EPCC Update and UK Exascale Project

HPC User Forum – May 2021

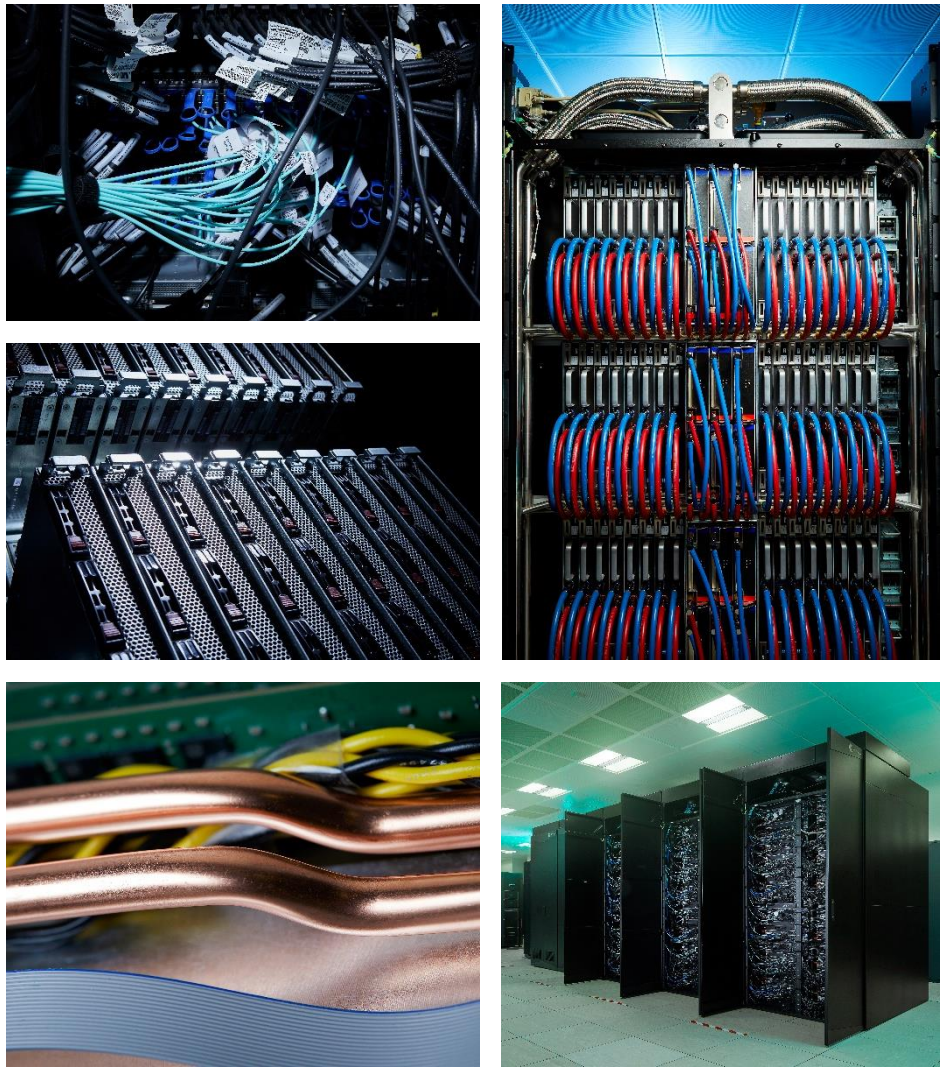
Professor Mark Parsons

EPCC Director
Associate Dean for e-Research

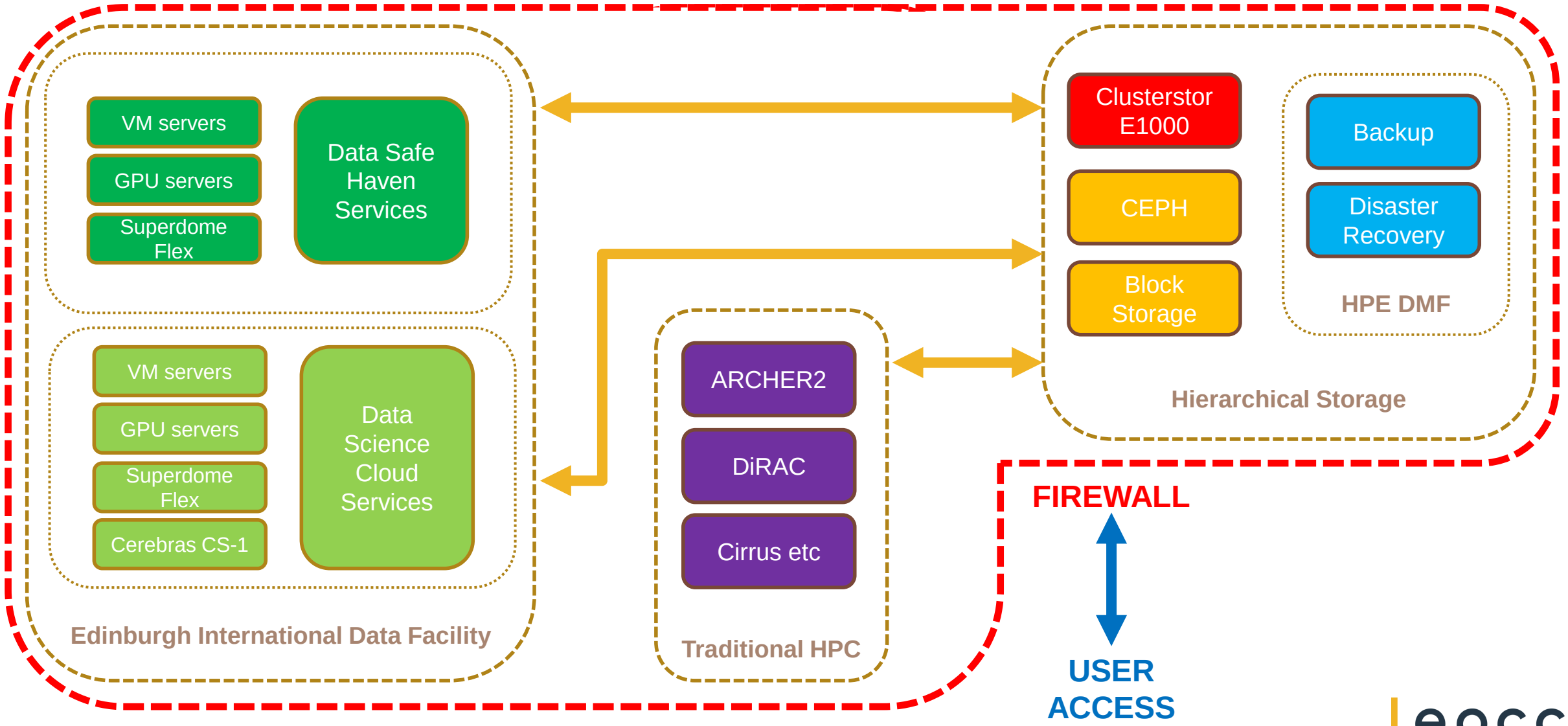
epcc

General EPCC update

- ARCHER2 has mostly arrived
- Cirrus Phase II running well
- Preparations at pace for hosting DiRAC Extreme Scaling service
  - ATOS Sequana XH2000 system
- New Computer Room 4 and new Data Centre Network completed
- £12m of Edinburgh International Data Facility equipment delivered and being installed

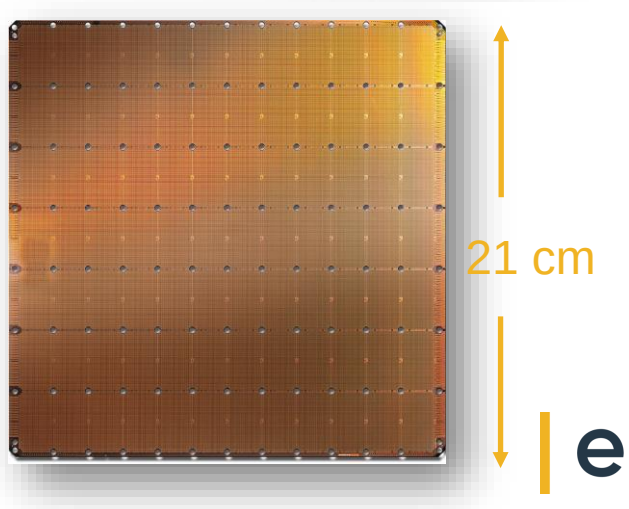


Edinburgh International Data Facility meets HPC



Finally … Cerebras CS-1 arrives at EPCC

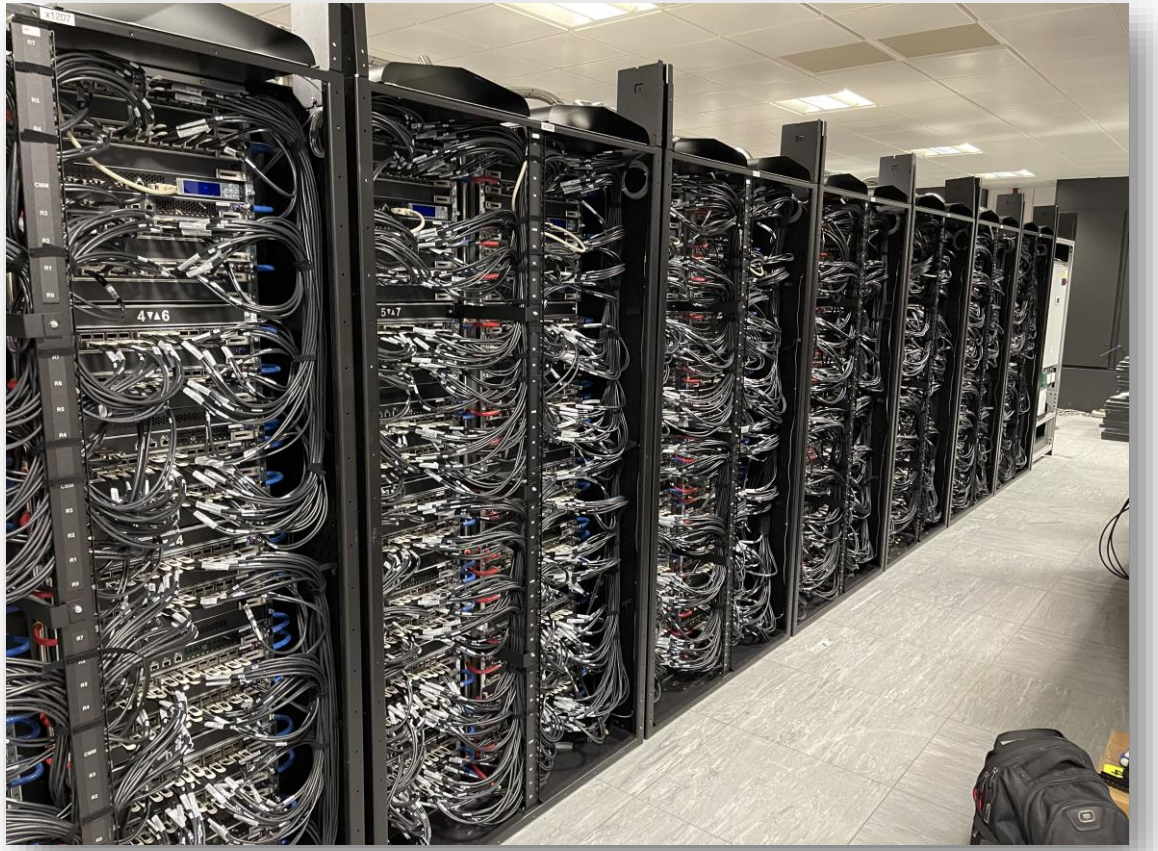
- EPCC has installed Europe's first Cerebras CS-1 – arrived March 2021
  - 400,000 AI cores optimised for sparse linear algebra – 1.2 trillion transistors
  - 18GB on-chip SRAM
  - 100 Pb/s internal interconnect with 1.2Tb/s ethernet connection to host
  - Focussed on TensorFlow and PyTorch etc
- Hosted within Edinburgh International Data Facility using 18TB SuperDome Flex
- Focussed on largest Deep Learning problems



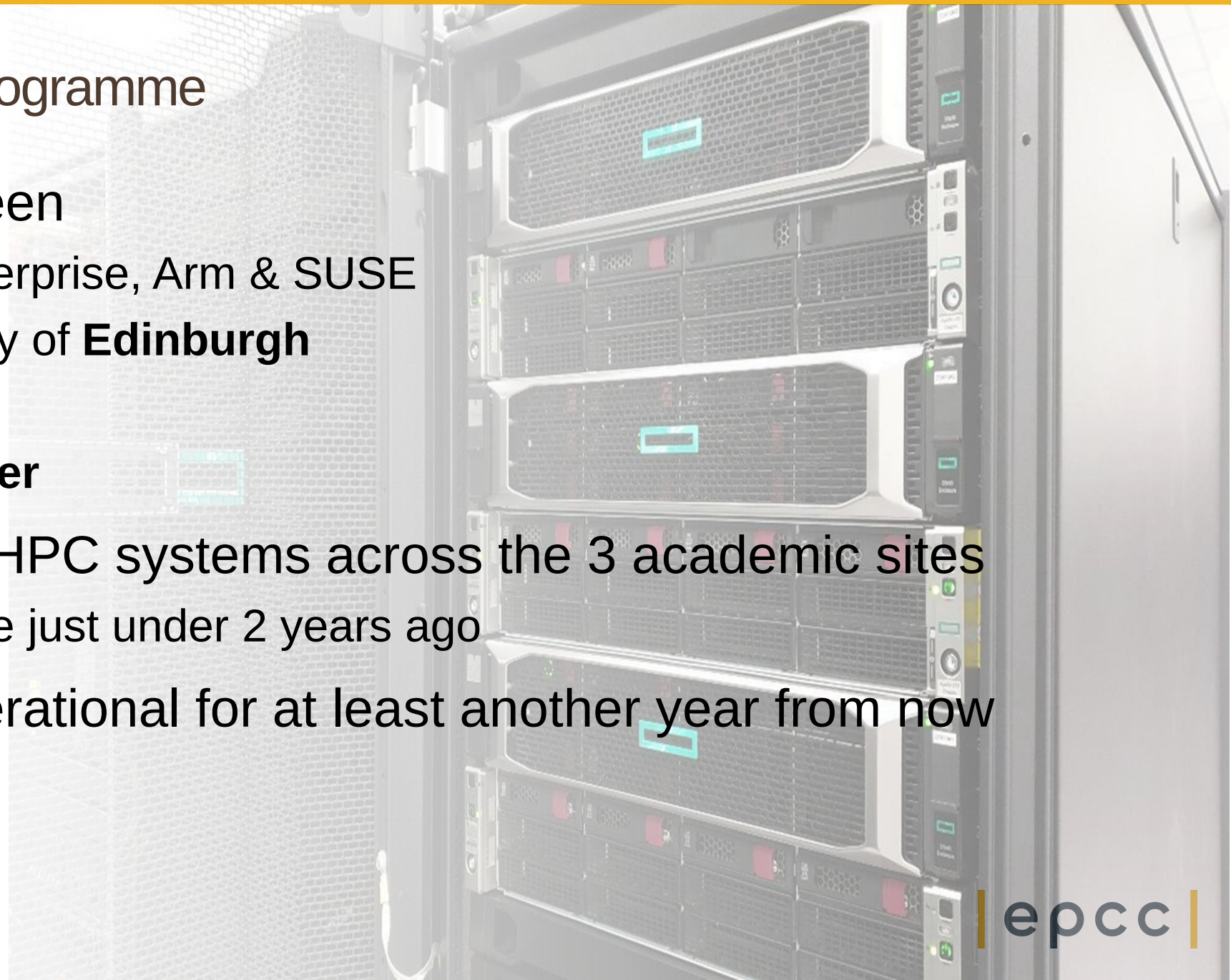
ARCHER2

- First 4 cabinets delivered in Summer 2020
- ARCHER 2 is an HPE Cray Shasta system
- 10x more powerful than ARCHER
- 748,544 AMD “Rome” cores in 5,848 compute nodes
- Each node has 2 x 64 core processors and 256GB of memory
- Remaining 19 cabinets delivered recently

ARCHER2 install – Day 4



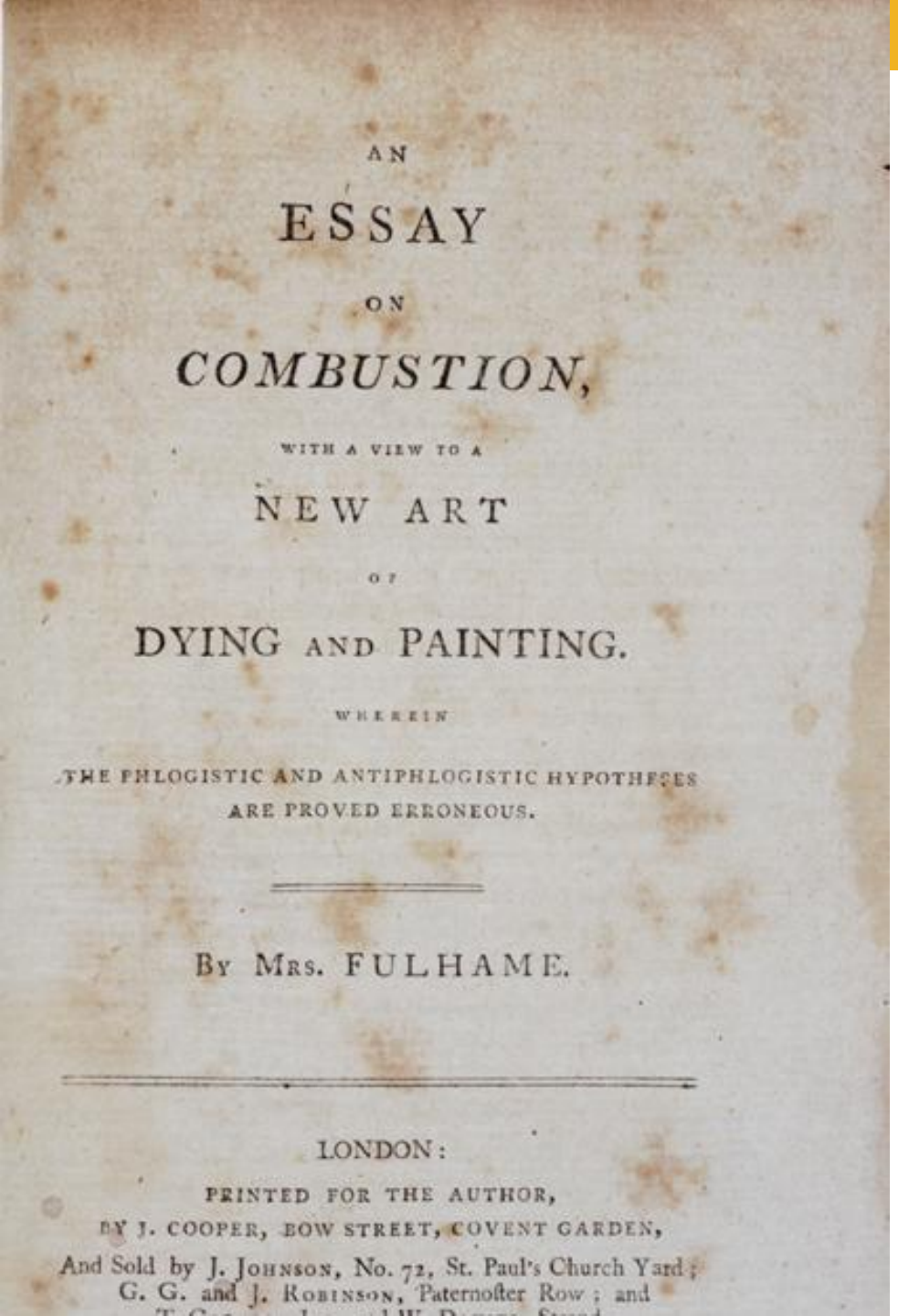
The CATALYST UK programme



- Collaboration between
  - Hewlett Packard Enterprise, Arm & SUSE
  - EPCC, The University of **Edinburgh**
  - University of **Bristol**
  - University of **Leicester**
- Deploy Arm-based HPC systems across the 3 academic sites
  - Installation took place just under 2 years ago
- Systems will be operational for at least another year from now

The HPE Catalyst Programme systems

- HPE Apollo 70
- The clusters at each university are largely identical
  - Designed, built and supported by HPE
- 64 nodes (i.e. 4096 cores) per cluster
  - 2 racks
  - Dual-socket nodes with Marvell ThunderX2 CPUs (32 cores per CPU)
  - 128GB (16 x 8GB) DDR4 per node
  - Mellanox EDR InfiniBand


AN
ESSAY
ON
COMBUSTION,
WITH A VIEW TO A
NEW ART
OF
DYING AND PAINTING.
WHEREIN
THE PHLOGISTIC AND ANTIPHLOGISTIC HYPOTHESES
ARE PROVED ERRONEOUS.
By Mrs. FULHAME.
LONDON:
PRINTED FOR THE AUTHOR,
BY J. COOPER, BOW STREET, COVENT GARDEN,
And Sold by J. Johnson, No. 72, St. Paul's Church Yard;
G. G. and J. Robinson, Paternoster Row; and

EPCC's system - Fulhame

- Fulhame was delivered and installed at the end of Nov'18
  - Named after **Elizabeth Fulhame** - 18th century Scottish chemist, first to describe the process of catalysis
- Our objectives for the system
  - Port applications
  - Try out different system software options
  - Measure performance
  - Report bugs and collaborate on resolving them

epcc

Original system setup

- SUSE Linux Enterprise (SLES) v12 SP3
- PBS Pro resource manager
- Small NFS storage, a mix of SSD and HDD

- The system worked very well out of the box – we were running a training course on the system less than a week after delivery!

Evaluating the Arm Ecosystem for High Performance Computing

Adrian Jackson
a.jackson@epcc.ed.ac.uk
EPCC, The University of Edinburgh
Edinburgh, United Kingdom

Andrew Turner
a.turner@epcc.ed.ac.uk
EPCC, The University of Edinburgh
Edinburgh, United Kingdom

Michèle Weiland
m.weiland@epcc.ed.ac.uk
EPCC, The University of Edinburgh
Edinburgh, United Kingdom

Nick Johnson
n.johnson@epcc.ed.ac.uk
EPCC, The University of Edinburgh
Edinburgh, United Kingdom

Olly Perks
olly.perks@arm.com
Arm
Warwick, United Kingdom

Mark Parsons
m.parsons@epcc.ed.ac.uk
EPCC, The University of Edinburgh
Edinburgh, United Kingdom

ABSTRACT

In recent years, Arm-based processors have arrived on the HPC scene, offering an alternative the existing status quo, which was largely dominated by x86 processors. In this paper, we evaluate the Arm ecosystem, both the hardware offering and the software stack that is available to users, by benchmarking a production HPC platform that uses Marvell's ThunderX2 processors. We investigate the performance of complex scientific applications across multiple nodes, and we also assess the maturity of the software stack and the ease of use from a users' perspective. This papers finds that the performance across our benchmarking applications is generally as good as, or better, than that of well-established platforms, and we can conclude from our experience that there are no major hurdles that might hinder wider adoption of this ecosystem within the HPC community.

CCS CONCEPTS

• **General and reference** → *Performance*; • **Computing methodologies** → *Massively parallel algorithms*; *Distributed programming languages.*

KEYWORDS

Benchmarking, Arm, ThunderX2, Marvell, Performance, Distributed Processing

ACM Reference Format:
Adrian Jackson, Andrew Turner, Michèle Weiland, Nick Johnson, Olly Perks, and Mark Parsons. 2019. Evaluating the Arm Ecosystem for High Performance Computing. In *Proceedings of the Platform for Advanced Scientific Computing Conference (PASC '19), June 12–14, 2019, Zurich, Switzerland.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3324989.3325722

PASC '19, June 12–14, 2019, Zurich, Switzerland

ACM ISBN 978-1-4503-6770-7/19/06…$15.00
https://doi.org/10.1145/3324989.3325722

1 INTRODUCTION

There is an established ecosystem of server-level processors suitable for computational simulation and machine learning applications built around traditional x86 architectures from processor manufacturers such as Intel and AMD. However, recently, alternative processor technologies have been developed; foremost amongst these are Arm based processors from manufacturers such as Marvell (ThunderX2), Ampere (eMAG), Huawei (Kunpeng 920), Fujitsu (A64FX) and Amazon (Graviton).

The first server class processor to be commercially available in large volume is the ThunderX2 processor from Marvell. The ThunderX2 processor uses the Armv8 instruction set and it has been designed specifically for server workloads. The design includes eight DDR4 memory channels to deliver measured STREAM triad memory bandwidth in excess of 220 GB/s per dual-socket node.

However, hardware only represents one part of the ecosystem that is required to deliver a usable High Performance Computing (HPC) platform for the varied workloads of computational simulation and machine learning applications. Operating system, compiler, and library support is required to provide a functional environment that supports large scale HPC applications and to ensure applications can both be easily ported to such new hardware as well as efficiently exploit it.

In this paper we will evaluate a range of computational simulation applications on a HPC system comprised of nodes with ThunderX2 processors connected together with an Infiniband network. Our paper makes the following contributions to deepening the understanding of the performance of a production HPC system that is based on the Arm ecosystem:

(1) We outline performance measurements of the interconnect network, using established MPI benchmarks, allowing us to assess the potential scaling performance of distributed memory applications.
(2) We present and evaluate the multi-node performance of scientific applications with varying performance characteristics and compare it to the established x86 ecosystem.
(3) We evaluate the portability of applications onto this new system, compared with equivalent systems based on other processor technologies.
(4) We discuss the causes for the performance and scalability results that have been observed, and based on this we draw

First results

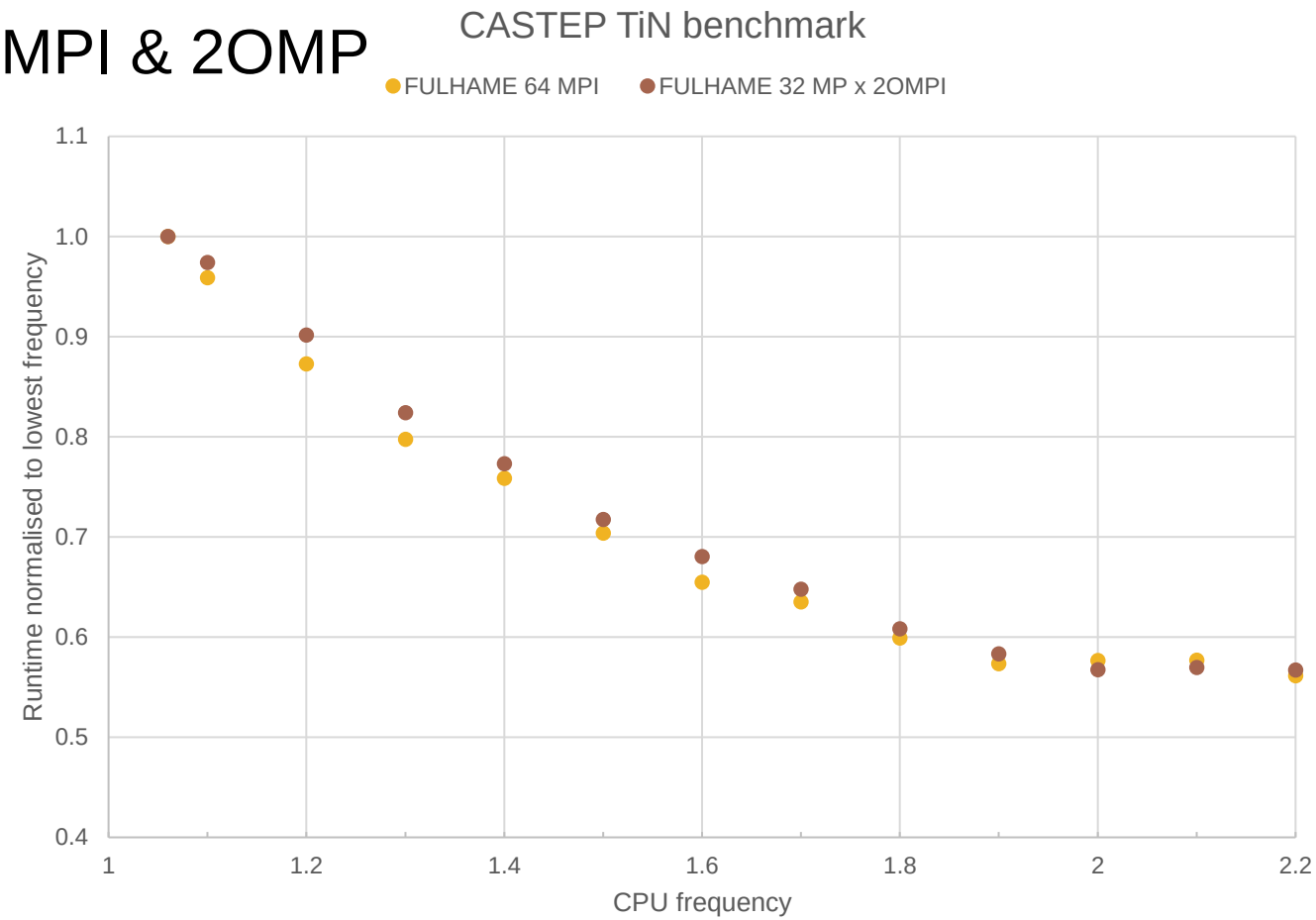
- Paper submitted a few weeks after system installation, published at PASC19 in June 2019
- Performance comparison of a range of applications up to 32 nodes
  - Comparing to Arm to established Intel-based systems

epcc

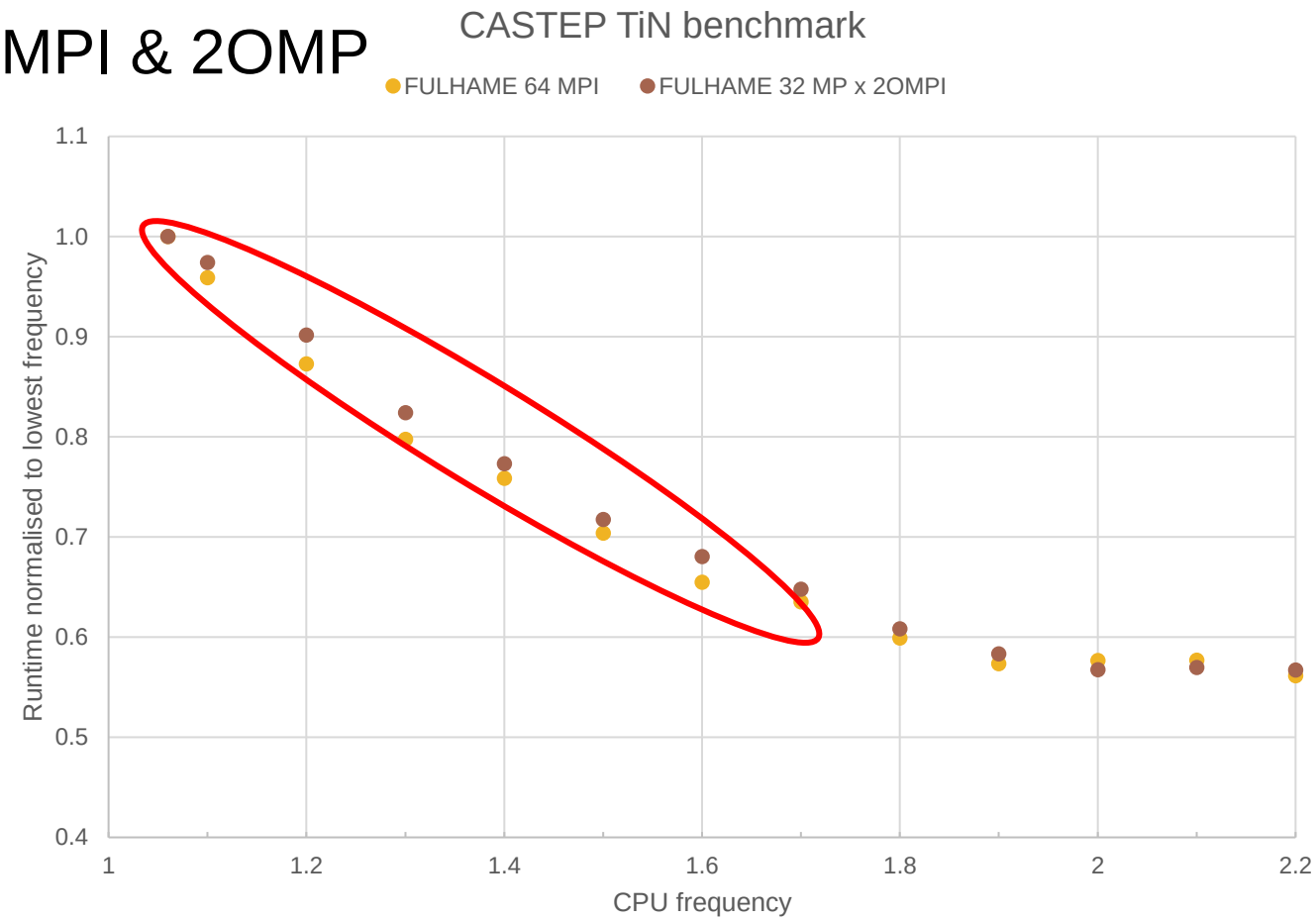
Drawbacks of original setup

- We wanted:
  - SLURM rather than PBS
    - Not support by SLES12 SP3 on ARM
  - Lustre storage
    - Not support by SLES12 SP3 on ARM
  - CPU frequency scaling and turbo states
    - Not support by SLES12 SP3 on ARM
- You should be able to spot a pattern…!
- What did we do? Upgrade to SLES15, which (in theory) supports all of the above!

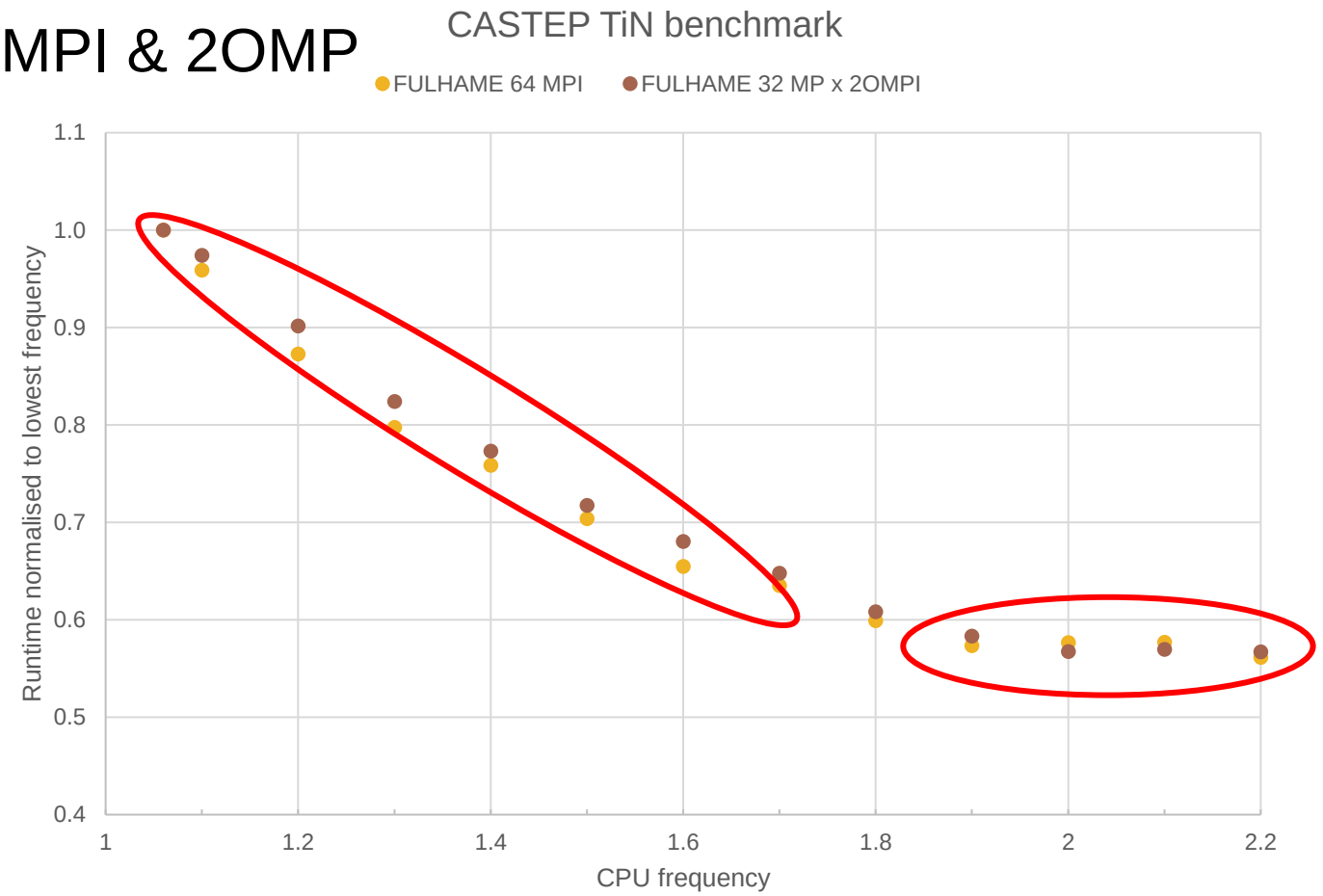
- Impact of frequency scaling on **CASTEP computational chemistry code**
- single node, TiN benchmark
- 64 MPI vs 32 MPI & 2OMP



- Impact of frequency scaling on **CASTEP computational chemistry code**
- single node, TiN benchmark
- 64 MPI vs 32 MPI & 2OMP

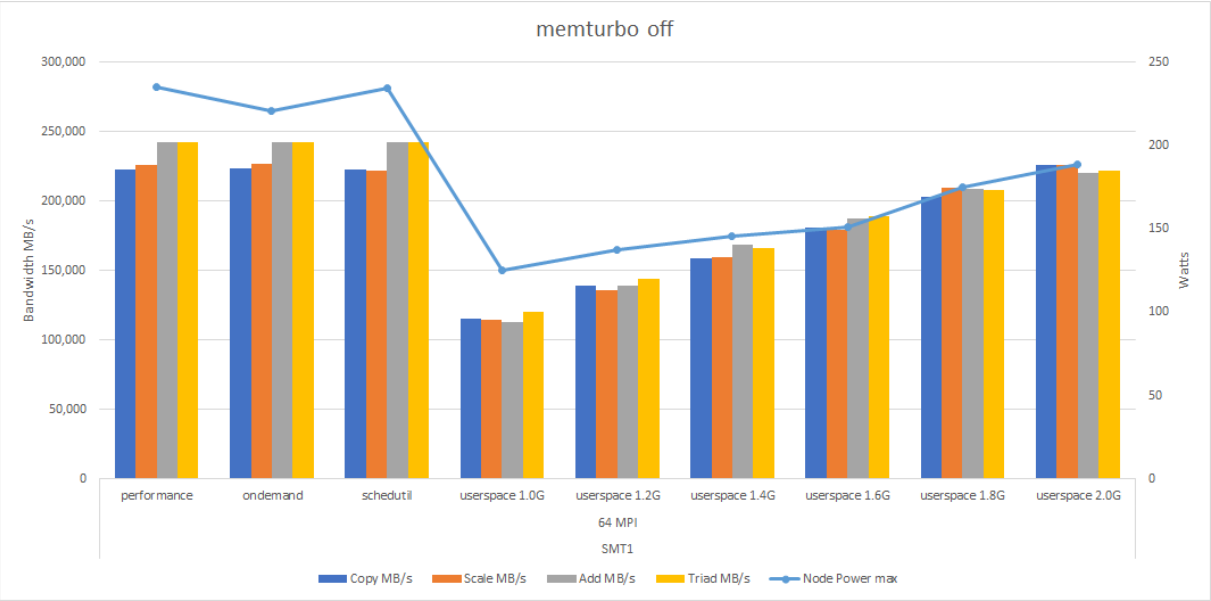


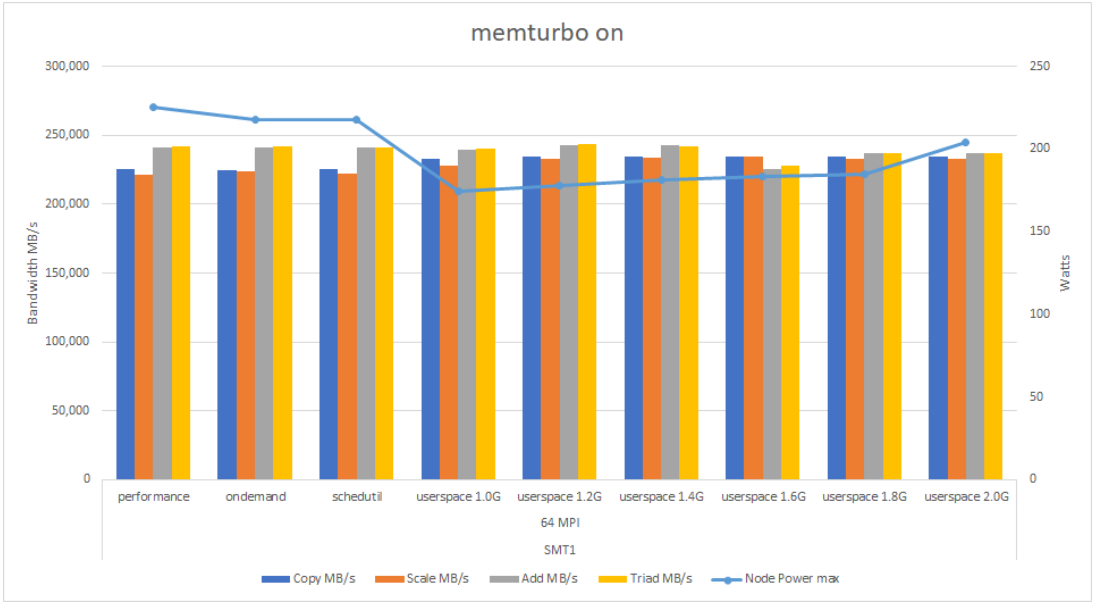
- Impact of frequency scaling on **CASTEP computational chemistry code**
- single node, TiN benchmark
- 64 MPI vs 32 MPI & 2OMP



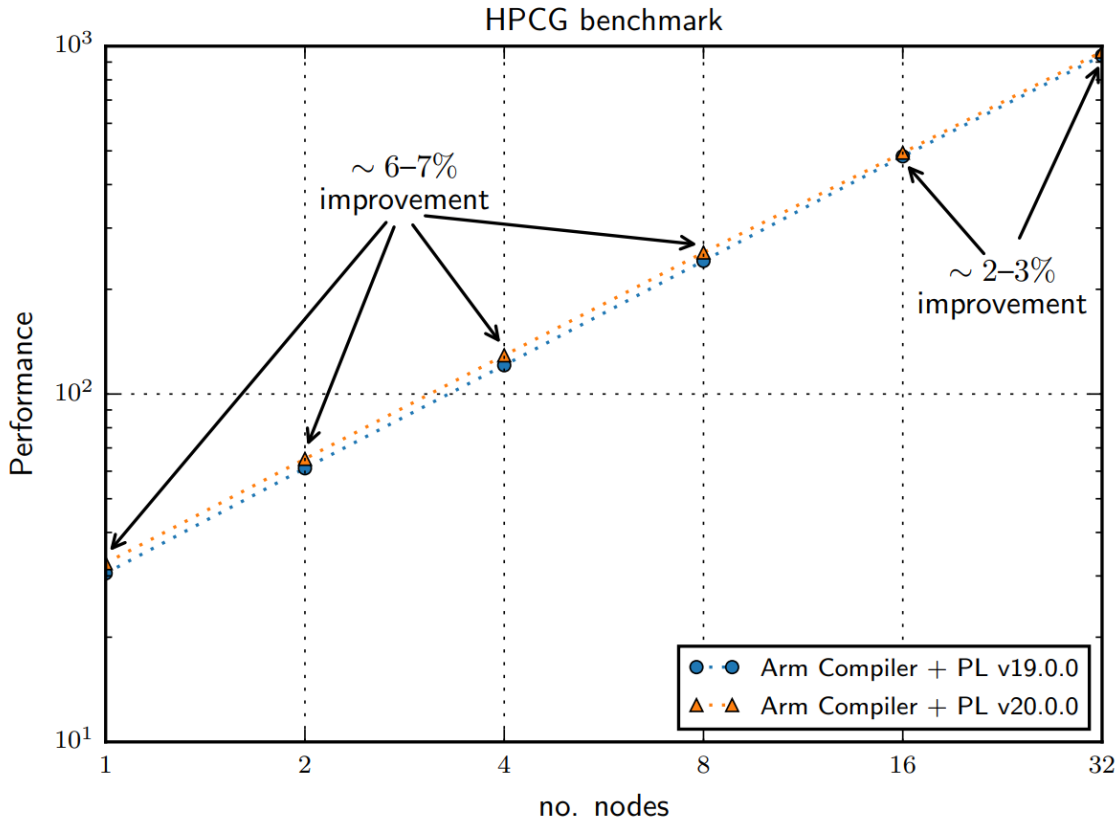
memturbo

	memturbo == off		memturbo == on	
CPU frequency governor	CPU max frequency	Memory subsystem frequency	CPU max frequency	Memory subsystem frequency
performance	2.5GHz	2.2GHz	2.2GHz	2.3GHz
ondemand	2.5GHz	2.2GHz	2.2GHz	2.3GHz
schedutil	1.7GHz	1.7GHz	2.2GHz	2.3GHz
userspace 1.0G	1.0GHz	1.0GHz	1.0GHz	2.3GHz
userspace 1.2G	1.2GHz	1.2GHz	1.2GHz	2.3GHz
userspace 1.4G	1.4GHz	1.4GHz	1.4GHz	2.3GHz
userspace 1.6G	1.6GHz	1.6GHz	1.6GHz	2.3GHz
userspace 1.8G	1.8GHz	1.8GHz	1.8GHz	2.3GHz
userspace 2.0G	2.0GHz	2.0GHz	2.0GHz	2.3GHz





HPCG – compiler evolution



- Arm compiler & PL v19.0.0 + OpenMPI 4.0.0
- Arm compiler & PL v20.0.0 + OpenMPI 4.0.2

- Compiler options

```
-O3 -ffast-math -funroll-loops
-fopenmp -std=c++11 -ffp-contract=fast -
mcpu=native
```

Recent EuroHPC announcements

- Finland (CSC) will host Lumi by mid-2021
  - 550 Petaflops, €145 million
  - Supplied by HPE
  - AMD EPYC CPUs + AMD GPUs

- Italy (CINECA) will host Leonardo
  - 248 Petaflops, €120 million
  - Supplied by ATOS
  - Intel Icelake CPUs + NVIDIA A100 GPUs

- Spanish announcement to follow …

Update on UK Exascale Project

- Since February 2020 new role as Director of Research Computing for Engineering & Physical Sciences Research Council
- Developing funding case for UK Exascale Project for Government
- UK Supercomputing Science Case completed in 2020
- Business Case for funding going to Government Autumn 2021
- Procurement in during 2022, manufacturing and installation 2023, final changes to hosting environment 2023
- Planned service opening April 2024

Exascale Project Specific 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 with regard to industry use for research
  - Pay-per-use production access will be supported
  - Specific support for SMEs
- System should be **operational by ~~2023~~ 2024**

Hosting – Advanced Computing Facility

- The ‘ACF’ will host the WCDI
- Opened 2005
- Purpose built, secure, world class data centre
- Houses wide variety of leading-edge systems
- Major expansion in 2013
  - 8 MW power upgrade
  - 550m^2 machine room + 850m^2 plant room
- New £20m expansion opened Dec ‘20
  - Further machine room and another 30MW

H y

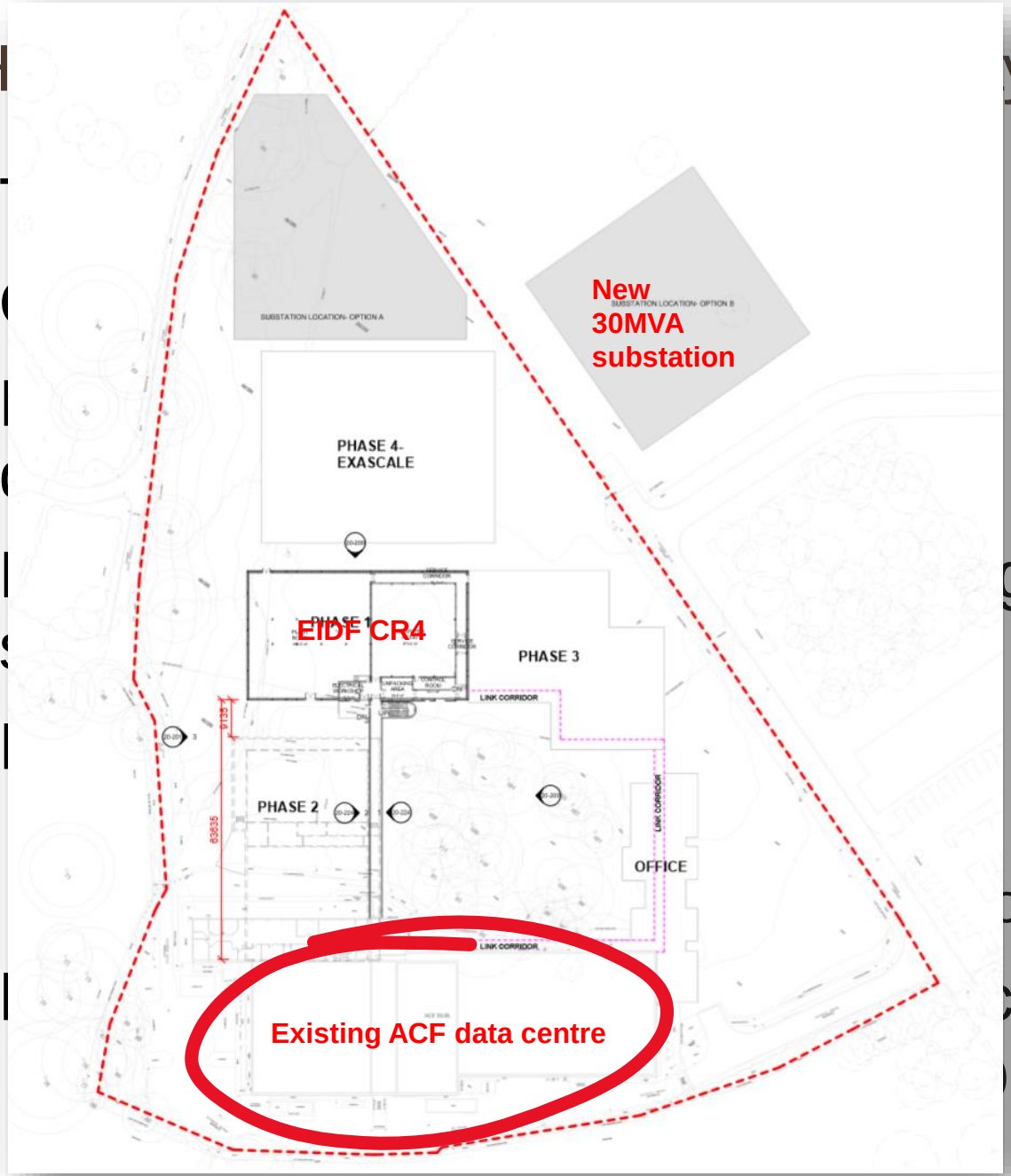
- -
- (
- I
- I
- I
- I

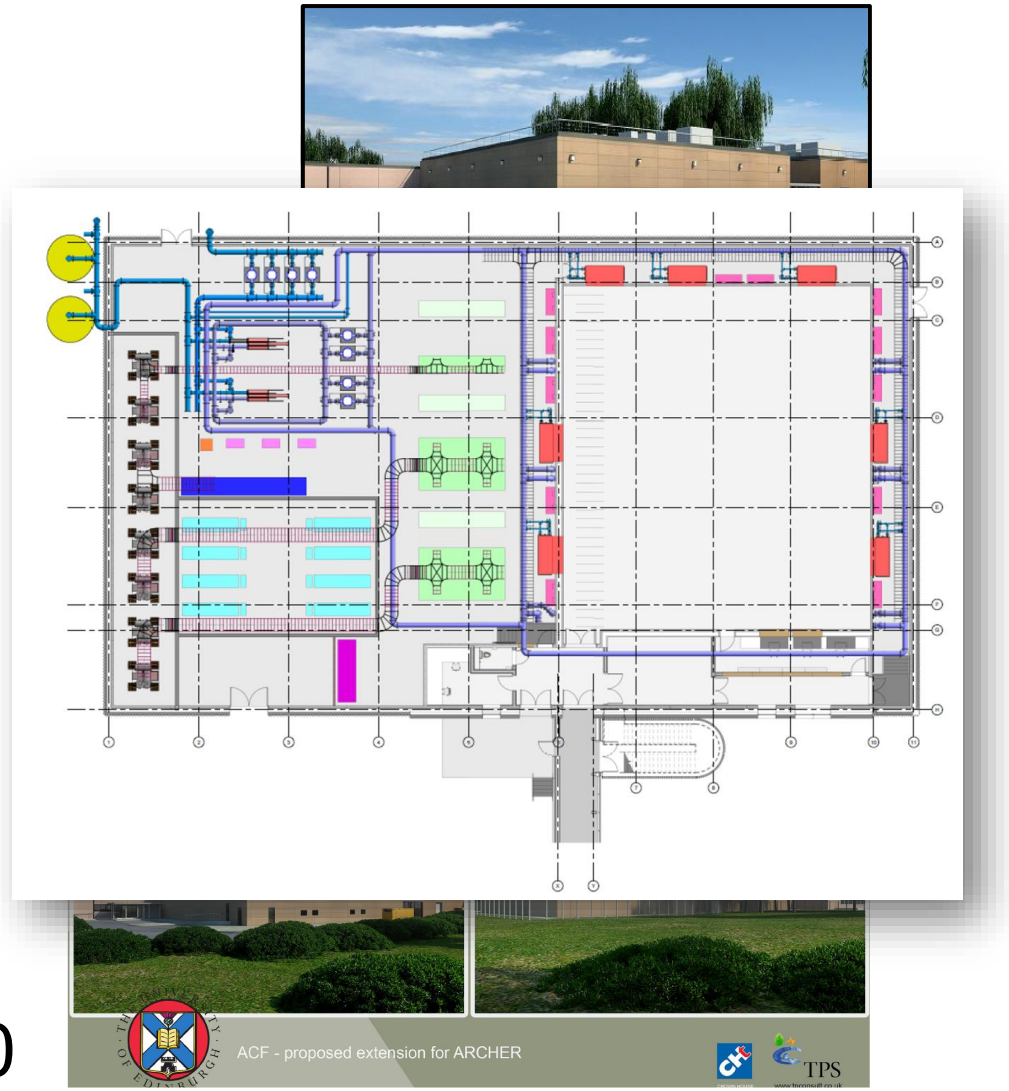
ge

oom

‘20

MW





epcc

H y

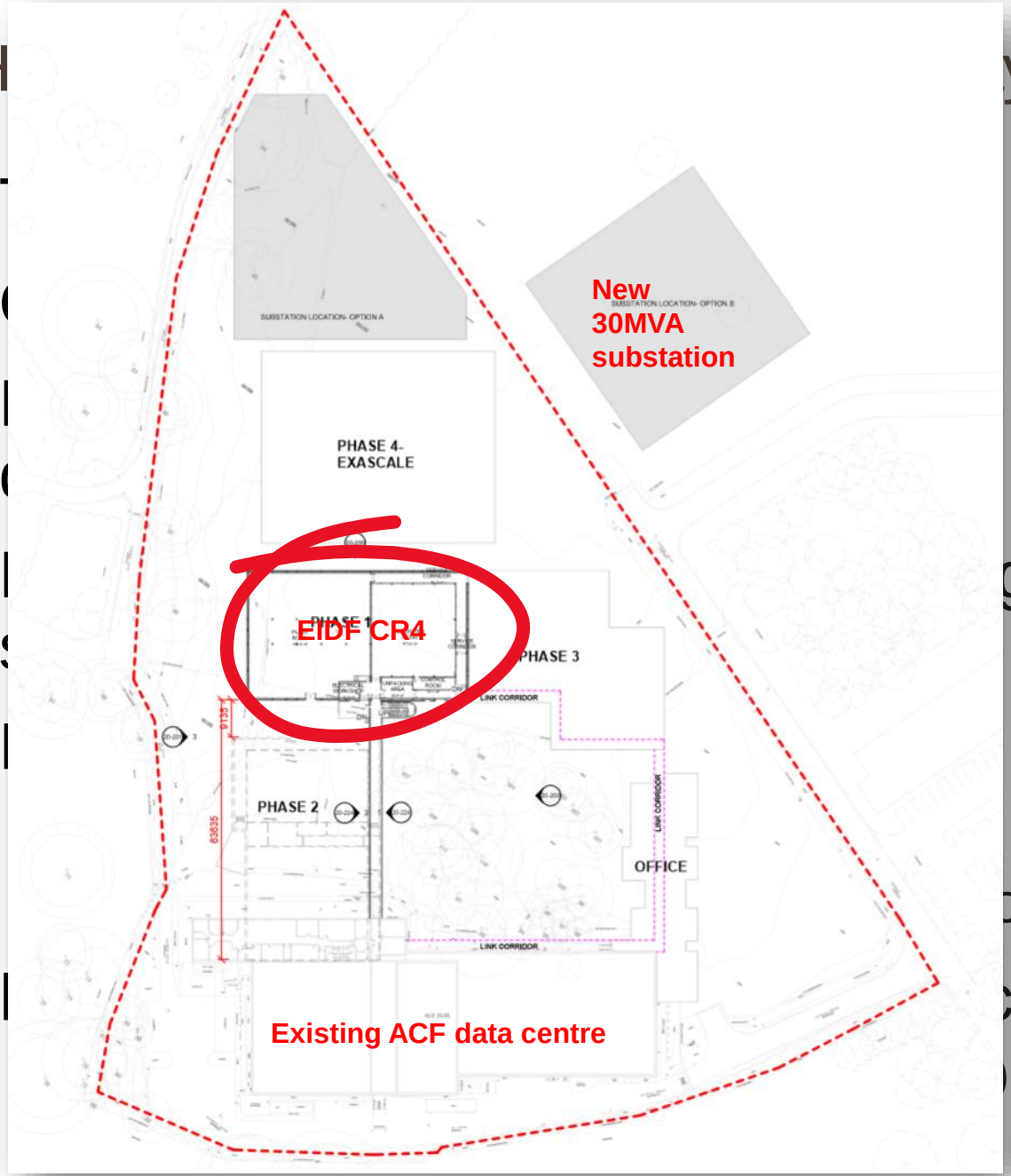
- -
- (
- I
- I
- I
- I

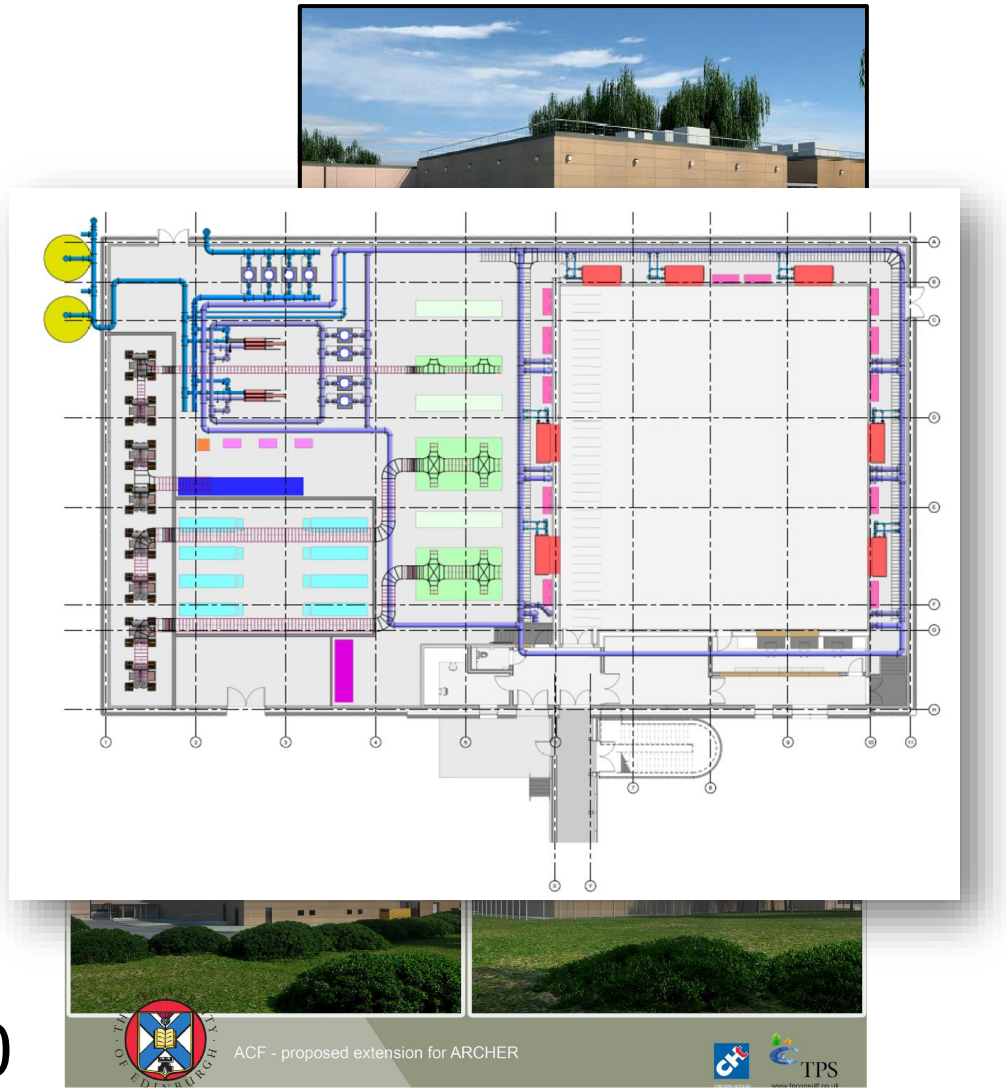
ge

oom

‘20

MW





H y

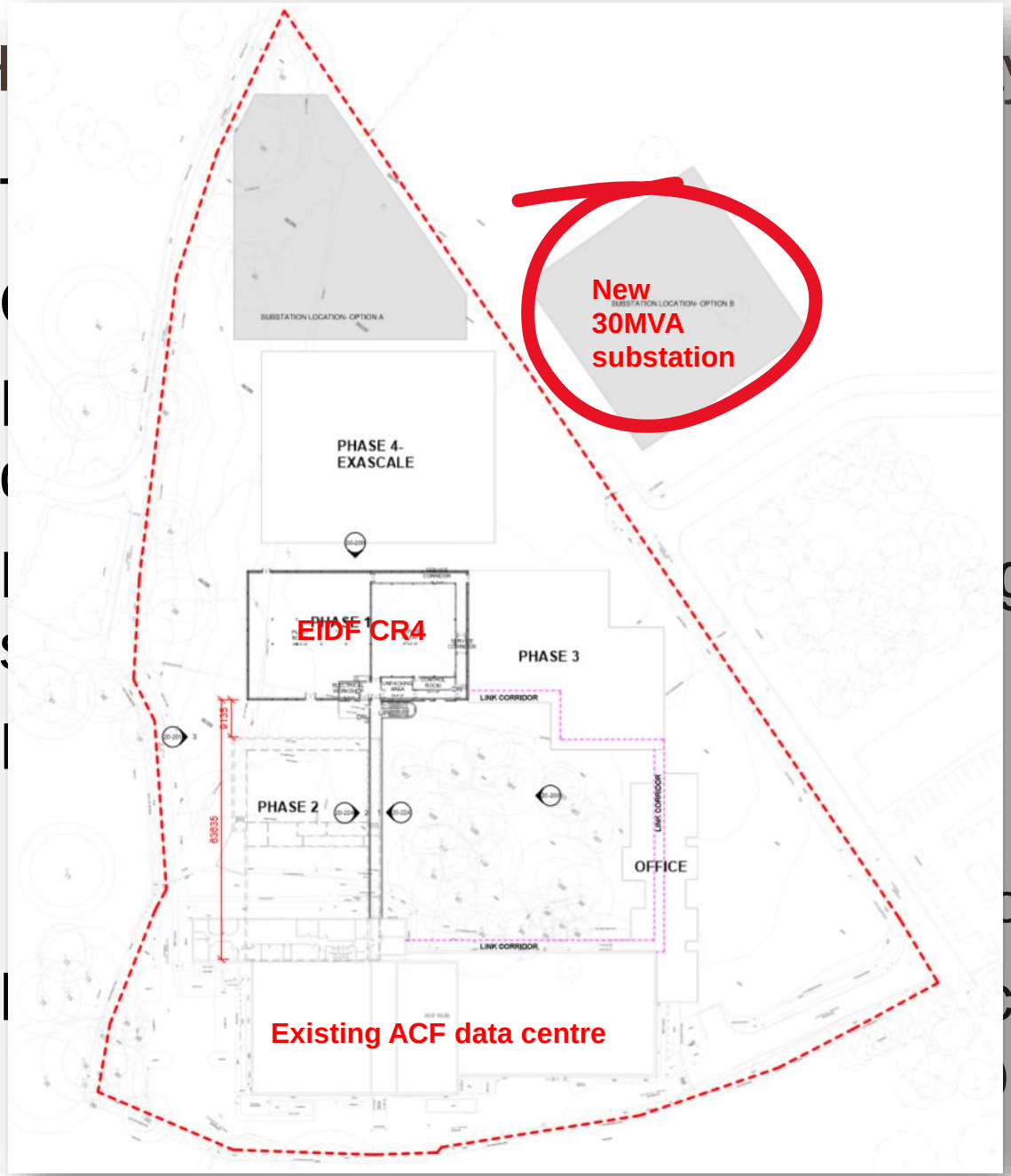
- -
- (
- I
- I
- I
- I

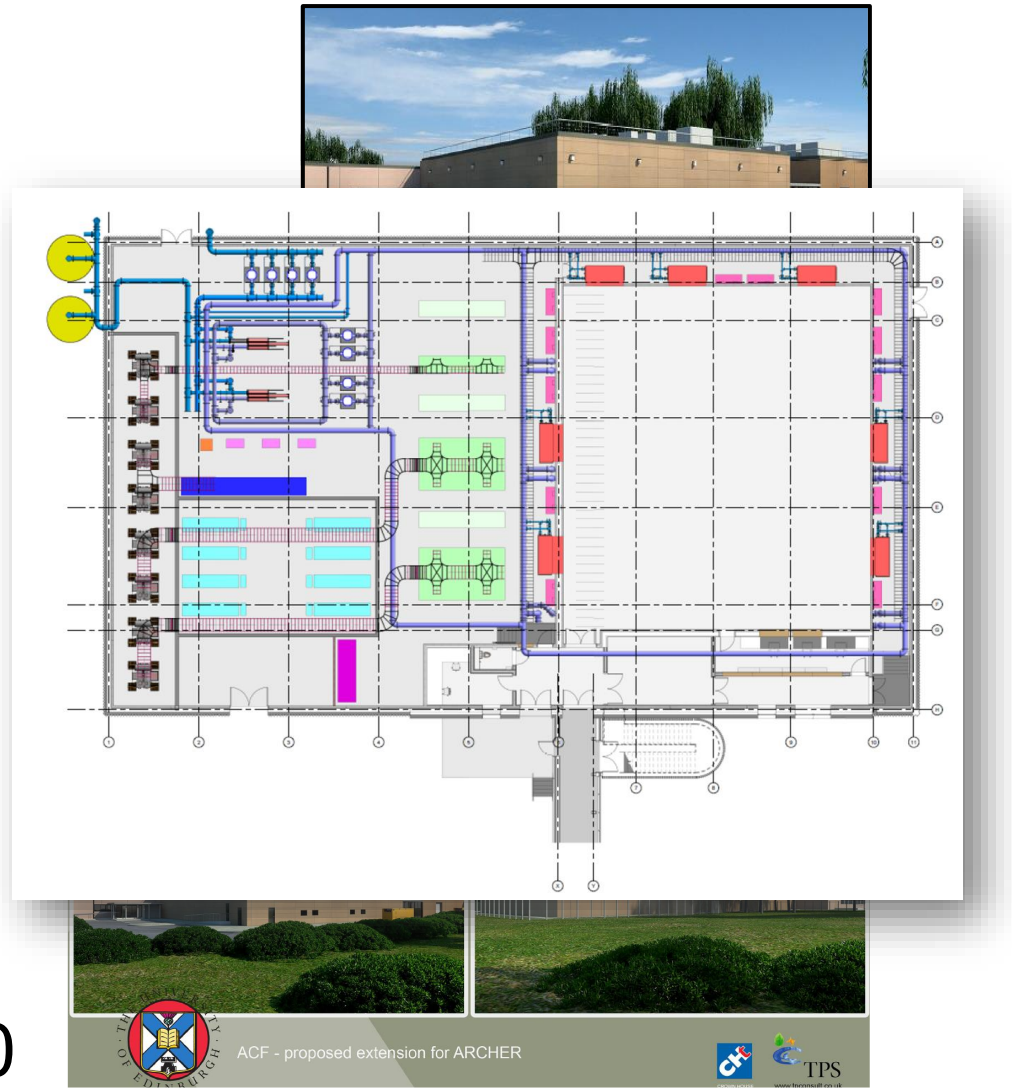
ge

om

‘20

MW





epcc

Infrastructure takes time and money

£20m – New computer room
£8m – 30MW additional power

Opened Dec 2020


epcc

Exascale Technology – Japanese versus American model

- Japanese model has attractions but difficult to sell to Government
- Lower peak performance but much longer science lifetime

HPCG Results (June 2019 and June 2020)

Rank	Site	Computer	Cores	HPL Rpeak	HPL Rmax (Pflop/s)	Top500 Rank	HPCG (Pflop/s)	Rmax to Rpeak	Fraction of Rmax
1	Riken CCS JAPAN	Fugaku	7,299,072	513.9	415.5	1	133.7	80.9%	32.17%
2	DOE/SC/ORNL USA	Summit	2,414,592	200.8	148.6	2	2.9	74.0%	1.97%
5	Riken CCS JAPAN	K-Computer	705,024	11.3	10.5	22	0.6	93.2%	5.73%

Exascale Technology – Japanese versus American model

- Japanese model has attractions but difficult to sell to Government
- Lower peak performance but much longer science lifetime

HPCG Results (June 2019 and June 2020)

Rank	Site	Computer	Cores	HPL Rpeak	HPL Rmax (Pflop/s)	Top500 Rank	HPCG (Pflop/s)	Rmax to Rpeak	Fraction of Rmax
1	Riken CCS JAPAN	Fugaku	7,299,072	513.9	415.5	1	133.7	80.9%	32.17%
2	DOE/SC/ORNL USA	Summit	2,414,592	200.8	148.6	2	2.9	74.0%	1.97%
5	Riken CCS JAPAN	K-Computer	705,024	11.3	10.5	22	0.6	93.2%	5.73%

Exascale Technology – Japanese versus American model

- Japanese model has attractions but difficult to sell to Government
- Lower peak performance but much longer science lifetime

HPCG Results (June 2019 and June 2020)

Rank	Site	Computer	Cores	HPL Rpeak	HPL Rmax (Pflop/s)	Top500 Rank	HPCG (Pflop/s)	Rmax to Rpeak	Fraction of Rmax
1	Riken CCS JAPAN	Fugaku	7,299,072	513.9	415.5	1	133.7	80.9%	32.17%
2	DOE/SC/ORNL USA	Summit	2,414,59	200.8	148.6	2	2.9	74.0%	1.97%
5	Riken CCS JAPAN	K-Computer	705,024	11.3	10.5	22	0.6	93.2%	5.73%

Exascale Technology – Japanese versus American model

- Japanese model has attractions but difficult to sell to Government
- Lower peak performance but much longer science lifetime

HPCG Results (June 2019 and June 2020)

Rank	Site	Computer	Cores	HPL Rpeak	HPL Rmax (Pflop/s)	Top500 Rank	HPCG (Pflop/s)	Rmax to Rpeak	Fraction of Rmax
1	Riken CCS JAPAN	Fugaku	7,299,072	513.9	415.5	1	133.7	80.9%	32.17%
2	DOE/SC/ORNL USA	Summit	2,414,592	200.8	148.6	2	2.9	74.0%	1.97%
5	Riken CCS JAPAN	K-Computer	705,024	11.3	10.5	22	0.6	93.2%	5.73%

Exascale Technology – Japanese versus American model

- Japanese model has attractions but difficult to sell to Government
- Lower peak performance but much longer science lifetime

HPCG Results (June 2019 and June 2020)

Rank	Site	Computer	Cores	HPL Rpeak	HPL Rmax (Pflop/s)	Top500 Rank	HPCG (Pflop/s)	Rmax to Rpeak	Fraction of Rmax
1	Riken CCS JAPAN	Fugaku	7,299,072	513.9	415.5	1	133.7	80.9%	32.17%
2	DOE/SC/ORNL USA	Summit	2,414,592	200.8	148.6	2	2.9	74.0%	1.97%
5	Riken CCS JAPAN	K-Computer	705,024	11.3	10.5	22	0.6	93.2%	5.73%

Exascale Technology – Japanese versus American model

- Japanese model has attractions but difficult to sell to Government
- Lower peak performance but much longer science lifetime

HPCG Results (June 2019 and June 2020)

Rank	Site	Computer	Cores	HPL Rpeak	HPL Rmax (Pflop/s)	Top500 Rank	HPCG (Pflop/s)	Rmax to Rpeak	Fraction of Rmax
1	Riken CCS JAPAN	Fugaku	7,299,072	513.9	415.5	1	133.7	80.9%	32.17%
2	DOE/SC/ORNL USA	Summit	2,414,592	200.8	148.6	2	2.9	74.0%	1.97%
5	Riken CCS JAPAN	K-Computer	705,024	11.3	10.5	22	0.6	93.2%	5.73%

Exascale Technology – Japanese versus American model

- Japanese model has attractions but difficult to sell to Government
- Lower peak performance but much longer science lifetime

HPCG Results (June 2019 and June 2020)

Rank	Site	Computer	Cores	HPL Rpeak	HPL Rmax (Pflop/s)	Top500 Rank	HPCG (Pflop/s)	Rmax to Rpeak	Fraction of Rmax
1	Riken CCS JAPAN	Fugaku	7,299,072	513.9	415.5	1	133.7	80.9%	32.17%
2	DOE/SC/ORNL USA	Summit	2,414,592	200.8	148.6	2	2.9	74.0%	1.97%
5	Riken CCS JAPAN	K-Computer	705,024	11.3	10.5	22	0.6	93.2%	5.73%

Exascale Technology – Japanese versus American model

- Japanese model has attractions but difficult to sell to Government
- Lower peak performance but much longer science lifetime

HPCG Results (June 2019 and June 2020)

Rank	Site	Computer	Cores	HPL Rpeak	HPL Rmax (Pflop/s)	Top500 Rank	HPCG (Pflop/s)	Rmax to Rpeak	Fraction of Rmax
1	Riken CCS JAPAN	Fugaku	7,299,072	513.9	415.5	1	133.7	80.9%	32.17%
2	DOE/SC/ORNL USA	Summit	2,414,592	200.8	148.6	2	2.9	74.0%	1.97%
5	Riken CCS JAPAN	K-Computer	705,024	11.3	10.5	22	0.6	93.2%	5.73%

Exascale Technology – Japanese versus American model

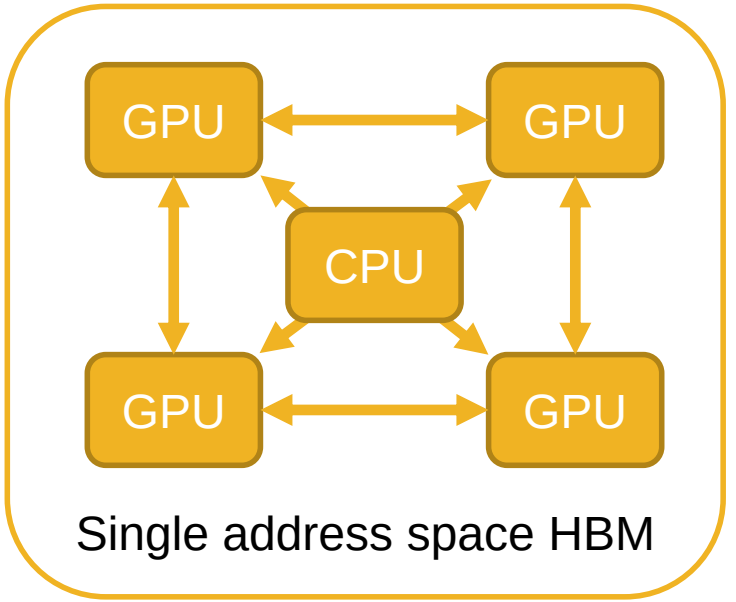
- Japanese model has attractions but difficult to sell to Government
- Lower peak performance but much longer science lifetime

HPCG Results (June 2019 and June 2020)

Rank	Site	Computer	Cores	HPL Rpeak	HPL Rmax (Pflop/s)	Top500 Rank	HPCG (Pflop/s)	Rmax to Rpeak	Fraction of Rmax
1	Riken CCS JAPAN	Fugaku	7,299,072	513.9	415.5	1	133.7	80.9%	32.17%
2	DOE/SC/ORNL USA	Summit	2,414,592	200.8	148.6	2	2.9	74.0%	1.97%
5	Riken CCS JAPAN	K-Computer	705,024	11.3	10.5	22	0.6	93.2%	5.73%

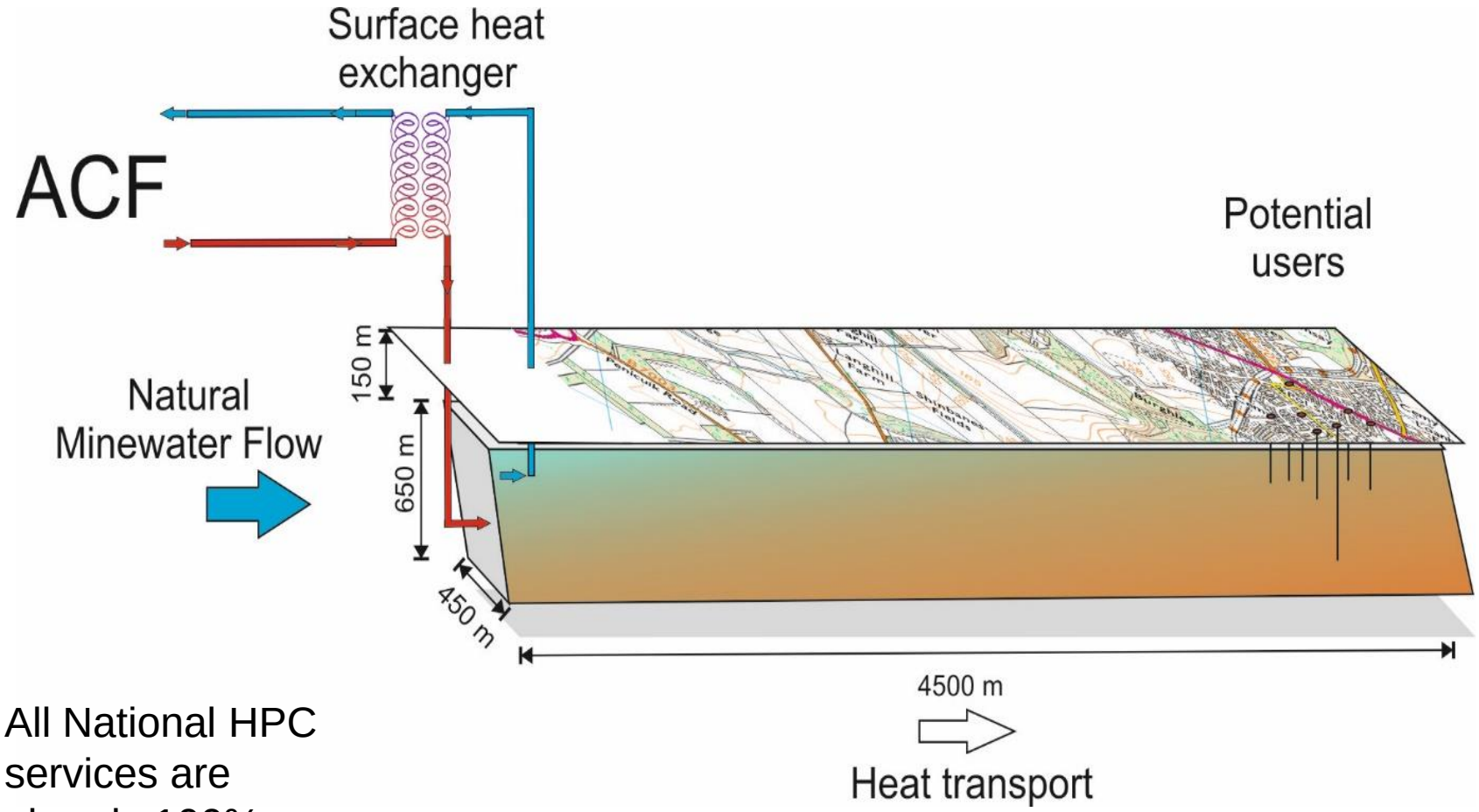
Technology –recent Exascale vendor briefings

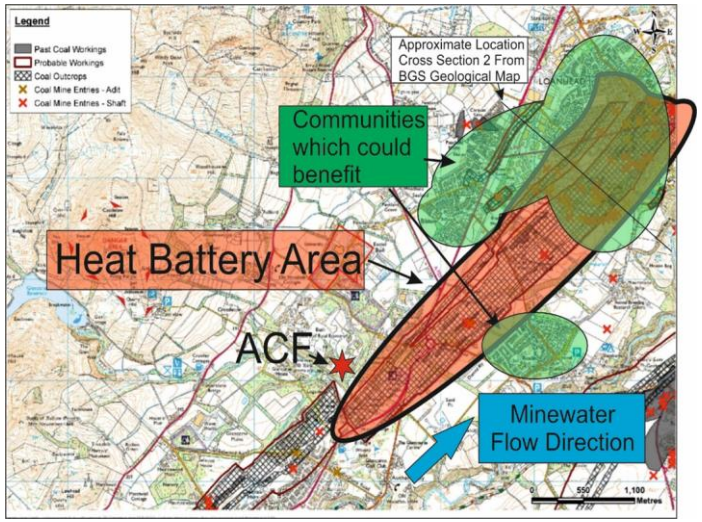
- Memory is changing
  - Many Exascale blades include HBM
    - Some designs have no DRAM at all
  - But recently LPDDR5 is being mentioned more
- Four-way competition for CPUs and/or GPUs
  - Intel versus AMD versus Arm versus NVIDIA
- GPUs market is broadening
  - AMD is strongly competing with NVIDIA
- Cabinet energy densities are rocketing
  - Today's 80-100KW cabinets will be eclipsed by cabinets at 300KW+
- Multicore CPUs are also getting AI Deep Learning features



Aiming for Net Zero

Bilston Glen Colliery, 670m, 15.0C, Minewater
Monktonhall, 866m, 25.5C, Rock
Lady Victoria, 768m, 18C, Minewater





All National HPC services are already 100% Green Electricity

epcc

Final question from Hyperion …

- **Exascale computers are very expensive. What happens if the UK doesn't fund this effort?**

- Simply put, ARCHER2 will need to be replaced in 2025 and its successor will not be an Exascale system
- This will have knock effects for UK scientific and industrial research
  - The economic benefits of an Exascale investment will be missed
  - UK science will have to apply for limited access to overseas resources
  - Many UK companies will simply have no access because of data security and other issues
- Surely the world's sixth largest economy should be investing in the future?