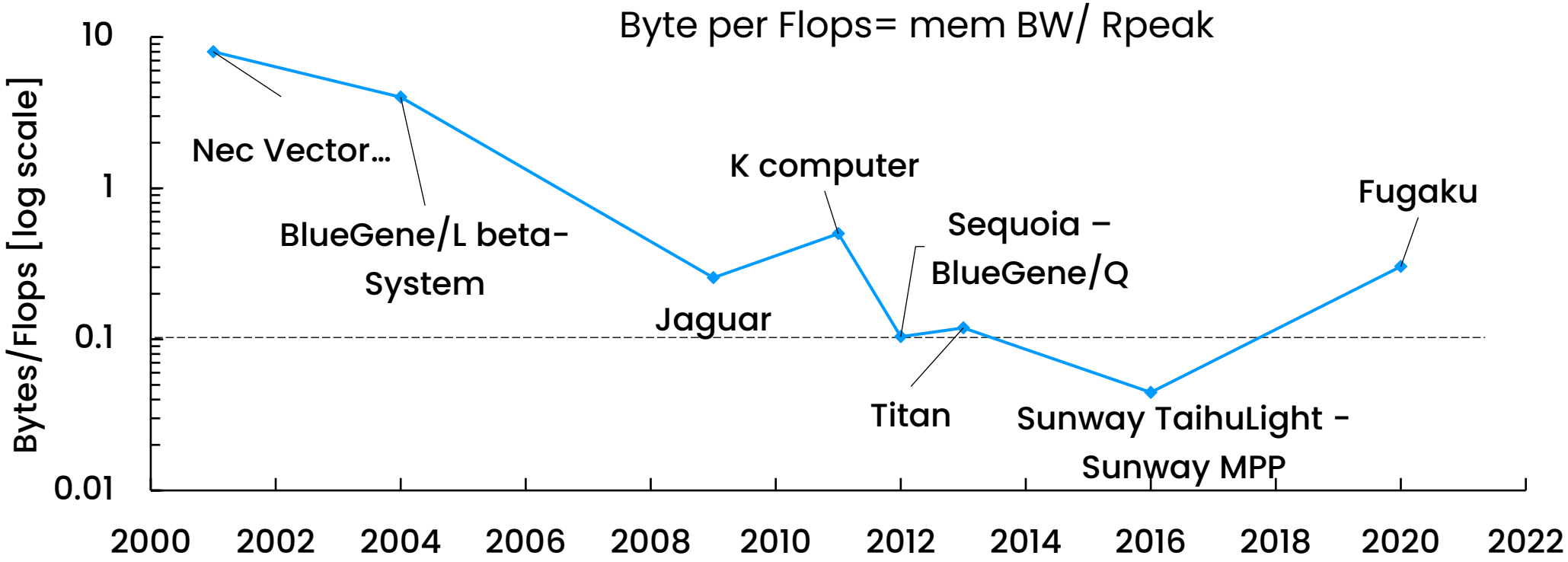


Byte per FLOP decrease rapidly.

Memory bound applications wait for data → Efficiency loss

Byte/FLOP ratio

Application efficiency requires a balanced design between memory transfers and compute power



Let’s build an efficient CPU for real applications!

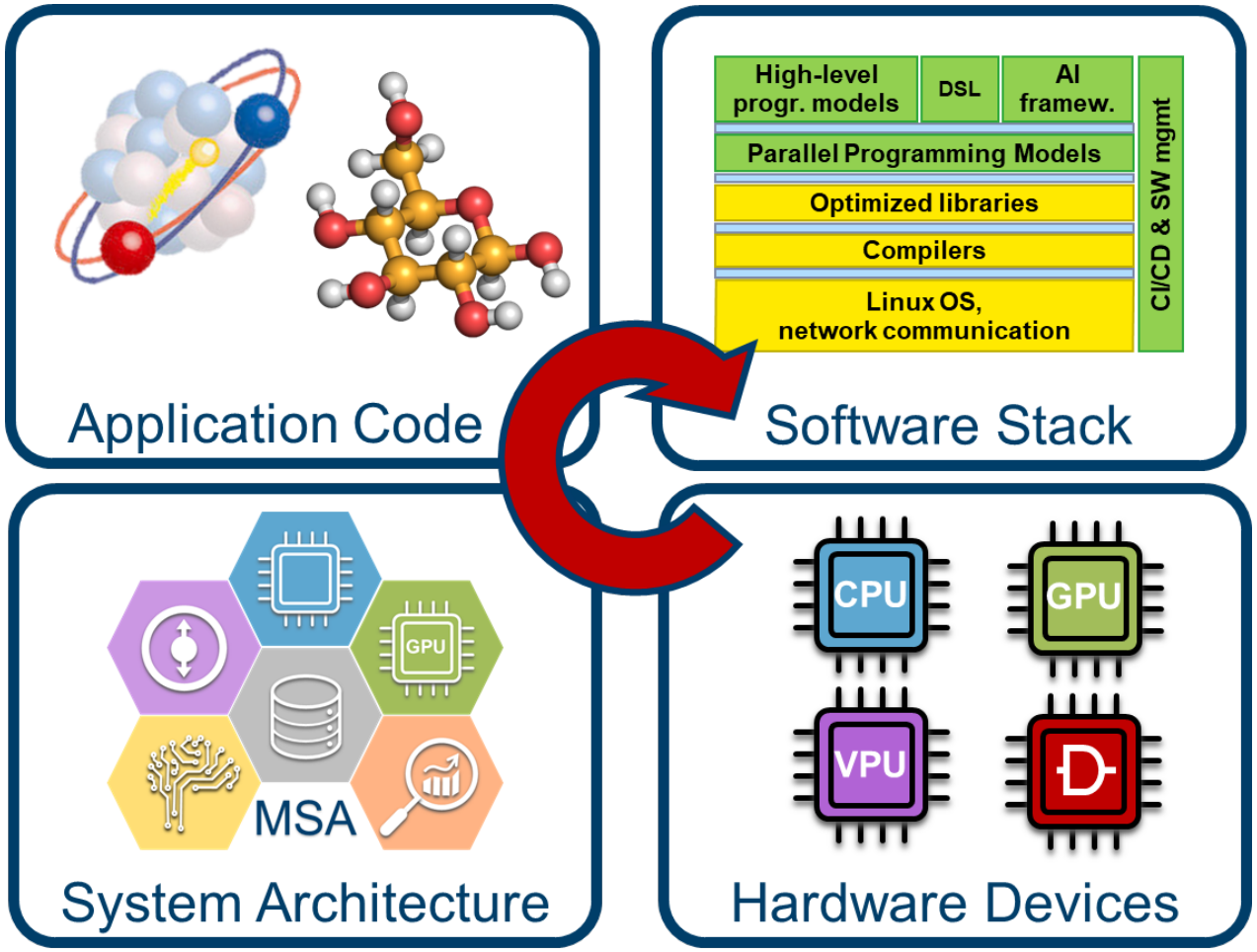
Co-design – my personal definition

Study interaction between

- application code,
- system software,
- hardware components,
- and system architecture

to find the configuration at each of those four levels that bring overall best

- performance and
- energy efficiency

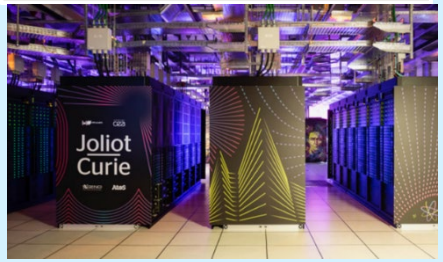


EPI: Co-design and Validation

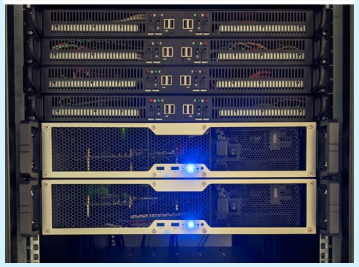
European Processor Initiative

Hardware platforms

Arm Reference Platforms

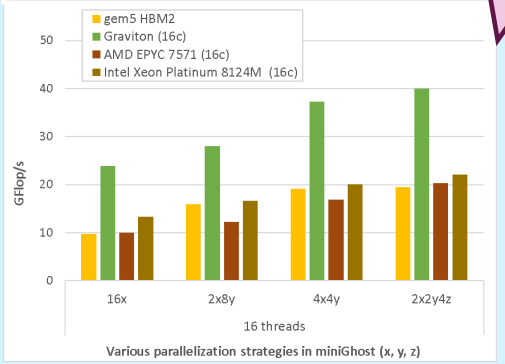


RISC-V SDVs

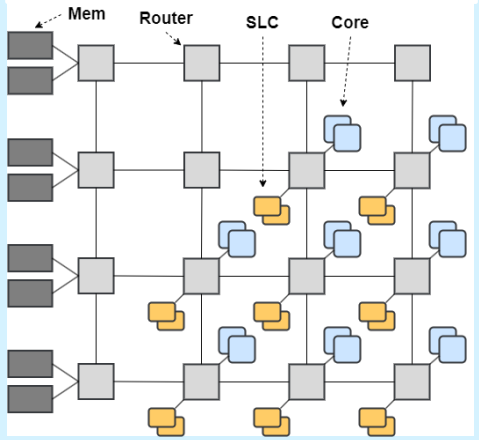


Validation and Co-design

Validation



Co-design

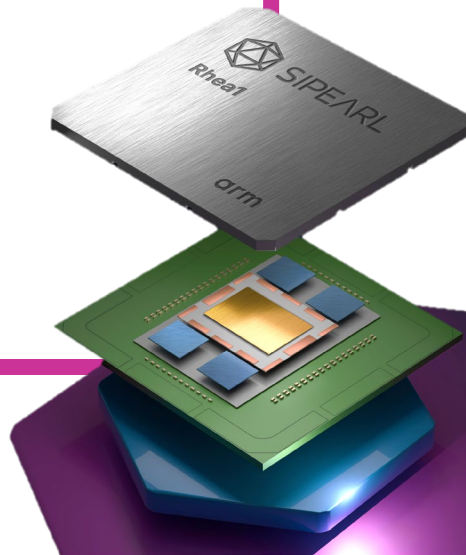


Multi-level:

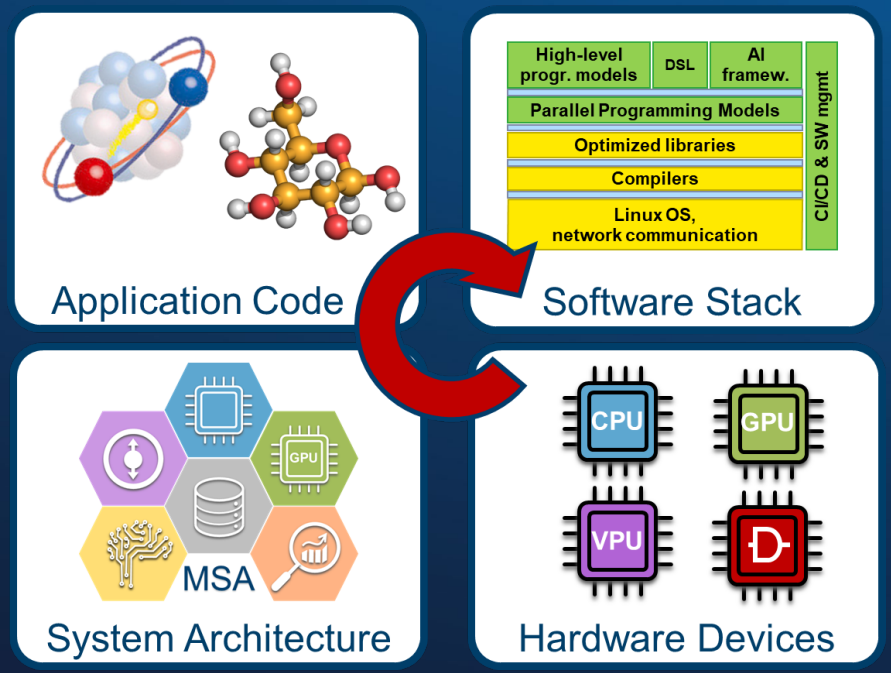
- benchmark suite
- models & simulators

Chip architects

Future Chips



Co-design at all levels



- System design ➔ Define configuration of each partition
- System integration ➔ Direct Liquid Cooling (DLC) coupling with data center cooling and waste-heat reuse
- Processor design ➔ Define the most efficient hardware for real-world applications (e.g. vector length, SIMD units, NOC layout)
- Operations ➔ Scheduling and resource management to maximize resource utilization
- Programming models ➔ Portability of applications on heterogeneous hardware (Caution!: portability vs. maximum performance)