Exascale + Neo Exascale: What's Next?

SC23
November 2023

www.HyperionResearch.com
www.hpcuserforum.com

Bob Sorensen

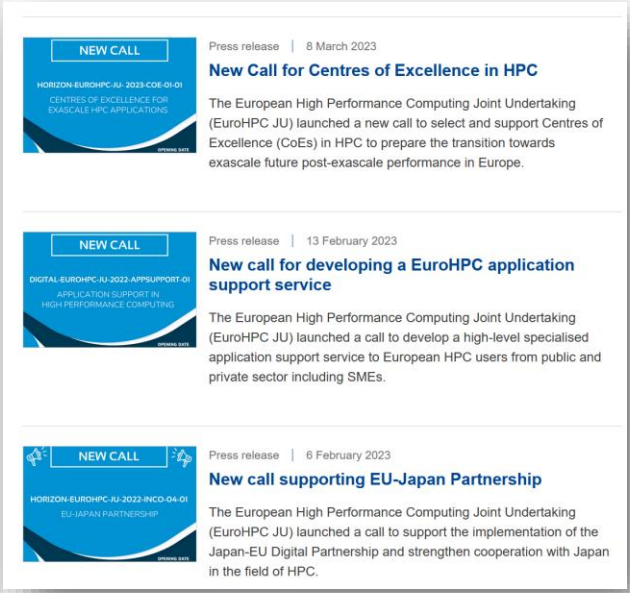
No Exascale Talk Complete Without the Top 500

On June 2023 list, US-5, EU-2, China-2, Japan-1

Rank	System/Site	# Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	Frontier - HPE Cray, DOE/SC/Oak Ridge National Laboratory, United States	8,699,904	1,194.00	1,679.82	22,703
2	Supercomputer Fugaku - RIKEN Center for Computational Science, Japan	7,630,848	442.01	537.21	29,899
3	LUMI - EuroHPC/CSC, Finland	2,220,288	309.1	428.7	6,016
4	Leonardo - EuroHPC/CINECA, Italy	1,824,768	238.7	304.47	7,404
5	Summit - DOE/SC/Oak Ridge National Laboratory, United States	2,414,592	148.6	200.79	10,096
6	Sierra - DOE/NNSA/LLNL, United States	1,572,480	94.64	125.71	7,438
7	Sunway TaihuLight - National Supercomputing Center in Wuxi, China	10,649,600	93.01	125.44	15,371
8	Perlmutter - DOE/SC/LBNL/NERSC, United States	761,856	70.87	93.75	2,589
9	Selene - NVIDIA Corporation, United States	555,520	63.46	79.22	2,646
10	Tianhe-2A - National Super Computer Center in Guangzhou, China	4,981,760	61.44	100.68	18,482

EuroHPC: Driving EU HPC Progress

The central focus of advanced computing in Europe



- Jointly funded by its 33 members with a budget of around EUR 7 billion between 2021-2027
- Develop, deploy, extend, and maintain in the EU a world-leading federated, secure and hyper-connected supercomputing, quantum computing, service and data infrastructure ecosystem

EuroHPC JU Activities and Plans

Varying workloads: varying architectures

- **An array of HPCs accessible to EU membership:**
  - **LUMI** in Finland (#3 on Top 500 list)
    - 375 Pflop/s sustained, 550Pflop/s TPP
    - Lumi-G (GPU centric),
    - Lumi-C (CPU only),
    - Lumi-D (Data Analytics: large memory),
    - Lumi-K partitions (containers, cloud services)
  - **LEONARDO** in Italy (#4 on Top 500 list)
    - 249 Pflop/s sustained, 323 Plop/s TPP
    - CPU partition (9 Pflop/s), GPU partition (240 Pflop/s)
  - **Vega** in Slovenia
    - 6.3 Pflop/s sustained, 10 Pflop/s TPP
    - CPU partition (960 nodes), GPU partition (60 nodes)
  - **MeluXina** in Luxembourg
    - 12.8 Plop/s, sustained, 18.3 Plop/s TPP
    - CPU partition (570 nodes), Accelerator Module (220 nodes)
  - **Discoverer** in Bulgaria
    - 4.5 Pflop/s sustained, 5.9 Pflop/s TPP
    - CPU only (1128 nodes)

EuroHPC JU Activities and Plans

Varying workloads: varying architectures

- **An array of HPCs in play:**
  - **Karolina** in the Czech Republic
    - 9.6 Pflop/s sustained 15.7 Pflop/s TPP
    - Four processing main partitions:
      - Standard numerical simulations: 720 servers
      - Accelerator: 70 servers with 8 GPUs per server
      - Large data set processing partition: 24 TB shared memory
      - Cloud service provider: 36 servers
  - **Deucalion** in Portugal
    - 7.22 Pflops sustained, 10 Pflops TPP
    - ARM partition (1632 nodes), x86 partition (500 nodes), accelerated partition (33 nodes)
  - **MareNostrum5** in Spain (still TBD)
    - 205 Pflops sustained, 314 Pflops TPP
    - GPP (general purpose partition), ACC (Accelerated partition,) NGT GPP (Next Gen Tech GPP partition), NGT ACC (Next Gen Tech ACC)

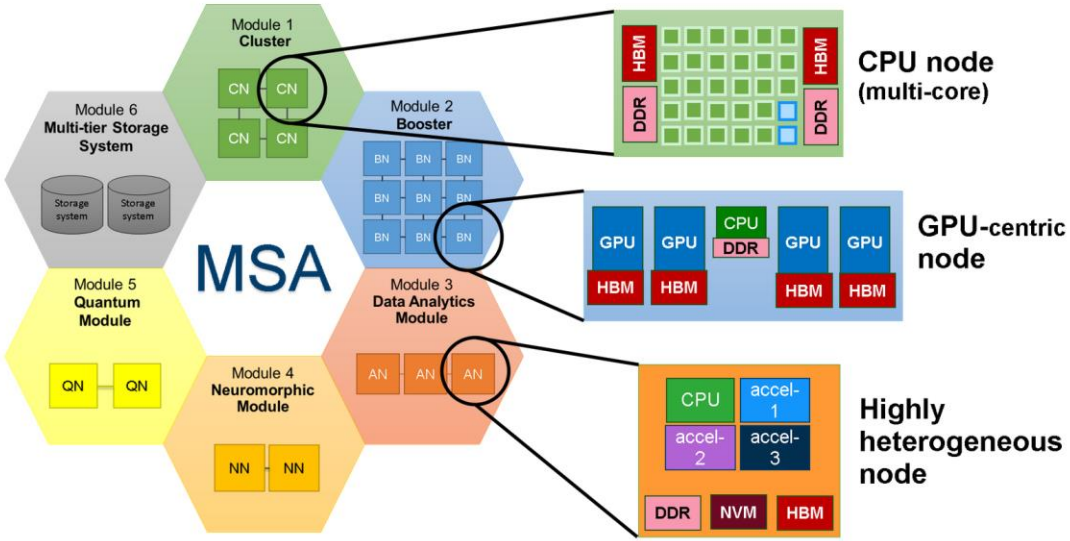
EU Plans for First Exascale System

JUPITER (Joint Undertaking Pioneer for Innovative and Transformative Exascale Research

- **Planned for 2024**
- **Installed at the Julich Supercomputing Center Germany**
  - Already hosts:
    - JUWELS Booster (#12), Module 1 (#93)
    - JURECA Data Centric Module (#61)
- **Average power is anticipated to be up to 15 megawatts**
- **Overall system and operation costs: 500 million euros**
    - 250 million euros: EuroHPC JU
    - 250 million euros in equal parts by the German Federal Ministry of Education and Research (BMBF) and the Ministry of Culture and Science of the State of North Rhine-Westphalia (MKW NRW)

EU Plans for First Exascale System

Is Jupiter the prototype for neo exascale systems?



https://www.fz-juelich.de/en/ias/jsc/about-us/structure/divisions/technology-division/next-gen-arch-proto/msa

- **Based on Eviden's Bull Sequana XH3000 technology**
- **Partnered with ParTec AG**
  - Specializes in the development and manufacture of modular supercomputers and quantum computers
- **SiPearl Arm-based Rhea Processors (fab TSMC))**
- **NVIDIA GPU (fab TSMC)**

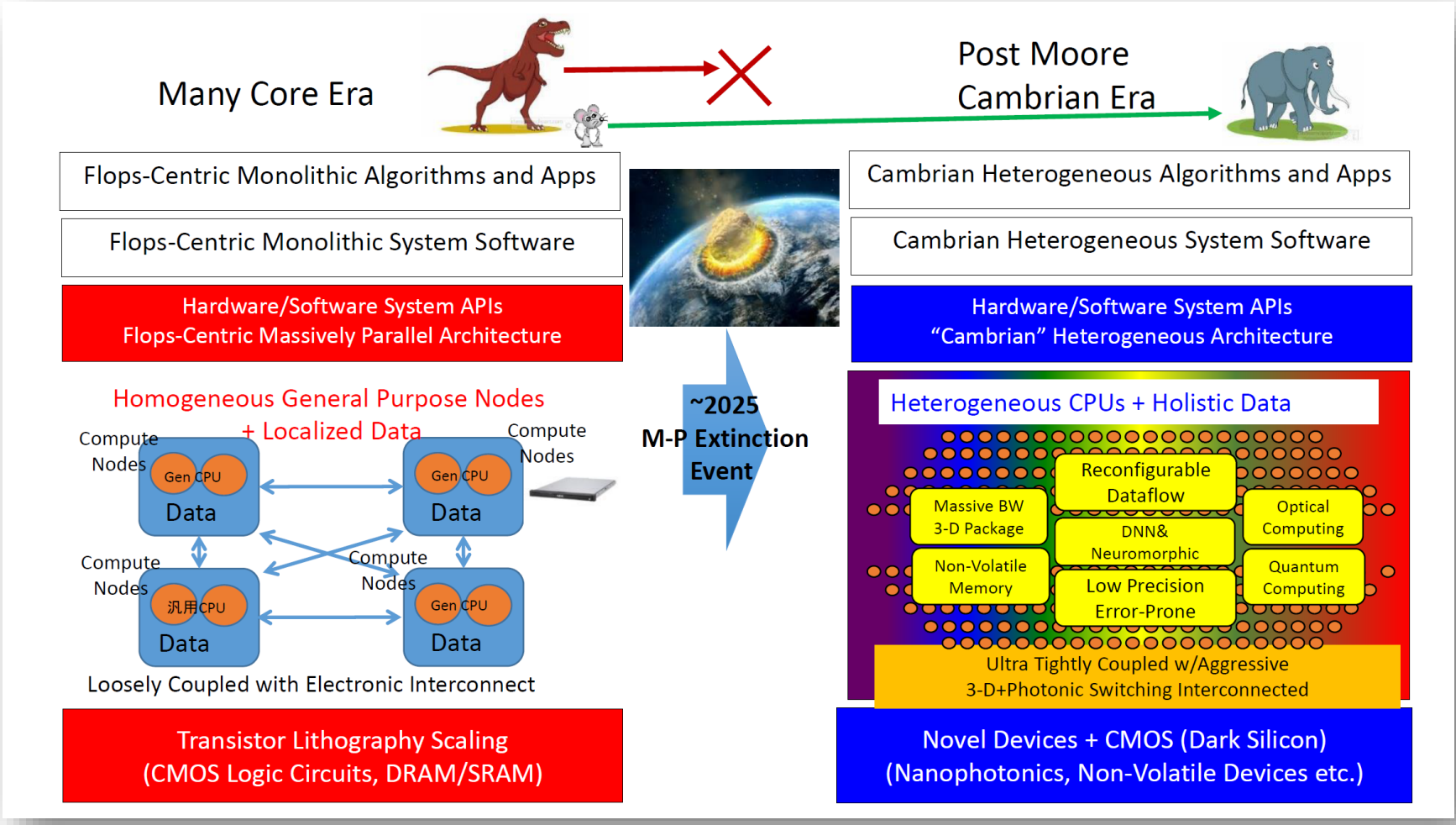
UK Plans for First Exascale System

50X performance of ARCHER2, UK's fastest system

- **Systems will be installed at EPCC's Advanced Computing Facility in Edinburgh**
- **Installation of the first phase is due to begin in 2025**
- **Part of a £900 million investment to upgrade the UK's next-generation compute capacity**
  - Speculations call for a £500 million budget for the HPC
- **More details to follow….**

Japan's Plan for 'Fugaku NEXT'

Not a straight-line projection



Satoshi Matsuoka, Director Riken RCCS ASCAC Presentation

China Exascale Status

- **No official announcements**
- **Last new Chinese appearance in Top 10 list was June 2018**
  - #7 Sunway TaihuLight in Wuxi (2016)
  - #10 Tianhe-2A at NUDT (2018)
  - Probably no new announcements this time
  - Most likely politically, not technologically, motivated
- **Strong evidence of at least five or more other Chinese systems that could make Top 10 list today**
- **Work, however, is being done**
  - Presentation on Wednesday: 5 ExaFlop/s HPL-MxP Benchmark with Linear Scalability on the 40-Million-Core Sunway Supercomputer

Rank	Site	Computer	Cores	HPL-AI (Eflop/s)	TOP500 Rank	HPL Rmax (Eflop/s)	Speedup
1	DOE/SC/ORNL	Frontier	8,730,112	9.9507	1	1.1940	8.3
2	EuroHPC/CSC	LUMI	2,174,976	2.168	3	0.3091	7.0
3	RIKEN	Fugaku	7,630,848	2.000	2	0.4420	4.5

Near-Term US Exascale Status

Three systems over two (or more) years with budget of ~ $1.8 billion

- **Frontier: DOE Office of Science: Oak Ridge National Laboratory**
  - First US exascale system in US
  - June Top 500 List: Rpeak = 1.68 Eflop/s, Rmax = 1.1 Eflop/s
  - 21 MW to run LINPAC
  - Cray Shasta with AMD EPYC CPU and AMD Radeon Instinct GPUs
  - Full user operations January 2023 (some delay)
- **Aurora: DOE Office of Science, Argonne National Laboratory**
  - **60MW**, ~ 1Eflop/s DP sustained, 2Eflop/s TPP
  - Cray Shasta architecture with Intel Xeons and Intel Xe GPU
  - On the list this week?
  - Delivery in late 2023, acceptance in 2024 (delayed at least X months)
- **El Capitan: DOE NNSA's LLNL**
  - ~ 2 Eflop/s
  - Cray Shasta architecture with AMD EPYC processors, next generation Radeon Instinct GPUs
  - Fully deployed in 2024

US Exascale Plans Going Forward

A new US Government procurement paradigm?

CHARTING A PATH IN A SHIFTING TECHNICAL AND GEOPOLITICAL LANDSCAPE:
POST-EXASCALE COMPUTING FOR THE NATIONAL NUCLEAR SECURITY ADMINISTRATION

FINDING 2.1: Semiconductor manufacturing is now largely in the hands of offshore vendors who may experience supply-chain risk; U.S. sources are lagging.

FINDING 2.2: All U.S. exascale systems are being produced by a single integrator, introducing both a technical and an economic risk.

FINDING 2.3: The joint Exascale Computing Project created a software stack for moving systems software and applications to exascale platforms, but although DOE has issued an initial call for proposals in 2023, there is not yet a plan to sustain it.

FINDING 2.4: Cloud providers are engaged in hardware and software innovations and will have more market influence in both technology and talent but are not aligned with NNSA requirements.

National Academies of Sciences, Engineering, and Medicine. 2023. *Charting a Path in a Shifting Technical and Geopolitical Landscape: Post-Exascale Computing for the National Nuclear Security Administration*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26916.

HYPERION RESEARCH

QUESTIONS?

bsorensen@hyperionres.com

Insufficient facts always invite danger.

- Spock, *Stardate: 3141.9.*