HPCs, ATLAS Perspective

Andrej Filipcic
Jozef Stefan Institute, Ljubljana, SI

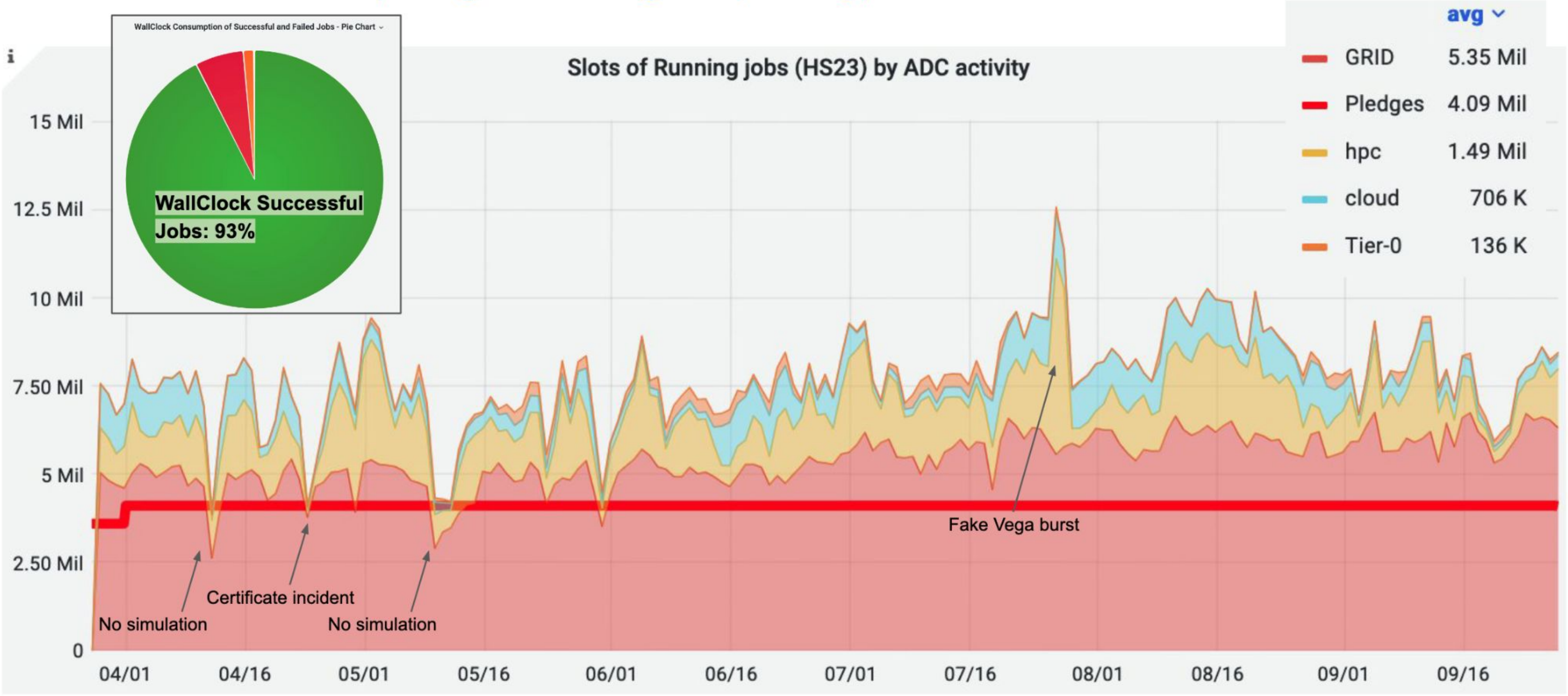
EuroHPC
Joint Undertaking

ATLAS & HPCs

- Used in production for 20 years
  - Mostly Nordic HPCs since the beginning
  - Distributed NDGF-T1 -> site managed data transfers: data preplaced/uploaded outside the running job
  - Several PRACE HPCs used for limited amount of time
  - CSCS PizDaint demonstrated for Tier0-like processing of B-stream triggers when CERN T0 was not sufficient
  - Experiment with Chinese HPCs: used 4 top HPCs for Geant4 for few years, not any more
  - US HPCs: Mira/Argonne for extensive Event Generation, Cori/NERSC and Titan/ORNL for Geant4 and more
- In most cases, usage was opportunistic through approved projects and limited in time and consumed CPU hours

ATLAS compute resources

^
- Grid sites providing quite a lot of efficient (and over-the-pledge) CPU
- HPC and cloud (Google and Vega especially) continue to contribute



Today

- Extensive usage of EuroHPC resources, mostly Vega/SI and Karolina/CZ
  - Vega runs all workflows including user analysis through PanDA
  - All HPCs included in central production system through Harvester and arcControlTower
  - Closed HPCs execute payload in fat containers, open HPCs use cvmfs
- Nordic resources: some stayed on HPCs, some migrated to national cloud infrastructure
- Other EU HPCs:
  - MareNostrum4/ES
  - CSCS PizDaint, ALPS
- NERSC Perlmutter, TACC Frontera
- UM6P/Morocco HPC, intentions to become Tier-2 in the future
- More expected next year, Leonardo, DE HPCs, …
- Some EU countries plan to provide significant WCLG pledges on HPCs

EuroHPC Joint Undertaking

OUR MEMBERS

- 33 participating countries
- The European Union (represented by the European Commission)
- 3 private partners

BDV BIG DATA VALUE ASSOCIATION

QuIC European Quantum Industry Consortium

ETP 4 HPC EUROPEAN TECHNOLOGY PLATFORM FOR HIGH PERFORMANCE COMPUTING

EuroHPC Machines

- 6 operational EuroHPC systems, all ranking among the world's most powerful supercomputers, in:
  - Slovenia
  - Luxembourg
  - Czechia
  - Bulgaria
  - Finland
  - Italy

- 4 EuroHPC systems are underway in:
  - Spain
  - Portugal
  - Germany
  - Greece

https://eurohpc-ju.europa.eu/supercomputers/our-supercomputers_en

Petascale systems

Vega

Sustained performance:	6,9 petaflops
CPU:	AMD Epyc Rome
GPU:	Nvidia A100
TOP500 ranking:	#32 in EU; #106 globally (June 2021)
Vendor/model	Atos BullSequana XH2000
Operated by	IZUM, Maribor, Slovenia

MeluXina

Sustained performance:	[illegible]
CPU:	AMD Epyc Rome
GPU:	Nvidia A100
TOP500 ranking:	#10 in EU; #36 globally (June 2021)
Vendor/model	Atos BullSequana XH2000
Operated by	LuxProvide [illegible]

Karolina

Sustained performance:	9,13 petaflops
CPU:	AMD Epyc Rome
GPU:	Nvidia A100
TOP500 ranking:	#20 in EU; #69 globally (June 2021)
Vendor/model	HPE Apollo 2000Gen10 Plus and Apollo 6500
Operated by	IT4I, Ostrava, Czech Republic

Discoverer

Sustained performance:	4,45 petaflops
CPU:	AMD Epyc Rome
GPU:	-
TOP500 ranking:	#27 in EU; #91 globally (June 2021)
Vendor/model	Atos BullSequana XH2000
Operated by	PSB consortium, Sofia, Bulgaria

Petascale systems in numbers

33.83 Petaflops sustained (47.19 Petaflops Rpeak)

- **11** partitions
- **3401** CPU Nodes
- **332** GPU Nodes
- FPGA, Visualisation and Cloud capabilities
- **24PB** Lustre Storage
- **6802** AMD EPYC Rome CPUs
- **1616** Nvidia A100 GPUs

EUROHPC AND QUANTUM

HPCQS

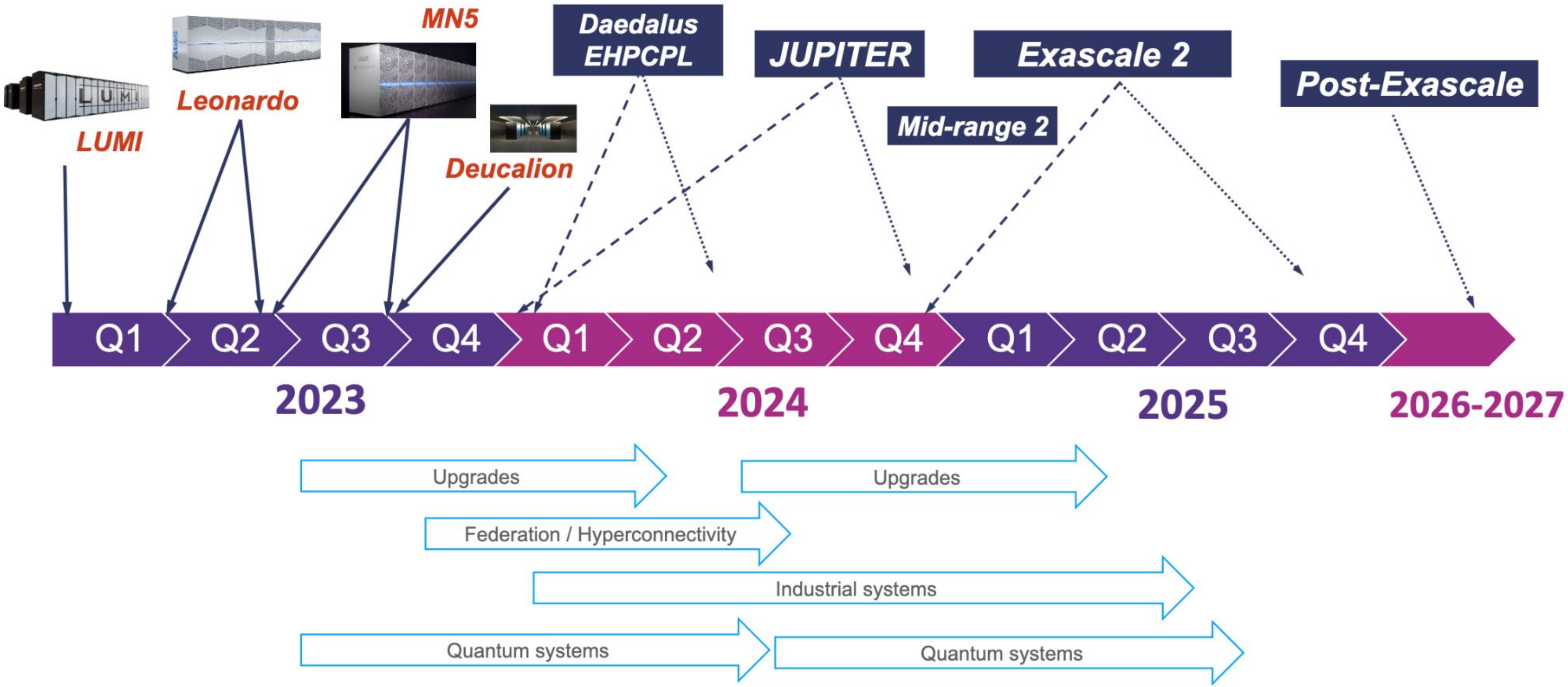
- The first EuroHPC initiative exploring quantum computing
- Launched in 2021 and running for 4 years
- HPCQS aims to integrate 2 quantum simulators, each controlling about 100+ qubits, into :
  - **Joliot Curie** (France)
  - **JUWELS** (Germany)
- French startup PASQAL will provide 2 Fresnel analog quantum simulators
- Incubator for **quantum-HPC hybrid computing**, unique in the world

PROCUREMENT OF QUANTUM COMPUTERS

- In October 2022, 6 sites were selected host and operate the first European quantum computers
- The selection includes IT4Innovations in Ostrava, CZ to host & operate LUMI-Q
- A diversity of quantum technologies and architectures is represented in this selection, giving European users access to many different quantum technologies

Is there any interest to explore quantum for generators?

EuroHPC Timeline



TOP 500 June 2023

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)	
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,699,904	1,194.00	1,679.82	22,703	0 + AMD/AMD
2	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899	ARM + 0/0
3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,220,288	309.10	428.70	6,016	AMD + AMD/AMD
4	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, Atos EuroHPC/CINECA Italy	1,824,768	238.70	304.47	7,404	Intel + Intel/NVIDIA
5	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148.60	200.79	10,096	0 + P9/NVIDIA
6	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94.64	125.71		0 + P9/NVIDIA
7	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93.01	125.44		Sunway + 0/0
8	**Perlmutter** - HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10, HPE DOE/SC/LBNL/NERSC United States	761,856	70.87	93.75		AMD + AMD/NVIDIA
9	**Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	555,520	63.46	79.22		0 + AMD/NVIDIA
10	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000, NUDT National Super Computer Center in Guangzhou China	4,981,760	61.44	100.68		Intel + 0/0

Existing HPCs and technologies

- Homogeneous HPCs are slowly going away, at least in EU
  - CPUs still needed by many users
  - Large increase in GPU demands, especially for AI
- Characteristics of EuroHPCs, 2-5 partitions
  - CPU partition: 100-200k cores (5M HS23), mostly AMD, some Intel (Leonardo, MN5), one AMD (Fujitsu)
    - Applications that have difficulties in porting to accelerators
    - Data intensive applications
    - Some have service partition for long lived user/group services (cloud like)
    - 2-4GB memory/core
    - Typically NO node local drive
  - GPU partition: NVIDIA in most cases, AMD MI250X on LUMI
    - 50-15k GPUs mostly A100, largely depending on investment (20-300M€)
  - Interconnect, Mellanox IB in most cases
  - Mostly Eviden (Atos), one HPE/Cray

Some EuroHPC guidelines

- Development/Benchmarking available an ALL EuroHPCs
  - Free of charge, no peer review, monthly calls, ~1M CPU , 50k GPU core hours on each HPC
  - Horizon eligibility criteria (except in France with embargo for some countries)
- In development, CI/CD for CoE (approved SW development), infrastructure could be used by others (HEP)
- Development support provided by HPC centers and upcoming EU EPICURE project
- 50% of HPC resources - EU Calls, long term allocations also possible
- National share - allocation through Hostiting Entity directly

Best practices on using HPCs

- Full node allocations
  - Some HPCs allocate cores, others only allow large jobs
  - Evgens should be at least multicore, possibly with MPI support - this can work on grid as well
  - Important to scale up to 1k cores in the future (per node)
- Jobs need to be shorter than 1 day, best is 6-12 hours
  - Faster turnaround for multi-user systems
- Most compute power will be on accelerators
  - Use it if available at runtime
  - API standardization still problematic, though apps like tensorflow are ported to NVIDIA and AMD - “easy” for users
- Effort in development should be spent on generators where HPC impact is significant. When grid is sufficient, there is no point to overdevelop
- Grid nodes are becoming very similar to HPC nodes in capacity, development would have much in common

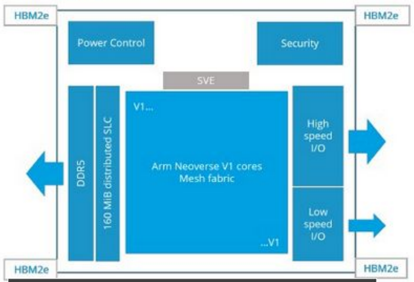
EuroHPC near future

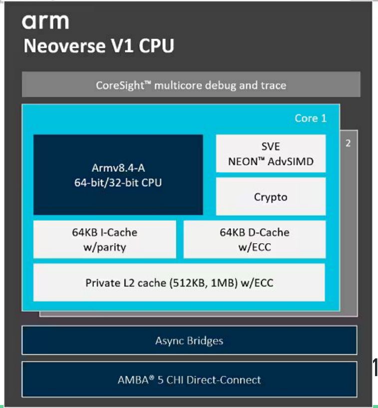
- JUPITER/Julich: 1st EU Exascale
  - EU Rhea-1 ARM64 CPU based on Neoverse V1 - SiPearl
  - NVIDIA G***
- Partnership with Japan on ARM CPU/GPU development
- Upcoming mid-range EuroHPCs (Greece, Sweden),
  - likely AMD/NVIDIA
  - Similar specs to others (1k CPU nodes, 200-300 GPU nodes)

With its high-performance energy-efficient Arm Neoverse V1 architecture, Rhea will meet the needs of all supercomputing workloads.

Core	- Arm Architecture Neoverse V1 cores - SVE 256 per core supporting 64/32/BF16 and int8 - Arm Virtualization Extensions
SoC	- Arm Mesh fabric - Advanced RAS support including Arm RAS extensions - Link protection for NOC and high-speed IO - ECC support for selected memory
Cache	- RAS supported for all Cache levels
Memory	ECC for memory and link protection for controllers - HBM2e - DDR-5
High Speed I/O	PCIe or CCIX/CXL: root and endpoint support
Other I/O	USB, GPIO, SPI, I²C...
Power Management	Power management block to optimize perf/watt across use cases and workloads.
Security Block Support	- Secure boot and secure upgrade - Crypto - True Random Number Generation

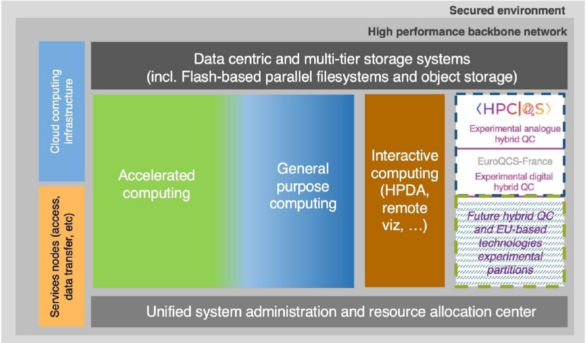
Block diagram



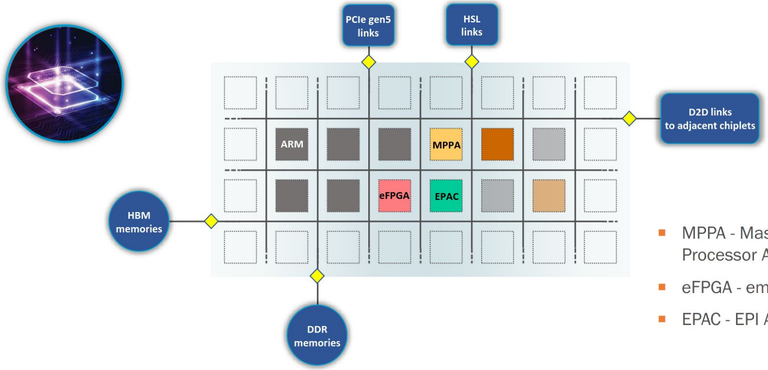


JULES VERNE/Genci (FR)

1. Fallback design, similar to Jupiter
   - Rhea-2 ARM CPU (designed finished)
   - GPU: NVIDIA or AMD or something else
2. Uniform partition with accelerated ARM, similar to NVIDIA Grace-Hopper
   - Unified memory
   - CXL enabled: virtual partitioning, eg CPU cores are allocated to CPU partition, accelerator to GPU partition - splitting compute, network, memory on the same node and allocating to different use cases
   - Estimated of 1.2M scalar cores and 20k GPUs



EPI - ARM Roadmap



Arm Research Summit 2019, 15/09/2019, Austin TX

- MPPA - Massively Parallel Processor Array
- eFPGA - embedded FPGA
- EPAC - EPI Accelerator

EPI - RISC-V

- The other EuroHPC option based on RISC-V
- Prototypes delivered already, though not even close to desired HPC performance
- Timeline for 1st chip in 2026/27
- Post Exascale EuroHPC is foreseen to be based on RISC-V, with fallback to ARM

HPC trends (in EU)

- Today: mostly x86_64 + NVIDIA, with some other similar architectures
  - Experimenting with FPGAs on some HPC centres
- 2024/25:
  - X86_64 might fade away, but will still be present for limited functionality, eg services, cloud infrastructure, management and monitoring, but limited compute power
  - ARM CPUs will likely get the largest share, especially if Jupiter HPC works well
  - NVIDIA and AMD GPUs will dominate
  - Development/testing partitions with exotic hardware on many EU HPCs (non-production hardware)
  - On CPUs, most performance will come from vector extensions (AVX512, SVE256) - vector parallelism is crucial
- 2026-27:
  - ARM with scalar and accelerated capabilities likely to become the standard (SiPearl, NVIDIA, Fujitsu/Fugaku2)
  - RISC-V might become production ready, too early to say
  - Power* might be important in US, but EU will likely push for its own CPUs
  - Quantum Computer on EuroHPC centers might be usable

Conclusions

- CPU architecture will be more diverse in the next 5 years, ARM taking a significant share, potentially with RISC-V
- Accelerated part will likely become embedded, CPU or GPU only chips will be deprecated
- ARM and RISC-V extensible architecture will support addons of FPGAs and other dedicated components (eg I/O ...)
- In EU, significant investment (7B€) in HPCs till 2027. Member states will likely build HPCs in cooperation with EuroHPC, less funding for national only HPCs
- At least in EU, fast development access to all HPCs resources is provided, also to prototype partitions with testing hardware
- It's time to think on quantum computing, it might be usable in 3-4 years.