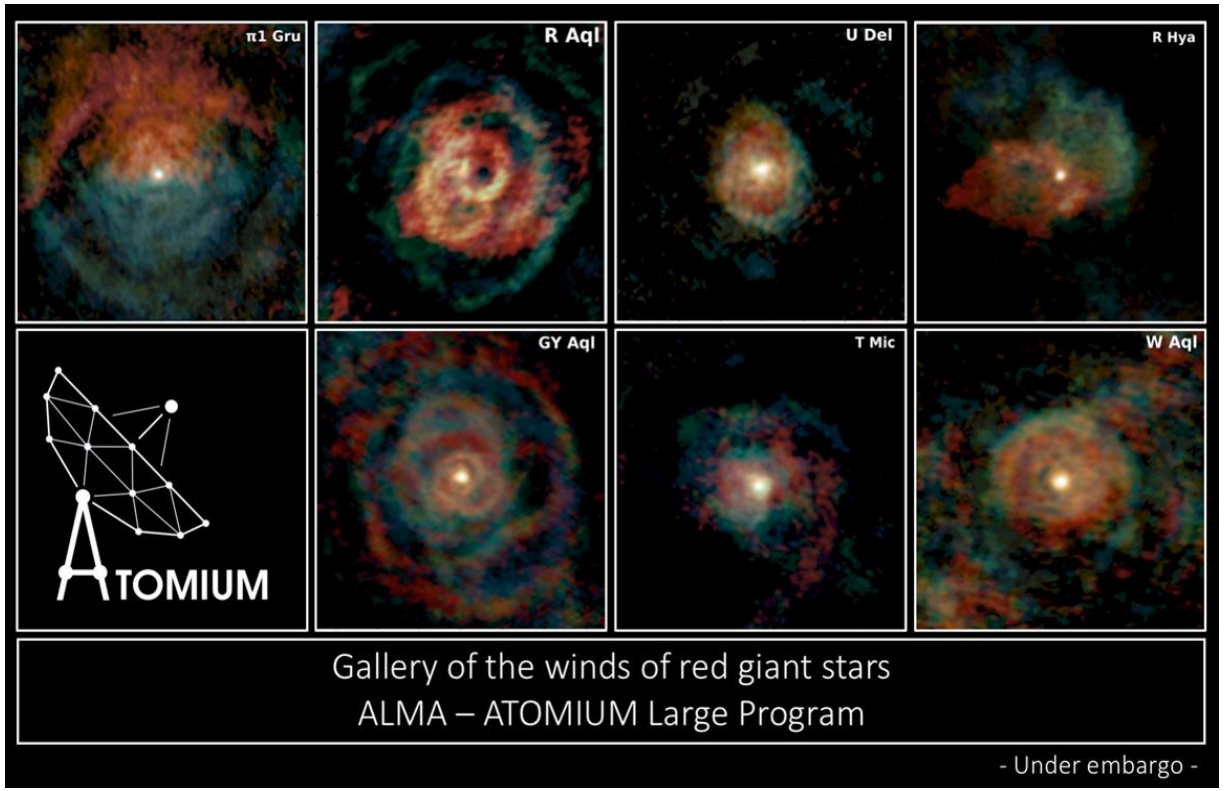


Massively Parallel Computing

Jeremy Yates (DiRAC, ExCALIBUR)
Mark Wilkinson (DiRAC, IRIS)
Tom Griffin (DiRAC, Scientific Computing Division)
Alison Kennedy (Hartree Centre)

UKRI Roadmap: Where do we lie

Using the approach taken by ESFRI the UKRI Infrastructure roadmap is structured as sectors:

- Biological sciences, health and food (BBSRC lead, MRC)
- Environment (NERC lead, Met Office)
- Energy (EPSRC)
- Physical Sciences & Engineering (STFC lead, EPSRC)
- Social sciences, arts and humanities (ESRC Lead, AHRC)
- Computational & e-infrastructures (EPSRC lead)
- https://www.ukri.org/research/infrastructure/

UKRI e-Infrastructure Landscape

What is an e-Infrastructure

- **e-Infra facilities**:
  - Provide generic facilities to a wide range of users from many research fields
  - Have expert user support and help
- **e-Infra associated with other facilities:**
  - Provide tailored data storage/analysis tools/high-throughput computing for their users
  - Have expert user support and help
- **e-Infra associated with research centres/institutes**
  - Often have `in-house` e-infrastructure tailored to support their research
  - May be dependent on access to e-infrastructure facilities for larger-scale jobs

Research Drivers for Massive Parallel Computing

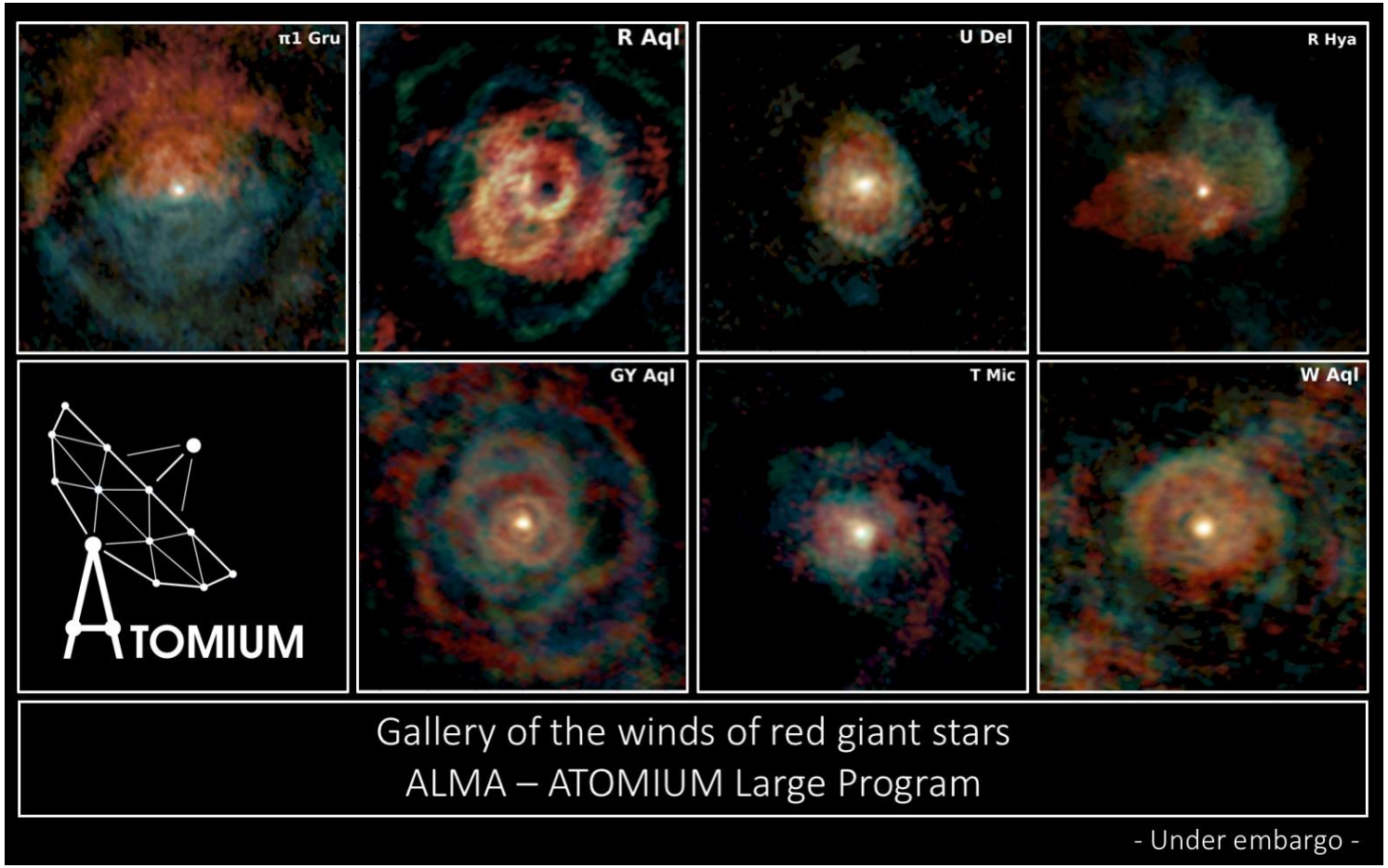
- Computational requirements of models are increasing due to
  - Increased resolution: running models based on existing understanding but at finer scales.
  - Increased complexity: introducing new processes into models to reflect progress in theoretical understanding; often needed to match resolution increases.
  - Coupling of models: multi-physics, multi-scale modelling; ultimate goal is real-time, `whole system modelling`.
  - Quantification of modelling uncertainty using large ensembles of simulations to provide robust statistics
- Direct numerical simulation and modelling increasingly a core research activity across all UKRI research areas: significant growth in new fields adopting supercomputing approaches (e.g. deep learning on social science data sets).
- Growing requirement for simulations and modelling concurrent with experiments or observations so that models evolve in line with data acquired. Experimental/observational facilities therefore need access to local Tier 2 computing capabilities as well as the option to burst out to the Tier 1 or Tier 0 national supercomputing facilities.
- Data analytics increasingly depend on more sophisticated algorithms as data volume, variety and complexity all increase towards exascale. Convergence of high throughput computation and supercomputing approaches as the data volume and complexity grows
- AI: implications for both hardware and software

Themes

Theme	Ingredients
Software and Skills	• UKRI strategy to emphasise centrality of software • Access to RSE-type support for new/all communities • Quality control support/tools (CI etc) • New software – exascale, AI etc • Co-design programme (maths, computer science, etc.) • Training: across career stages; coordination
Network and Access Management	• Janet • Access management tools • Cloud: computing and data storage
Industry	• Proper programme to support collaboration, involvement
Computing	• Sustained investment in supercomputing tiers, aiming towards exascale by 2030 • Sustained investment in research computing for facilities/ research centres • Technology foresight/test-bed programme
Data eInfrastructure	• Sustained investment in data centres to deal with ever-increasing capacity requirements • Data Curation and Data Management

Specific Why – Some Stellar Astrophysics

- Mass loss from Red Giants
- Non linear, non local process
- MHD, Chemistry, Material Science, Radiative Transfer
- Calibration (Signal Processing), Interferometry, Image Analysis
- AL/ML to reduce and process data – learning calibration techniques that were done by eye
- AL/ML to identify physical structures
- High RAM Clusters to reduce data
- Lower Clusters to model and interpret data



Data Motion – Sinks and Sources

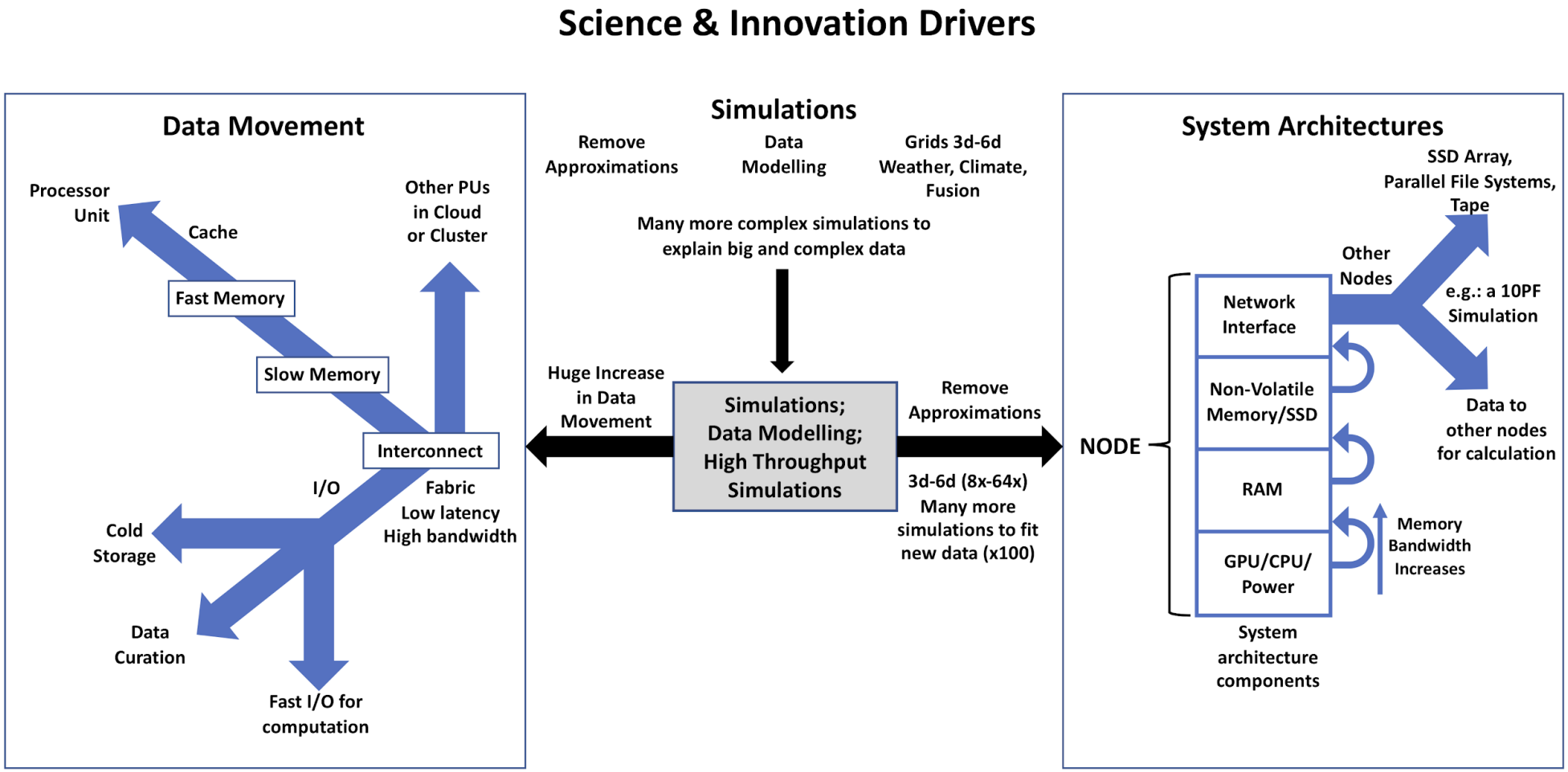
- A common mode in system design has been the idea of data motion between the various system components
- Too much focus has been on the Processor unit registers and their floating operations (flops) rating.
  - the Top500 bears little relationship to application performance
- The performance (bandwidth and latency) and size of the memory hierarchy, the storage hierarchy and the interconnect (Networking) itself have an equal, if not more, importance.
  - These get the data & instruction sets in, and then take data out.
  - They are a pinch points that need to be attended to in system and application design
- IO is a parallel process as well

Parallelism is when tasks literally run at the same time

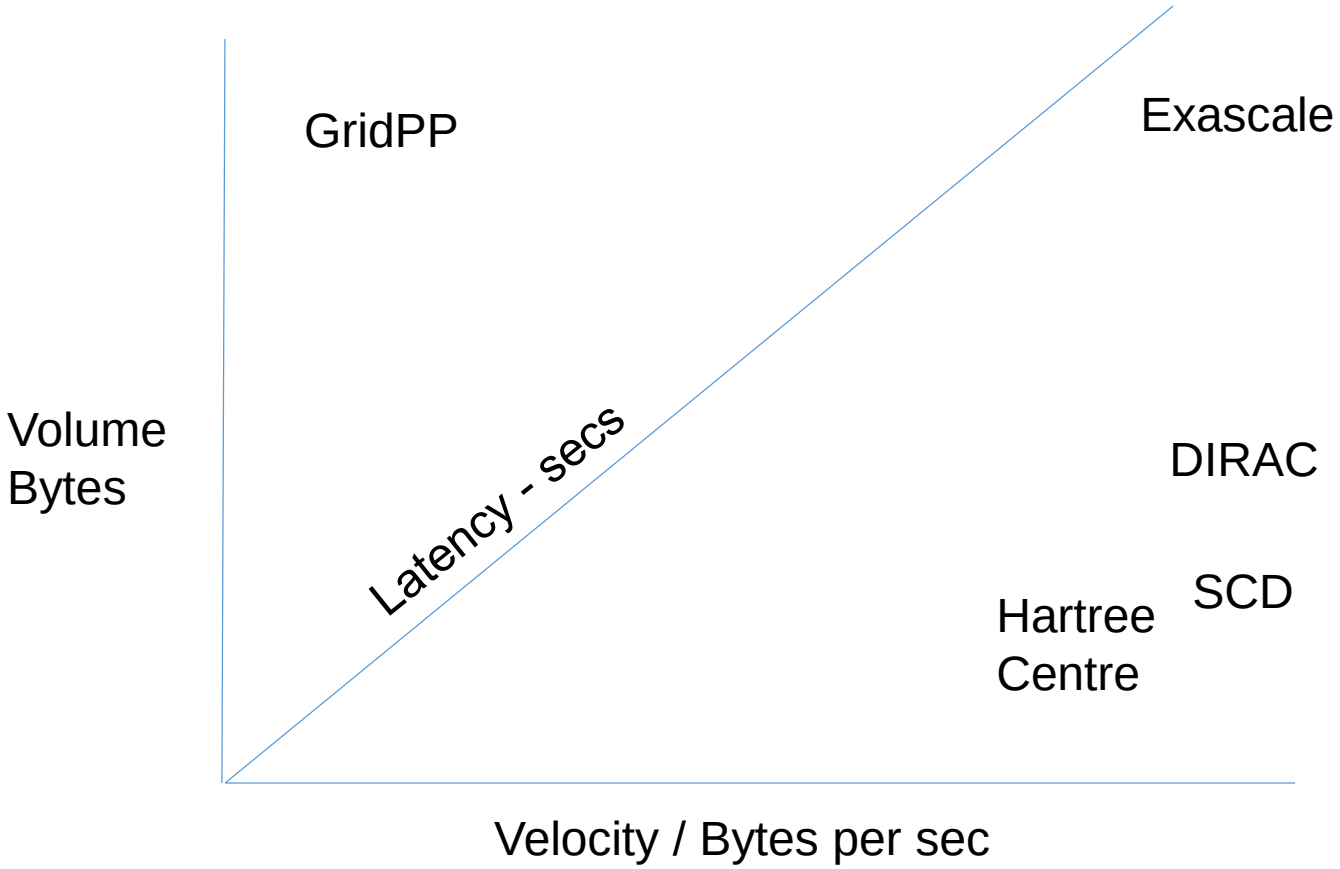
- A google search is probably the most common example
- Communications are key to parallelism

Parallelism	Workflow	Required Latency (microsecs)	Required Bandwidth (Gbs)	System Type	Inter-node communication level
Embarrassing loose	Parameter search Model fitting	100 10	1 10-100	Cluster	rare
Weak	CFD, SPH	1	50-100	Cluster	occasional
strong	LQCD	<1	100	Cluster	As busy as intra node
Cache Coherence	Strings?	<0.2	400	Single Node	As busy on the CPU

The Challenges We Face in System Design



Volume and Velocity and Latency



Common Model of understanding Problem need

A third axis is needed – latancy (delay)

Software

- **Software: STFC** needs a software infrastructure strategy that reflects both the central role played by software in the delivery of research and innovation and the diversity of the research code base
- It is essential that codes become more agile, able to harness the power of evolving hardware architectures.
  - Accelerators
- This agility is needed because the pace, and nature, of hardware evolution means that it is no longer possible to programme code for long term, un-modified usage. In many areas, software development work is already overdue to support the next generation of science calculations.
- Not just about applications Supporting accelerators
  - E.g. Networking programming to create new cheap SSD arrays

Accessibility

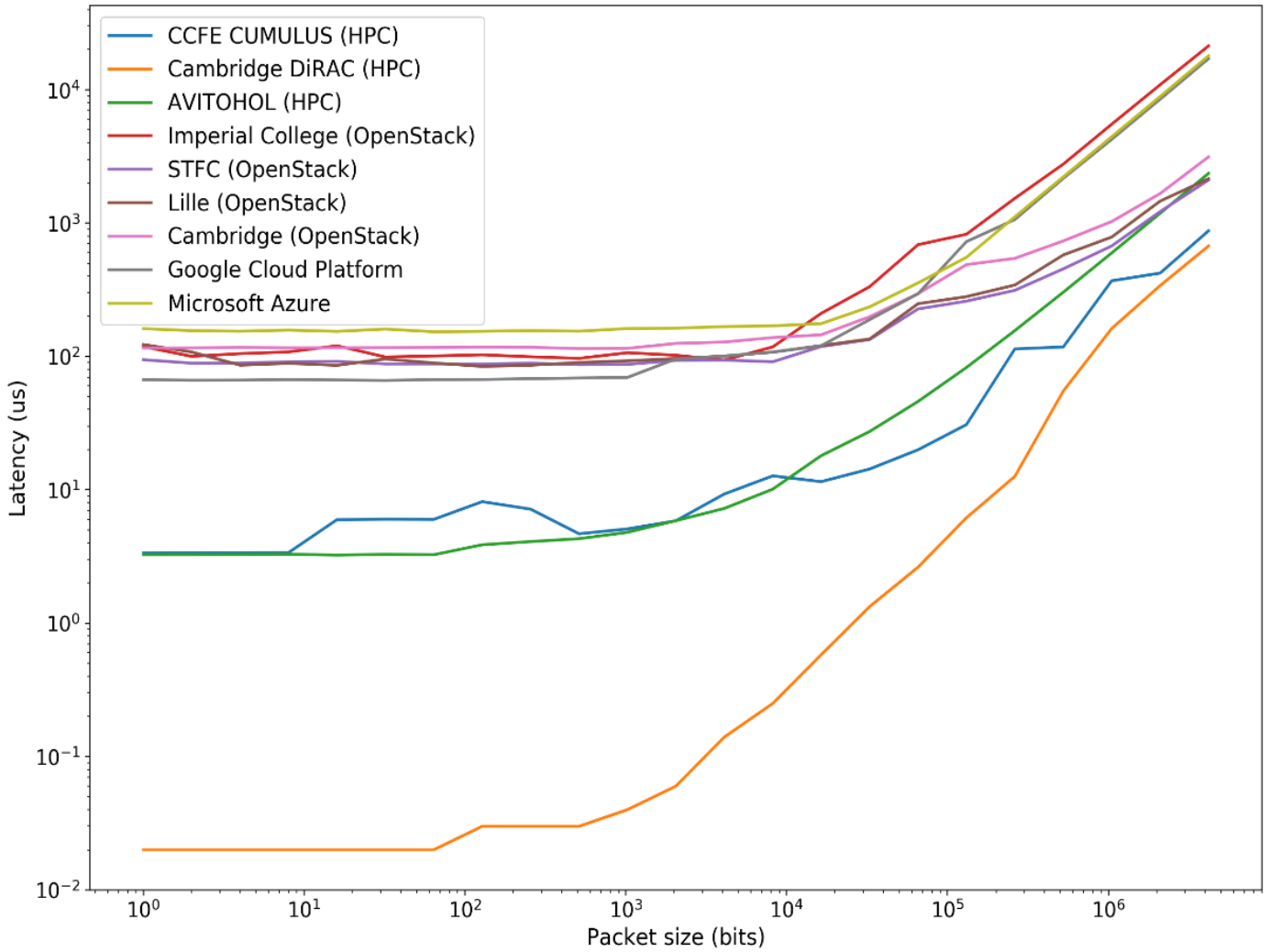
How does my application work – Creating a Virtuous Software Ecosystem & Community

- The LQCD community have been able to design in a quantitative manner systems based upon the operational requirements of key applications.
- We need a small-ish cadre of people who understand how key applications and libraries actually operate quantitatively on hardware
  - Just like for telescopes and accelerators
  - Why should we be different.
- They will be supported by people who write the software that makes the components actually work
- Data Motion is key to qualitatively understanding of how systems and apps work together
- Most people only need this level of understanding
  - Need to know the hardware characteristics and
  - how they work together.
  - What has the compiler done to my code and how it is used by the components
  - We should know how our experimental platforms work
  - Personally I'd like to see computational based research degrees describe the equipment and application performance in the same way that we do for experiments..

Cloud for massive parallel computing

- Cloud middleware development is needed to support resource-sharing and ensure optimal matching of science with supercomputing resources.
  - Portable workflow
  - Joining resources together
  - Not being limited by the batch queue
- Continued support for existing development work e.g. the OpenStack and containerisation work by IRIS/DiRAC/Cambridge/ is necessary to facilitate greater workflow movement between services and help communities from being locked into incompatible systems.
- We need to maintain a watching brief on the development of commercial and community cloud offerings to assess their relevance for supercomputing resource provision.
- However need to address low latency requirement

Simulations done by Andrew Lahiff CCFE



Need for Physical and Virtual Architecture Communities

- Know who is good at what
- Support key projects and people
- Co-Design projects are a good way of understanding how things work together and create communities around a particular activity/technology
- We've been quietly doing this in DiRAC for 9 years
  - Need to add accelerators and cloud technology
  - Both in progress
- We need to now address the application understanding

Forward Look - All

- Growing demands for modelling and simulation from the Facilities, Programmes (DiRAC) and the Hartree Centre
- Faster data rates: bigger data – more processing (tomography etc)
- Challenge of evolving codes to exascale architectures
- Increased use of ML and AI, as an alternative to ‘traditional’ processing, in combination, and as a driver of Supercomputing in itself
- Greater use of AI in simulations and data modelling
- Use of new storage, networking, memory, Processor+Accelerator, middleware and software technologies
- Use of Cloud/Grid technologies to design new workflows and make it easier to combine different resources for our numerical experiments
- New ways to for researchers access our systems e.g. Jupyter notebooks
  - Greatly expand the use of systems
- Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery – AI3SD

Our mission

Transforming UK industry by accelerating the adoption of high performance computing, big data and cognitive technologies (AI, ML, DL) through challenge-led research and innovation

Offering

Based on the five foundations of UK Industrial Strategy:

- Ideas - Collaboration across industry, UKRI and academia. Centre of excellence in translating research into innovative solutions

- People – Critical mass of cross-functional teams combining chemistry, materials, engineering and life science domain expertise (internal or via collaboration with SCD or academia) , with research software engineering, numerical and computational methods, working together for economies of scale and scope

- Infrastructure - Dedicated platform(s) for HPC, HPDA and AI applications and emerging technologies, providing research and innovation capabilities

- Business Environment – Driving productivity, increasing collaboration e.g. across supply chains, support to growing businesses

- Place – Sci-Tech Daresbury, Northern Powerhouse,
  fast growing digital cluster of local companies

Hardware Platforms

- Scafell Pike (funded from Hartree Phase 3) – Bull Sequana X1000 supercomputer
  - ~4 Pflop/s | Intel Xeon and Xeon Phi, GPUs
  - 25,728 Intel SkyLake cores
  - 55,680 Intel Xeon Phi cores

- JADE – (funded by EPSRC Tier 2, owned by Oxford)
  - Bull Atos NVIDIA DGX1 Deep Learning supercomputer
  - 176 NVIDIA V100 GPUs | Deep Learning frameworks
  - 630,784+ CUDA Cores

- Panther and Paragon (funded from Hartree Phase 3)
  - 512 POWER8 cores and 64 NVIDIA Kepler K80 GPUs (Panther)
  - 656 POWER8 cores and 82 NVIDIA Pascal P100 GPUs (Paragon)

- (Many data science and AI projects use commercial cloud e.g. Azure)

What we do

- **Collaborative R&D**
  We build a team to deliver a solution to a particular challenge

- **Platform as a service**
  Pay-as-you-go access to our compute power

- **Creating digital assets**
  License industry-led software applications we create with IBM Research (IROR programme)

- **Training and skills**
  Run specialist training courses and events

Scientific Computing Department

“to maximize the impact of scientific computing through our expertise, leadership and collaboration.”

Designing, deploying and operating large and complex computing and data systems
Supporting the research life-cycle by extracting insights and value from data

Creating algorithms and software to exploit future research computing infrastructure

Providing cross-domain expertise to develop, innovate, and sustain software, and related digital assets for research

Hardware

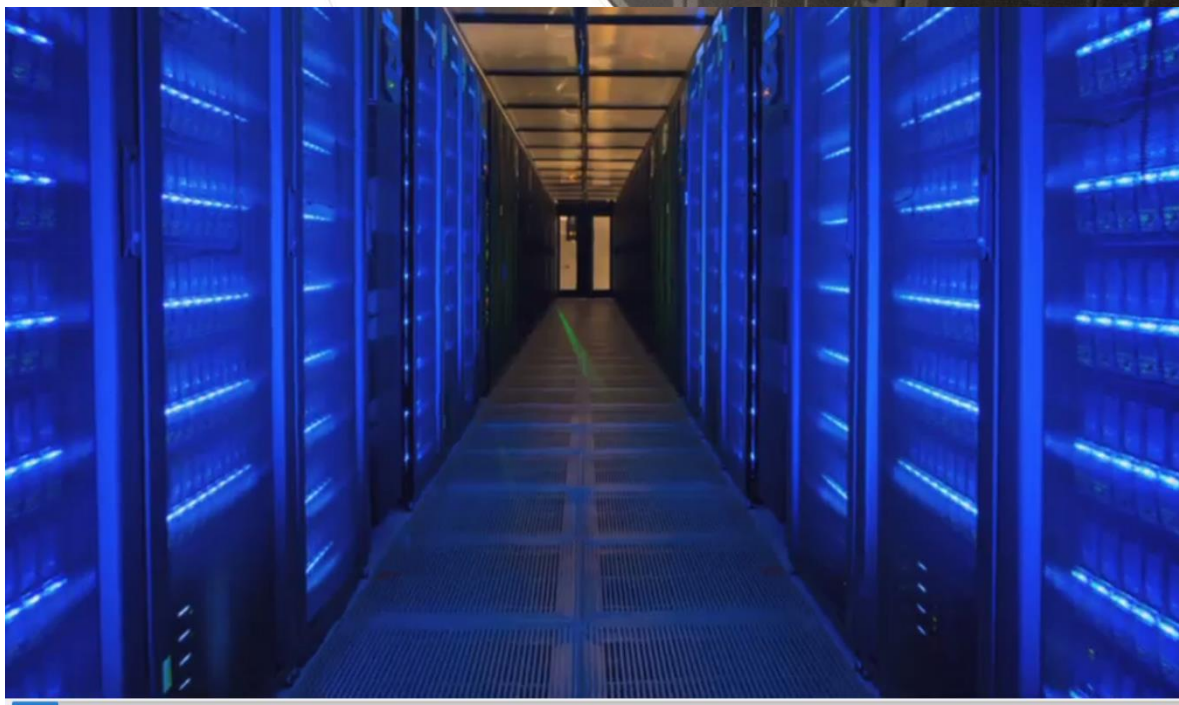
SCARF – open to STFC departments, facilities and their users.

General Purpose CPU and GPU cluster.
Demand **significantly** exceeds capacity

Annual, incremental investments since 2004

570 Nodes over 5 generations
12k CPU. Infiniband
30 GPU-nodes
1PB Panasas storage
2PB GPFS

+ PEARL Deep Learning Service – 2x NVidia DGX-2
+JASMIN Super Data Cluster (NERC communities)



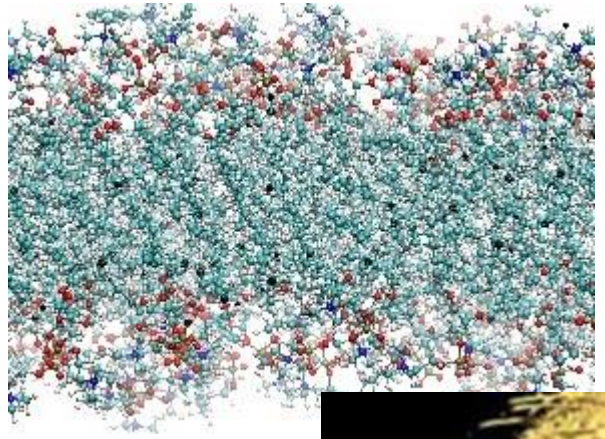
Software

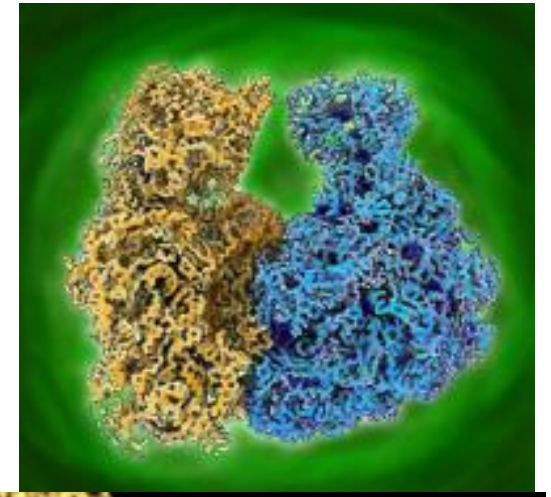
CoSeC

- Support for Collaborative Computing Projects (CCPs) and High End Consortia (HECs)

- computational scientists and specialists research software engineers who combine science domain knowledge with software engineering skills

- Wide range of areas covering bio, chemistry, physics and engineering



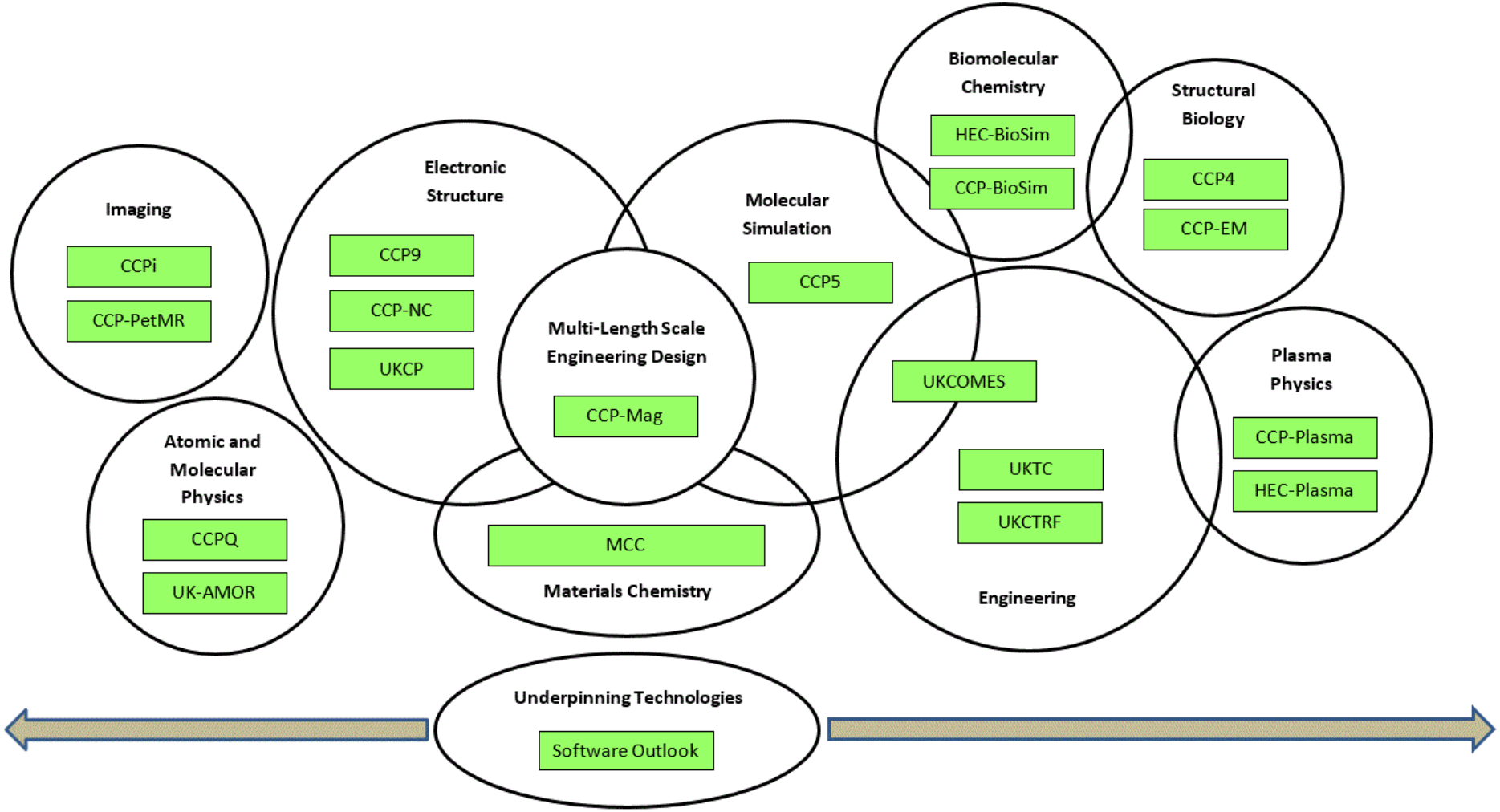


Communities

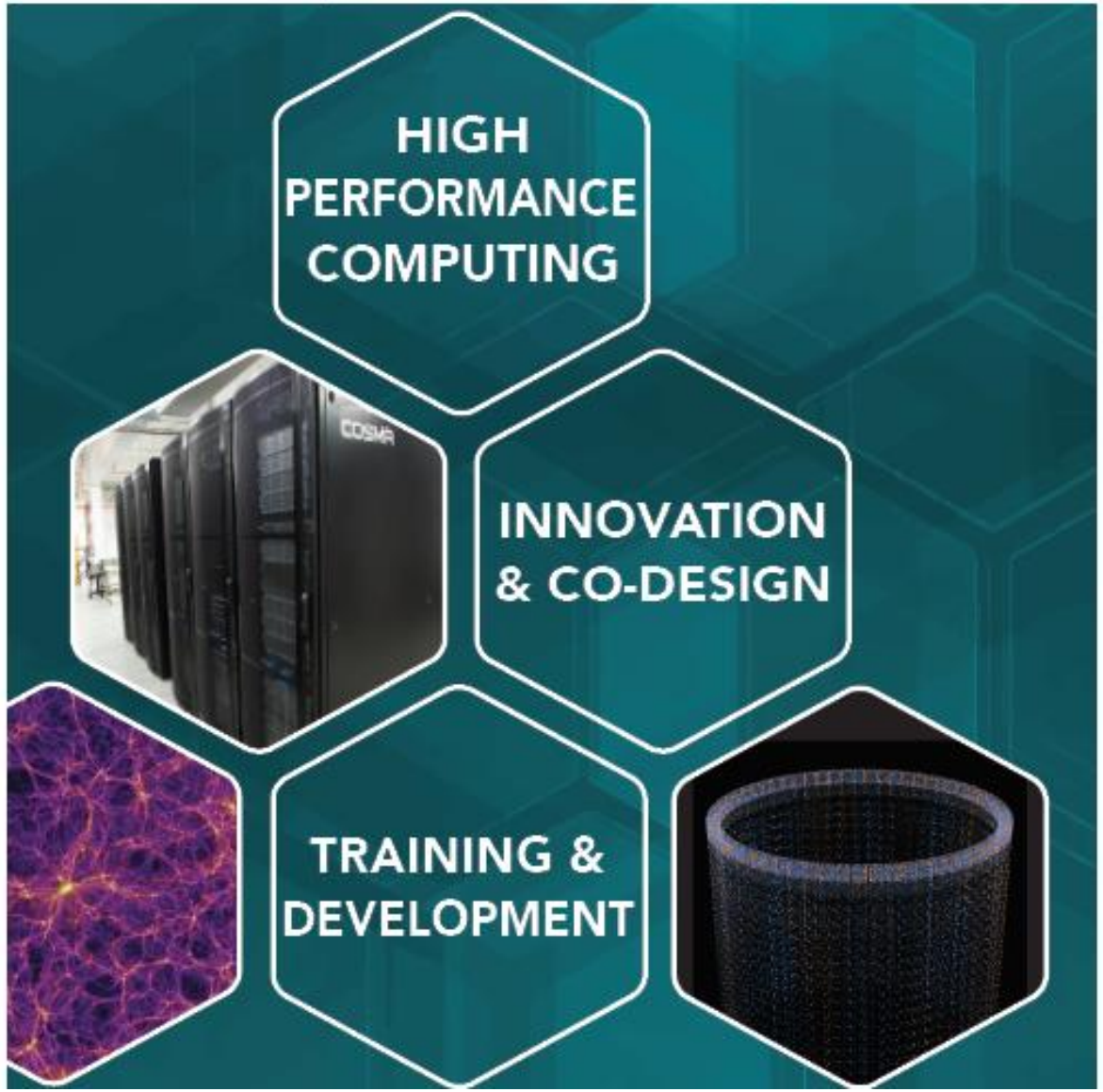
CCPs and HECs

Facilities staff and users
- ISIS, CLF, Diamond

Broad range of science areas
- Materials science
- Bio
- Imaging
- Engineering



DiRAC


HIGH PERFORMANCE COMPUTING
INNOVATION & CO-DESIGN
TRAINING & DEVELOPMENT

DiRAC

Diverse science cases require heterogenous architectures

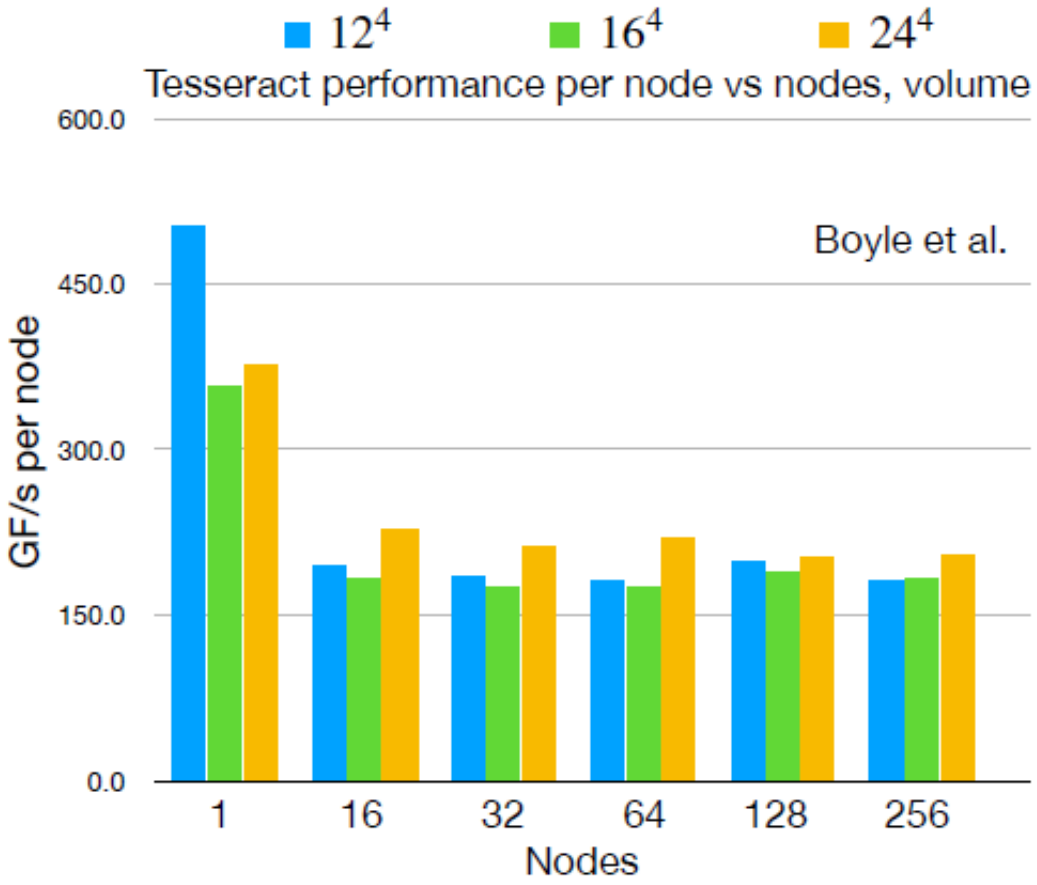
Extreme Scaling "Tesseract" (Edinburgh)	Data Intensive "DIaL" and "CSD3" (Leicester & Cambridge)	Memory Intensive "COSMA" (Durham)
		
2 Pflop/s to support largest lattice-QCD simulations	Heterogeneous architecture to support complex simulation and modelling workflows	230 TB RAM to support largest cosmological simulations

DiRAC

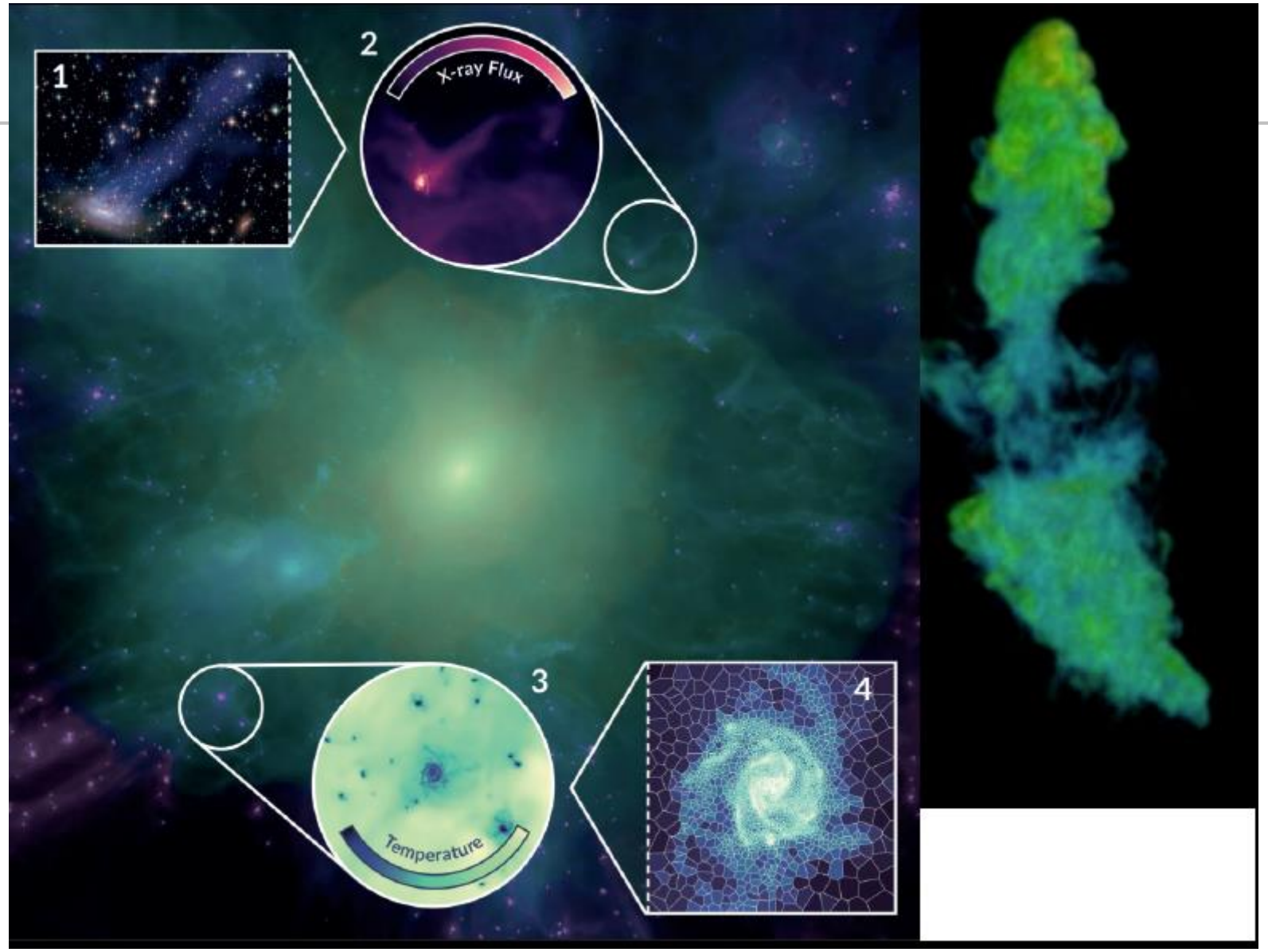
Extreme Scaling: Tesseract

- Example of co-design in action - matching application to network topology
- Embed 2^n QCD torus inside hypercube so that nearest neighbour comms travel single hop: 4x speed up over default MPI Cartesian communicators on large systems

$\Longrightarrow$ customise HPE 8600 (SGI ICE-XA) to use 2^4 nodes per leaf switch



- 16 nodes (single switch) delivers bidirectional 25 GB/s to every node (wirespeed)
- 512 nodes topology-aware bidirectional 19 GB/s
- 76% wirespeed using every link in system concurrently
- Tesseract: 1468 Intel Skylake 24-core nodes (>1.8 PF); 32 Nvidia V100 GPUs


1
2
X-ray Flux
3
Temperature
4

DiRAC

Innovation and co-design

- Lowering the bar for academic engagement with industry
  - Crucial to maximise science productivity of services and secure funding
- Focus on projects that benefit both science programme and industry
- Proof-of-concept systems:

- Pilot systems:

- Co-design from chip-level to system level:

DiRAC

Training

- DiRAC provides ***access*** to training from wide pool of providers
  - Workshops: Software Design & Optimisation; MPI programming
  - Hackathons and CodeCamps:
  - 6-month innovation placements for PhD students and early-career PDRAs
- Facility training goals:
  - maximise DiRAC science output through more efficient software
  - flexibility to adopt most cost-effective technologies
  - future-proofing our software and skills
  - contributes to increasing skills of wider UK economy

DiRAC

- Delivering HPC resources for the UK theory communities in particle physics, astrophysics, cosmology and nuclear physics
- Our goal is to maximise the science our researchers can carry out
- This is achieved through:
  - Engagement in hardware and software co-design
  - Enhanced training
  - Research software engineering support

(Better systems) + (Better software) = Better science

@DiRAC_HPC
dirac.ac.uk

A Final Thought - AI

- Current AI applications will give me 42
  - But what does that mean?
  - What do we understand by this?
  - How did it get this answer?
  - In what sense is this answer usable?
- The recent UKRI AI workshops identified Explainable AI as a requirement for both research and application
- To implement this we need to go from Deep Thought to the Earth in terms of system size and design
- Laptop to Exascale – let’s hope the Planning Department has dealt with it…..