



**UK Research
and Innovation**

The UK Exascale Project Summary

6th September 2023

Professor Mark Parsons, Project Director
EPSRC Director of Research Computing



Exascale supercomputing

- Supercomputing advances by setting itself targets
- For more than 10 years the international supercomputing community has been targeting the Exascale
- An Exascale supercomputer is capable of 10^{18} (a billion billion) calculations per second – 1 Exaflop/s
- By way of comparison, ARCHER2 the current National Supercomputer is 50 times less powerful providing 20 Petaflop/s.
- We entered the Exascale age with the US's Frontier system in 2022
- A number of countries are now installing Exascale systems. The first European system will be in Germany at FZ Jülich in 2024/5.

The image shows a server rack with the word "FRONTIER" in large white letters. The letter "O" is replaced by a red and white compass rose. The background is a dark blue server rack with many cables and lights.

FRONTIER

Frontier is the new No 1 Top 500 system

- Frontier became first Exascale system on Top500 at ISC 2022 conference in Germany
- HPE Cray EX system with AMD Trento CPUs and MI250X GPUs
- Slingshot-11 network
- 1.194 Exaflop/s sustained on HPL
- 37,888 GPUs
- 606,208 cores
- 22.7 MW

Future of Compute review

- Review report published on 6th March “Tech Moment”
- Published alongside Science & Technology Framework
- Report makes 10 recommendations split into 3 themes
 - Theme 1: Unlock the world-leading, high-growth potential of UK compute
 - Theme 2: Build world-class, sustainable compute capabilities
 - Theme 3: Empower the compute community
- Only two recommendations focus on immediate investment – Exascale and AI Research resources



Future of Compute review

“Have a long term rolling 10-year strategy”

“Invest in infrastructure for compute and AI”

- Review report published 5th March
- Published alongside Science & Technology Framework
- Report makes 10 recommendations split into 3 themes
 - Theme 1: Unlock the world-leading, high-growth potential of UK compute
 - Theme 2: Build world-class, sustainable compute capabilities
 - Theme 3: Empower the compute community
- Only two recommendations focus on immediate investment – Exascale and AI Research resources

“Invest in software and skills”

Department for
Science, Innovation
& Technology

Independent report

Independent Review of The Future of Compute: Final report and recommendations

Updated 6 March 2023

Contents

Introduction from the expert panel

Report outline

List of recommendations

Proposed timeframe

Glossary of terms

1. The significance of compute for the UK
2. The international landscape of compute
3. The demand for compute in the UK
4. Meeting the UK's compute needs



Budget 2023 announcement

- 3.91 Further investment is needed in infrastructure for research and innovation. **Powerful computing capability is an essential component of being a global hub for innovation and achieving the UK's ambition to be a science superpower. Compute is also essential to progress in Artificial Intelligence research.** However, according to the independent Future of Compute Review, published last week, the UK's most powerful computer ranks just 28th in the world. The Review also found that the UK's AI community has immediate requirements for large-scale, accelerator-driven compute to remain internationally competitive.
- 3.92 In line with two of the key recommendations of the Future of Compute Review, the government will invest, subject to the usual business case processes, in the region of **£900 million to build an exascale supercomputer and to establish a new AI Research Resource**, with initial investments starting this year. Together, these will provide significant compute capacity to our AI community and provide scientists with access to cutting-edge computing power. They will allow researchers to better understand climate change, power the discovery of new drugs and maximise our potential in AI.

Exascale Requirements from Government

- System should support both **traditional Modelling & Simulation** and **Artificial Intelligence / Deep Learning** applications
 - Technology choices may be impacted by this
 - But future technologies blur the distinction
- System should support both **scientific user communities** and **industry users**
 - A greater focus is proposed for industry research use
 - Pay-per-use production access will also be supported
 - Specific support for SMEs
- Initial 250 Petaflop system should be **installed in 2024**

20% ringfenced
for industry
projects

EuroHPC systems

- Finland (CSC) is hosting Lumi
 - 375 Petaflops (HPL) / 550 Petaflops (Peak)
 - €145 million
 - Supplied by HPE
 - AMD EPYC CPUs + AMD GPUs
- Italy (CINECA) will host Leonardo
 - 249 Petaflops (HPL) / 324 Petaflops (Peak)
 - €120 million
 - Supplied by ATOS
 - Intel Icelake CPUs + NVIDIA A100 GPUs



Operational
2022

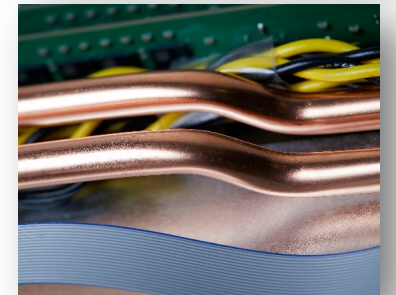
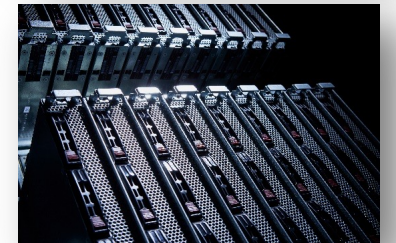


Operational
2023

UK
Researchers
can apply for
time on both
of these
systems

General design principles for UK Exascale Project

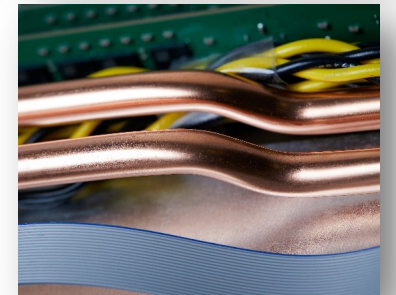
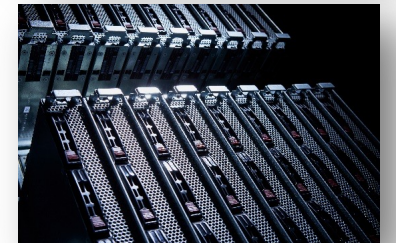
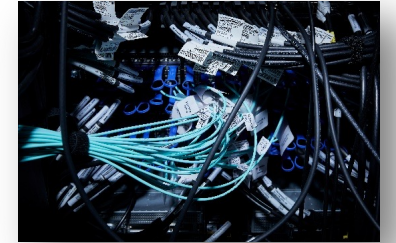
- Up to 25MW system + 5MW support and cooling
- Single tightly coupled system with two phases
- Compute power from GPU accelerated nodes (servers)
 - Target 1 Exaflop/s R_{MAX}
- Until 2026 accompanied by ARCHER2
 - Designed to provide attractive powerful resource for non-accelerated codes as they transition
- Large Software Programme envisaged
 - Multiple activities – Grand Challenge based to eCSE type activities
 - Lots of requirements gathering / consultation to do



General design principles for UK Exascale Project

- Up to 25MW system + 5MW supply
- Single tightly coupled system with
- Compute power from GPU accelerated nodes (servers)
 - Target 1 Exaflop/s R_{MAX}
- Until 2026 accompanied by ARCHER2
 - Designed to provide attractive powerful resource for non-accelerated codes as they transition
- Large Software Programme envisaged
 - Multiple activities – Grand Challenge based to eCSE type activities
 - Lots of requirements gathering / consultation to do

1 Exaflop/s on High Performance Linpack benchmark



Future of Compute Review – phased approach

- Phase 1: Immediately deliver hardware that supports a wide range of demands from research and business communities. This should provide at least 250 petaflops with enough performance and capacity to support current and future user requirements.
- Phase 2: Deliver hardware that has at least one exaflop of processing power ... This should be delivered no later than 2026, and within 2 years of phase 1 to maximise investment.
- Review proposed Phase 1 by 2024 and Phase 2 by 2026

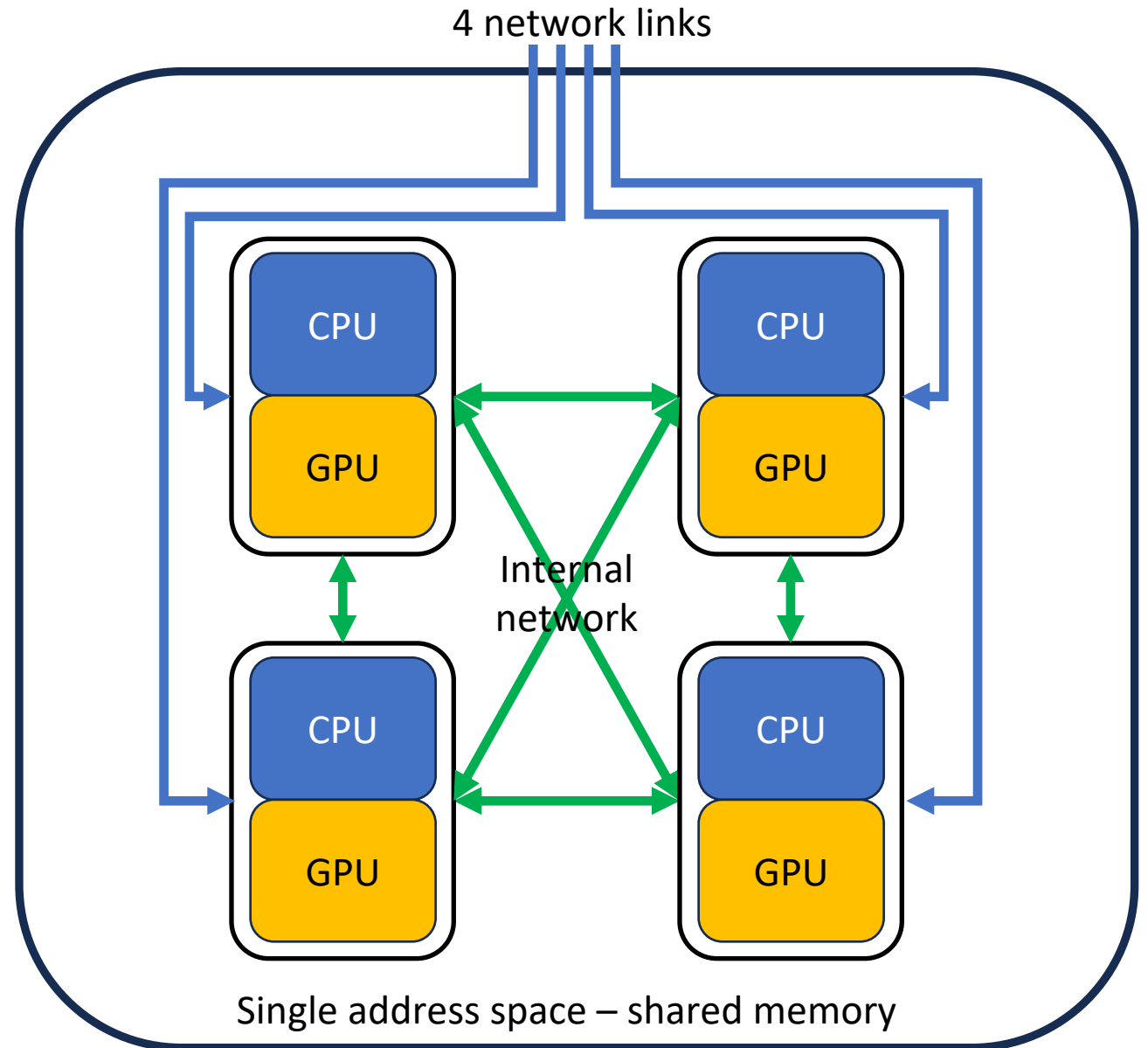
Exascale system parameters

- 1 Exaflop/s Top500 performance – in less than 25MW
- 6,000 nodes (approx.) – very similar to ARCHER2
- Each node must deliver 50X more performance than ARCHER2
- The route to Exascale is only possible using accelerators
- 1 node will have CPU cores + 4 GPUs
- Interconnect – 200 or 400Gbps
- In total the system will have circa. 24,000 GPUs

- **Key point: node level parallelism same as ARCHER2 – node performance 50X ARCHER2**

A “Typical” Exascale node

- 4 CPU/GPU modules
- Connected by very fast internal network
- 4 network links for interconnect topology
- Shared memory
- Key innovations
 - Very high-performance GPUs
 - Power performance ratio 10X better than ARCHER2
 - Single address space
 - Coherent shared memory
- Key challenges
 - NUMA effects
 - Programming models



Planned software programme

- UKRI delivered software programme for Exascale
 - Will benefit many parts of computational science ecosystem
- Outline plan - mixture of
 - Expanded eCSE programme – eCSE++
 - Longer software development projects for specific applications development – new applications / rewritten applications
 - Challenge driven software development – to put UK computational science at forefront internationally
- Keen to hear thoughts and ideas today

Summary

- The UK Exascale Project is being delivered by UKRI
- The system will be the new national HPC service
- A phased approach is being taken with full Exascale service planned by 2026 or 2027
- Building on the work of the ExCALIBUR programme and earlier EU-funded projects, a UK Exascale Software Programme is being developed by UKRI