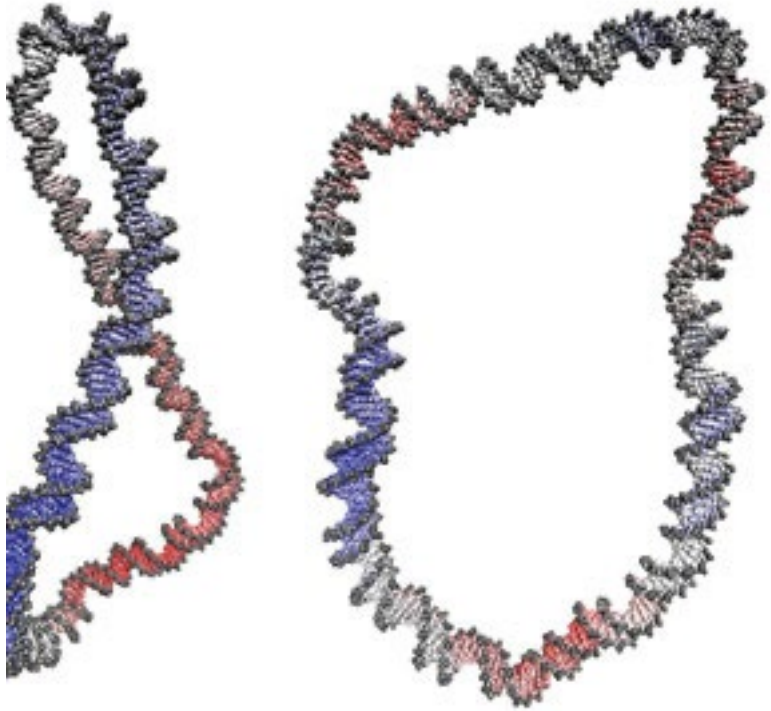
N8 CIR
COMPUTATIONALLY INTENSIVE RESEARCH

The UK Supercomputing landscape
February 2022

Alan Real, Director of Advanced Research Computing,
Durham University

Overview

- What do we mean by Supercomputing
- Funding councils and Supercomputing
- Supercomputing Tiers
- Specific facilities
  - ARC@Leeds
  - National Tier-2
  - National capability
- Access mechanisms
- What's next

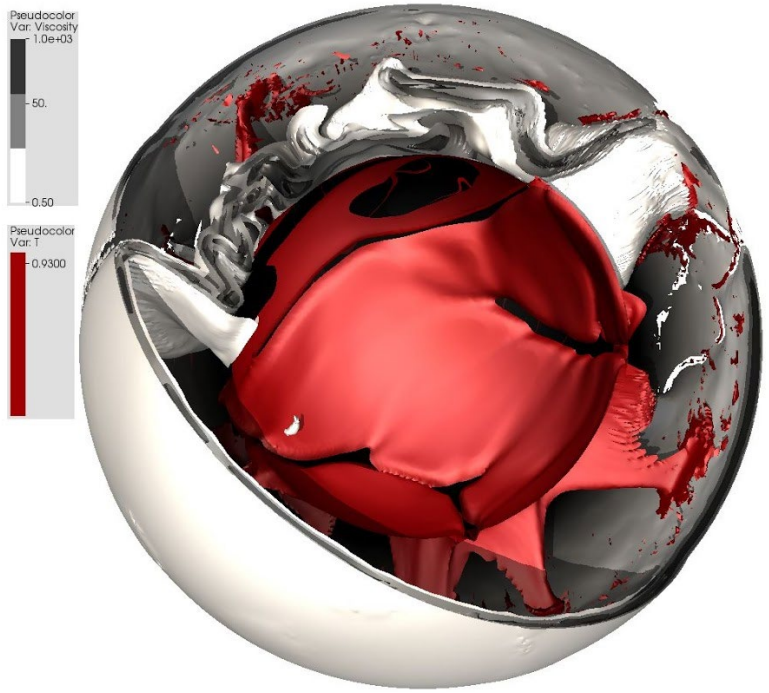


Why High-Performance Computing?

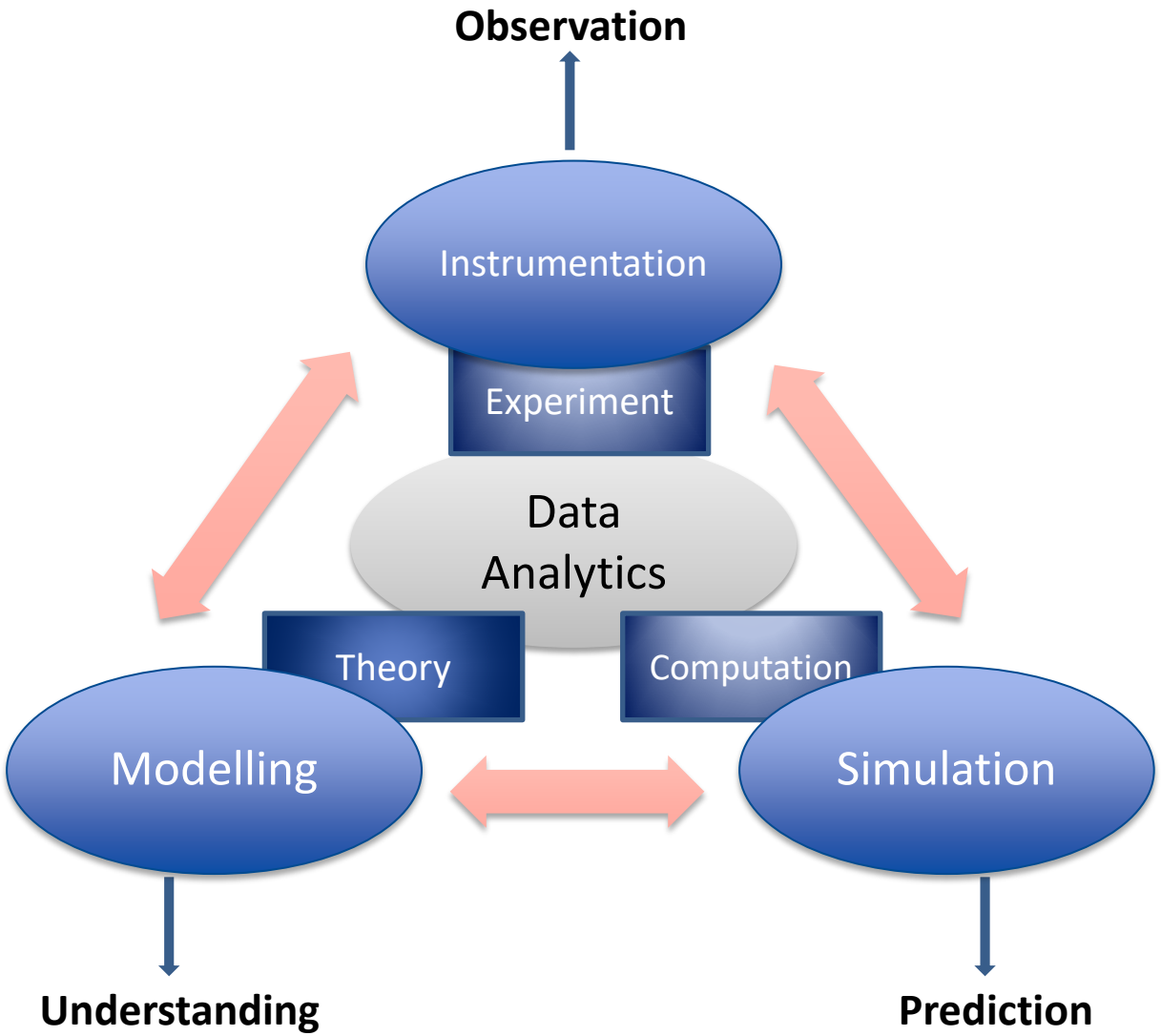
- Computational science `third leg` of scientific enquiry, alongside experiment and theory

- Computations may be mandatory because
  - Experiments are controversial (biology and medicine)
  - Experiments are too expensive (Multiphase flows in oil wells)
  - Experiments are dangerous (development of fires)
  - Experiments are illegal (radioactivity)
  - Experiments are impossible (climate)
  - Realize the full potential of investments in large infrastructures (e.g. CERN)

•Computation is now integral to many experiments, due to their data demands (e.g. Cryo-EM)



Tools of the modern scientist

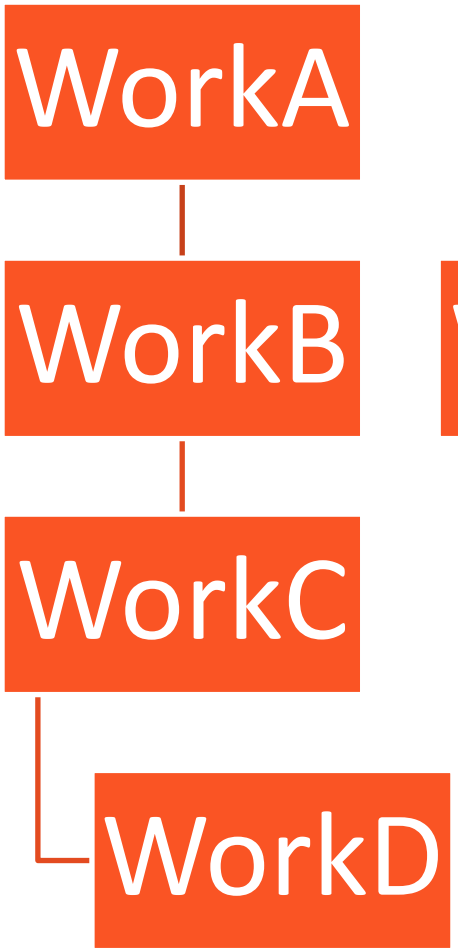


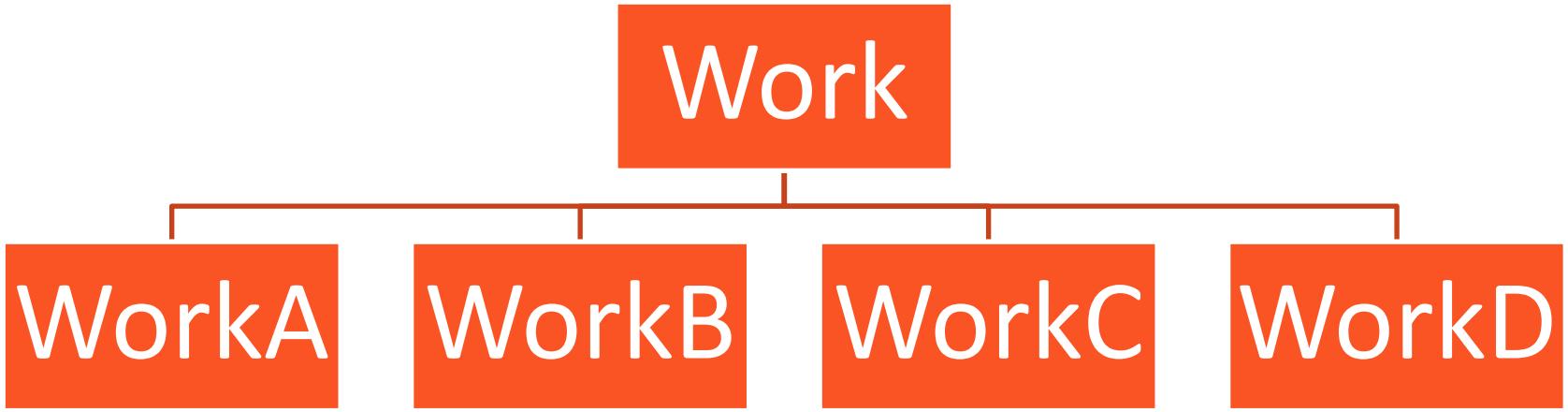
What do we mean by supercomputing?

- Enabling calculations that cannot be performed ‘on the desktop’.
- Our desktops are quite beefy nowadays
  - Particularly if they contain GPUs
- Cloud technologies allow an ensemble of these ‘desktop’ calculations.
  - Often by oversubscribing individual machines with many applications

- Supercomputing overcomes some of these limits
  - Allowing multiple machines to be used in a single calculation
  - Designed in response to scientific case, with unique combinations of
    - fast networks
    - High speed and large-scale filesystems
    - Often accelerated and allow multiple-GPU use.
    - Large memories

Supercomputing and parallel processing





Will take the time it takes to do Work (A + B + C +D)

Will take the time it takes to do the longest of work (A, B, C, D)

UNLESS:

Work A needs to tell Work B, C, D something or

Work B needs to tell Work C, or D something etc.

Longest of work (A, B, C, D) will includes any time needed to chat

Funding councils and supercomputing

- Officially, only the
  - Engineering & Physical Sciences (EPSRC)
  - Natural Environment (NERC)
  - Science and Technology Funding Councils fund supercomputing facilities.
- In reality
  - Most research councils require access to capability computing.
  - All part of National e-Infrastructure roadmap
  - Any grant can cost in time on supercomputing systems
  - Need to look at individual sites for access and costing detail

UK Research and Innovation

Natural Environment Research Council

Engineering and Physical Sciences Research Council

Science and Technology Facilities Council

Current UK supercomputing provision

Provider	Activity	Location(s)
Met Office	Weather forecasting and climate modelling	Exeter
ARCHER	Engineering, Physical Sciences, Natural Environment	Edinburgh Parallel Computing Centre (EPCC)
DiRAC	PPAN Theory	Cambridge, Durham, EPCC, Leicester
Hartree Centre	Industry and Commerce	Daresbury Lab (DL)
EPSRC Tier 2 services	Engineering, Physical Sciences	Cambridge, HPC Midlands+ (Warwick), Baskerville (Birmingham), GW4 (Bristol/Met Office), EPCC, JADE (at DL), UCL (JISC Southern DC), N8 (Durham), QUB
Research Data Facility	EPSRC, NERC, PPAN	EPCC
JASMIN (NERC)	Natural Environment	Rutherford Appleton Lab (RAL)
Earlham Institute	Life Sciences	Norwich

Supercomputing tiers

For completeness

- DiRAC www.dirac.ac.uk
- Particle Physics Astronomy & Nuclear Physics Science communities
  - Sites at Cambridge, Durham, Edinburgh and Leicester
  - E.g. COSMA facility in Durham, produces most detailed simulations of the Universe
- PRACE www.prace-ri.eu
  - 7 HPC resources across Germany, France, Italy, Spain & Switzerland
  - UK participation but excluded from new activities until 3rd country agreement in place
- INCITE www.doeleadershipcomputing.org
  - Access to US leadership systems (e.g. Summit/Sierra)
  - US Department of Energy Office for science
  - Any researcher can apply for resources
  - Competitive awards for large-scale computationally intensive grand-challenge projects

Tier-3

- Most research intensive universities have some form of HPC provision.
- Leeds has ARC www.leeds.ac.uk/arc
  - Free at the point of use
  - Regular investment cycle
  - X86 and X86+GPU system with high performance interconnect & filesystem
- Designed to suit a broad spectrum of needs
- Tends to be busy with long wait times
- Additional equipment can be costed into grants and integrated.

- Cambridge Tier-2 & DiRAC, MMM-hub, NI-HPC all built as extensions to Tier-3

Cambridge – initial system

- **Intel**
- 1152 * 32 core Intel Skylake + Intel Omnipath
- 342 * 64 core Intel Knights Landing

Nvidia

- 360 * Nvidia Pascal GPU + Mellanox EDR

Storage

- 500TB fast flash (500GB/s)
- 1PB disk (/home)
- 12PB disk (/rds)
- 10PB cold storage (tape)

Offer time allocation to partners

- EPSRC Tier2
- STFC DiRAC
- STFC IRIS
- University of Cambridge

Cambridge – 2020 upgraded system

https://www.hpc.cam.ac.uk/

- **Intel**
- 672 * 56 core Intel Cascade Lake + Mellanox HDR
- 1152 * 32 core Intel Skylake + Intel Omnipath
- 342 * 64 core Intel Knights Landing

Storage

- 1PB fast flash (1TB/s)
- 1PB disk (/home)
- 21PB disk (/rds)
- 10PB cold storage

Nvidia

- 480 * Nvidia ?? GPU + Mellanox HDR
- 360 * Nvidia Pascal GPU + Mellanox EDR

Offer time allocation to partners

- EPSRC Tier2
- STFC DiRAC
- STFC IRIS
- UK Atomic Energy Authority
- University of Cambridge (HDR-UK, Clinical med)

Joint Academic Data Science Endeavour (JADE)

- www.jade.ac.uk
  - Housed and operated by the Hartree centre, supported and managed by Oxford.
  - Initially 22 Nvidia DGX-1 with 8 P100 GPUs
  - Upgraded to V100 GPUs
  - Targeted towards Machine learning software packages.
  - Hardware potentially useful for Molecular Dynamics
  - Scales reasonably across GPUs within individual DGX nodes.

- Jade 2 upgrade in 2020 refresh
  - Around 50 DGX-1 MaxQ
  - Low-power & low-bandwidth version of the V100 GPU (~2/3 of performance of V100)
  - Consortium expanded to include ML and Biomolecular simulation communities.

MMM Hub upgrades + refreshes

Two Tier 2 HPC services:

- Thomas – MMM Hub
  - + £4m (hw), £0.5m (running)
  - Power paid by partner (inc. UCL) contributions
- Michael – Faraday Institution
  - + £1.5m (hw), £0.7m (running)

- Based on "Kathleen" Tier 3
  - Won by HPE
  - Intel Xeon Cascade Lake, 40 core nodes, 192 GB RAM, OPA, 5,760 core islands, 7,680 core initial deployment

- Thomas replacement system:
  - 23,040 cores of Xeon Cascade Lake
    - 1,440 core islands
  - Six high memory nodes
    - 3x 1.5TB, 3x 3TB
  - OPA
  - 1PB of storage
  - Fully in service by late May/early June

- Michael upgrade:
  - 8,160 *additional* Cascade Lake cores
  - New node type in that cluster
  - We're used to heterogeneous clusters from Tier3 HTC
  - Delivery mid-late Feb

Cirrus 1

HPE 8600 / SGI ICE XA

280 Compute Nodes

- 2x18 core Broadwell
- 256GB RAM

2 GPU Nodes

- 2x20 core Skylake
- 384GB RAM
- 4xNvidia Tesla V100-SXM2-16GB

- 400TB Lustre
- 26 FDR Infiniband Switches
- PBS Pro

- WoS Object Store available to all Tier 2 sites

2020 Upgrade

Two proposed purposes:

- Preparing for heterogeneity at Exascale
- Support for growth in Artificial Intelligence and Machine Learning

Achieved through:

- Expansion of GPU capability - 36xGPU Nodes
  - 2x20 core Cascade Lake
  - 384GB RAM
  - 4xNvidia Tesla V100-SXM2-16GB

Addition of a fast NVMe storage capability

- RP80 6.4TB NVMe drives
- **256TB usable**

- **Isambard**
- **10,752** Armv8 cores (168n x 2s x 32c)
  - **Cavium ThunderX2 32 core @2.5GHz**
- Cray XC50 ‘Scout’ form factor
- High-speed **Aries** interconnect
- Cray HPC optimised software stack
  - Compilers, math libraries, CrayPAT, …
  - Also comes with all the open source software toolchains: GNU, Clang/LLVM etc.
- Hosted for the Consortium by the Met Office in Exeter
- https://gw4.ac.uk/isambard/

Isambard 2 upgrades

- Q2 2020:
  - Will double the size to 21,504 cores, 336 nodes, making. Isambard 2 the largest Arm system in Europe (for now!)
  - Will also double the size of the storage to 1PB
  - Will refresh our architecture comparison system with the latest CPUs and GPUs from Intel, AMD and NVIDIA
- Q3 2020
  - Will add a Cray CS500 Fujitsu A64fx-based system to Isambard 2 in Q3 2020
  - 72 A64fx nodes in a cluster, IB interconnect

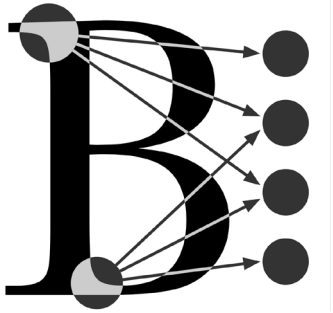
- Partnering with Cray to access a wide set of toolchains
  - Isambard 2 will be one of the first such systems in the world

NI-HPC
NORTHERN IRELAND
HIGH PERFORMANCE COMPUTING

N8

- www.ni-hpc.ac.uk
- Kelvin-2
- Queens University Belfast & Ulster University
  - 8000 AMD Rome cores
  - Dual 64-core CPU 768GB RAM nodes
  - 32 V100 GPUs
  - 4x 2TB nodes
  - EDR infiniband
- 65% for NI-based researchers
- 35% for UK-based researchers
- RSE support for partner organisations
- Fast track approval process for new users (20,000 wall clock hours)

- www.sulis.ac.uk
- Warwick led, with Birmingham, Nottingham, Queen Mary, Loughborough, Leicester.
  - 25,728 AMD Rome cores
  - Dual 64-core 2.25GHz CPU 0.5TB RAM nodes
  - 90 A100 GPUs (30 nodes with 3xA100 per node)
  - 4x 1TB nodes
  - HDR 100 infiniband
  - 2PB filesystem
- Focuses on ensemble calculations
- Access via Access to HPC call (website has advice)



- https://www.baskerville.ac.uk
- Birmingham led
  - 46 Compute nodes
  - 2x 36-core 2.4GHz Intel Xeon 8360Y
  - 4x A100 40GB GPUs [Nvlink]
  - 0.5TB RAM
  - 1TB NVMe
  - HDR infiniband
  - 0.5PB flash storage.
- Consortium access for: Birmingham, Turing, Diamond
- EPSRC Access to HPC call

Northern intensive Computing environment (Bede)

- https://n8cir.org.uk/supporting-research/facilities/bede/
- Durham-led, with DiRAC and Cambridge on behalf of N8 Research Partnership
- Designed to overcome limitations of accelerated computing.
  - Bridge between experiment and simulation
  - Prove scaling, exploit GPU/CPU memory coherence & distributed GPU
  - 38% National access, 62% N8.
- Around ~1 Pflop.
  - 32-node IBM Power 9 based system, 4GPUs per node (0.5TB/node), NVLink & EDR IB.
  - 4x T4 nodes, 2Pb/10Gb/s Lustre.

GROMACS single GPU:

- Small system ~ 25% quicker than JADE2
- Large system ~75% quicker than JADE2

- https://www.archer2.ac.uk/
- Peak performance: ~28 PFlops
  - CASTEP: 11x ARCHER throughput
  - CP2K: 9x ARCHER throughput
  - GROMACS: 13x ARCHER throughput
  - OpenSBLI: 9x ARCHER throughput
  - Met Office UM: 18x ARCHER throughput
- 5,848 compute nodes in 23 cabinets
  - Dual AMD ROME: total 748,544 cores
- 14.5 PB Lustre (4 file systems)
- 1.1 PB solid state burst buffer (*Gazelle*)
- 1 PB home (backed up) storage
- Cray Slingshot interconnect in Dragonfly topology

Access mechanisms

- Most systems run pump-priming access
  - Can be used to prepare larger allocation request
- ARCHER driving test can be used to gain ARCHER allocation
- EPSRC run Open Access calls twice per year for larger resource requests
  - https://www.ukri.org/opportunity/access-to-high-performance-computing/
  - These are subject to peer and technical review
  - Awarded according to ranking by panel
- EPSRC also award significant time via HEC consortia
  - HECBioSim access to Bede
  - Have their own award mechanisms
  - MMM-hub allocations proceed via MCC and UKCP consortia.
    - https://mmmhub.ac.uk/
- For ARCHER, the KAU calculator is your friend!
  - http://www.hecbiosim.ac.uk/archer/aucalculator

Future systems

- Large scale computing case (20th September 2021)
- https://www.gov.uk/government/publications/large-scale-computing-the-case-for-greater-uk-coordination/large-scale-computing-the-case-for-greater-uk-coordination-html-version
  - A range of well-coordinated supercomputing services with appropriate diversity to support the research programme of UKRI and UK industry
  - Sustainable long-term funding including staff and power
  - Support software engineering work and algorithm development
  - Deploy an Exaflop system, for genuine exascale science (ExaFlops and ExaBytes)
  - Align with HPC vendor technology roadmaps for effective exploitation of emerging tech.
- ExCALIBUR: www.excalibur.ac.uk
  - Initiative led by UK Met Office, UK Atomic Energy Agency and UKRI to develop software for Exascale
  - Hardware & Enabling software project run in conjunction with this.
  - Knowledge integration & cross cutting themes

Acknowledgements

- ARCHER: Alan Simpson (EPCC)
- EPSRC Tier-2 sites: Jo Beech-Brandt (EPCC), Jeffery Salmond (Cambridge), Vaughan Purnell (QUB), James Grant (Bath) / Simon McIntosh-Smith (Bristol), Owain Kenway (UCL)
- DiRAC: Mark Wilkinson, Clare Jenner & Jeremy Yates
- The N8 Partners