

The DiRAC-4 design process: from science to services

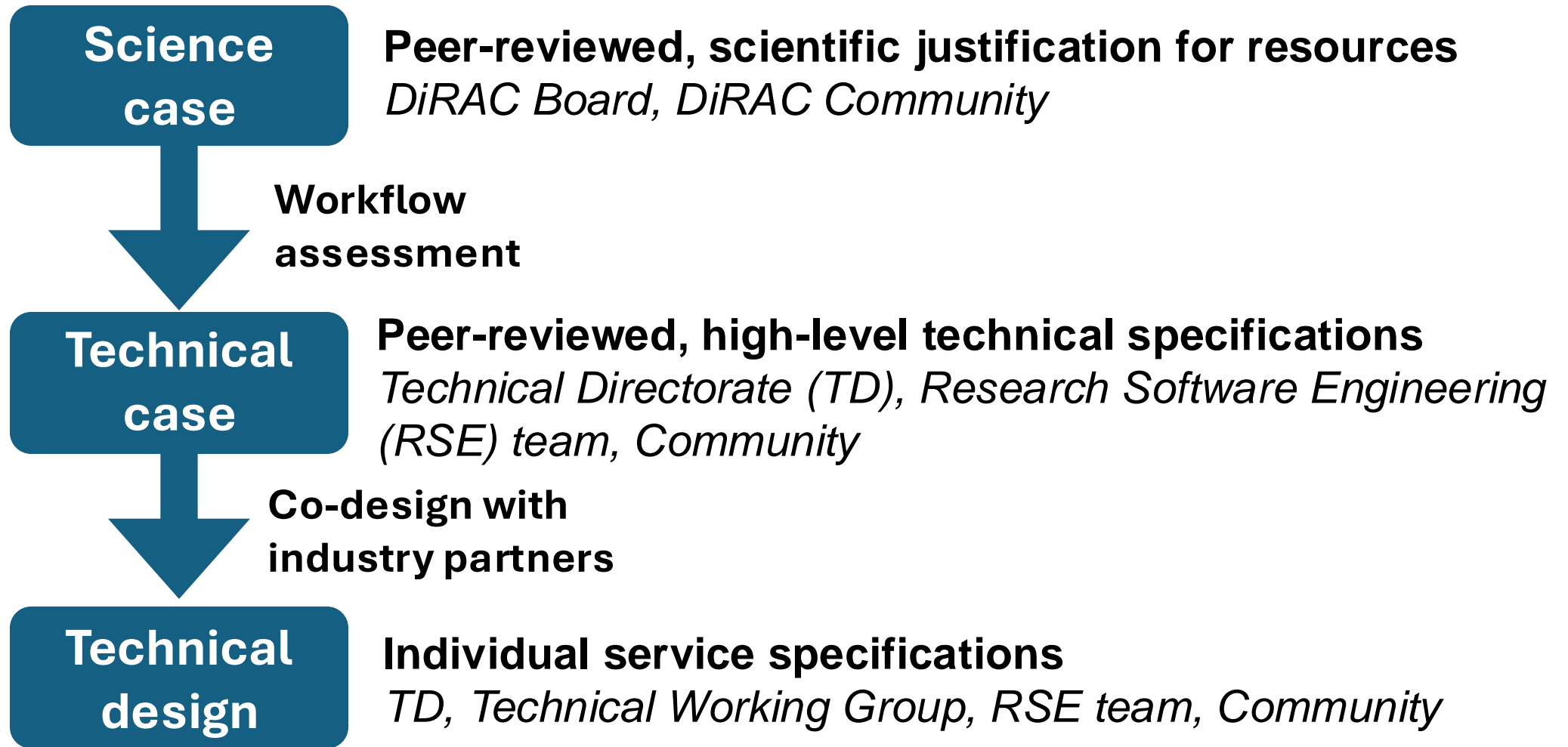
Mark Wilkinson
Director, STFC DiRAC HPC Facility

DiRAC-4 Design Workshop
York, 20th March 2025

Introductions

- Technical Directorate
 - In-person: Alastair Basden (Co-Chair), Antonin Portelli (Co-Chair), Kieran Leach, Jeremy Yates
 - Remote: Biagio Lucini
- Technical Working Group
 - In-person: Paul Walker
 - Remote: Wojciech Turek
- RSE team
 - In-person: Simon Burbidge (Co-lead), Ilektra Christidi (Co-lead), Mashy Green, Gokmen Kilic, Kacper Kornet, Miren Radia
 - Remote: Asif Muhammad

DiRAC Co-design Process – a multi-stakeholder partnership



- Community involved in all three phases of the co-design process

Co-design benefits & the importance of people

- Investment in people is vital for productive HPC services
- DiRAC services require specialist technical support for hardware and users
- RSE team supports code improvement and re-factoring, energy efficiency, co-design, procurement & training

Evolution of Grid code (Boyle et al.) performance on Tursa relative to Tesseract

Stage	1 node	% inc.	16 nodes	% inc.	speed up 512 tess
Measured	9.2	-	5.3	-	1.1
Committed	9.2	-	5.83	10%	1.22
Acceptance	9.65	5%	6.15	16%	1.28
Commissioning	12	30%	8.8	66%	1.83
Peak	12.9	40%	9.9	87%	2.06

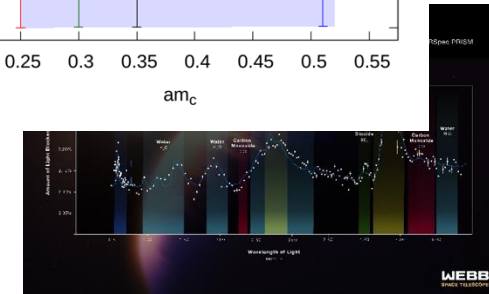
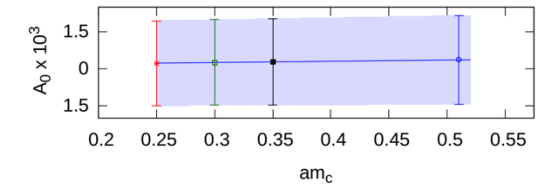
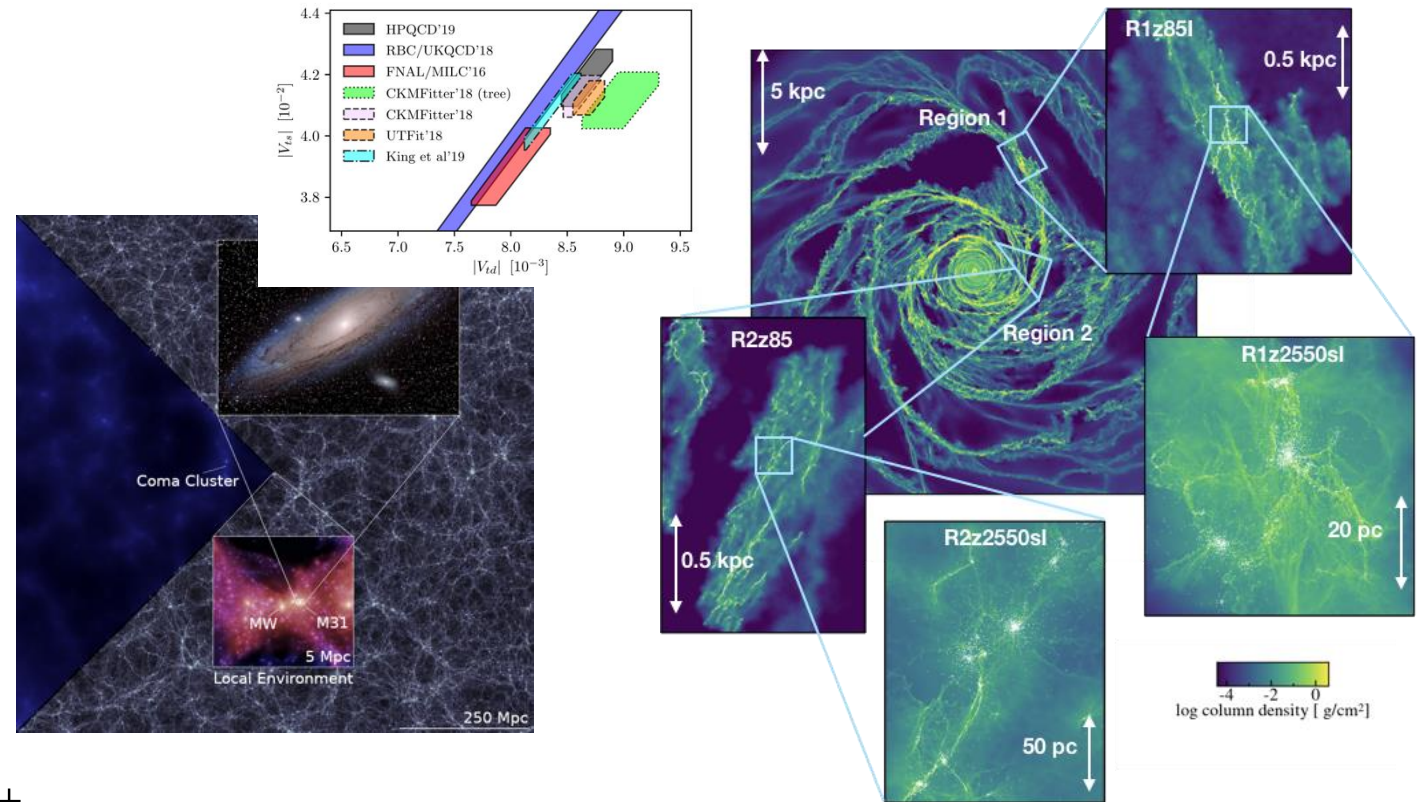
*James
Richings et al.*

- Tursa Extreme Scaling service (DiRAC@Edinburgh) provides ~9x the performance of its CPU-based predecessor for lattice QCD codes but uses just 0.8x the power.
- Cosma8 Memory Intensive service (DiRAC@Durham) is 4x more efficient for cosmological simulations than comparable systems in Europe – DiRAC services allow us to remain internationally leading
- Net Zero: Clocking down Tursa GPUs: ~15% energy saving with only ~5% performance loss for Grid code

DiRAC Science Programme 2024-28

DiRAC

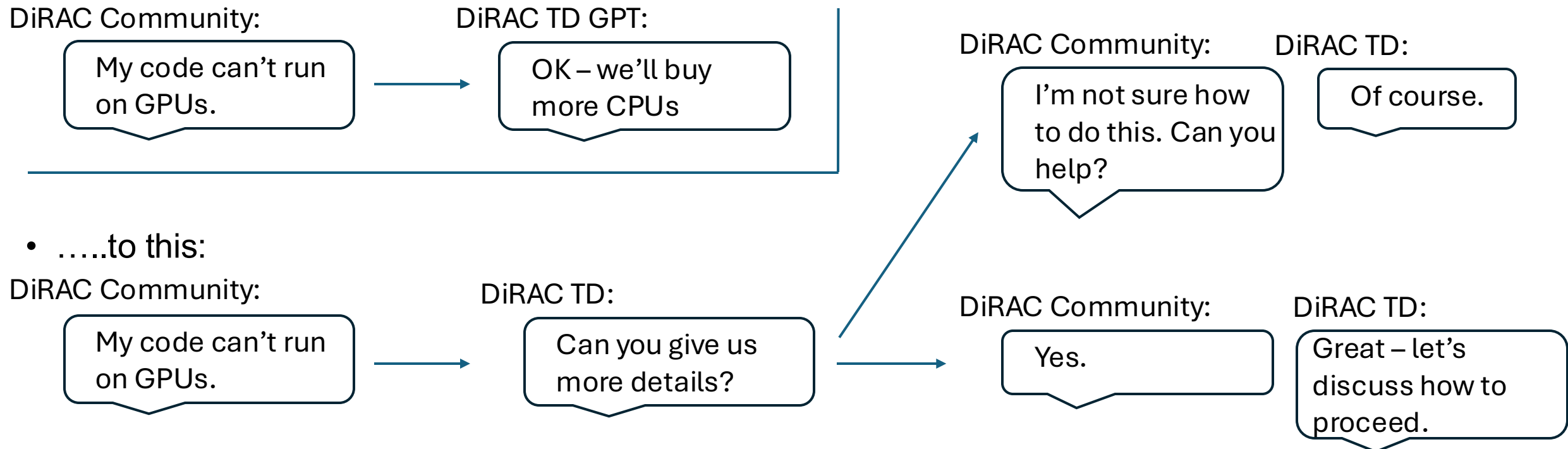
- Capability calculations include:
 - Galaxy formation
 - Lattice Quantum Field Theory
- Data Intensive calculations:
 - Gravitational waves
 - Gaia modelling
 - Precision cosmology
 - Planetary atmospheres
- Data challenges growing rapidly
 - Individual simulations generate 10Pb+
- Increasing use of AI/ML techniques to enhance simulation methods
 - At least 50% of fields are using or exploring AI over next 4 years.
 - DiRAC simulation data can also be used to train AI models.
- Action: check science case for accuracy – tell us now if there are issues



“Trust, but verify”

Evidencing the DiRAC-4 technical case

- Evidence-based development of compute services is at the core of DiRAC’s design philosophy
- The case for DiRAC-4 will be carefully scrutinised by funders (STFC, UKRI DRI, DSIT) and must be as robust as possible
- Assertions made without data to back them up will undermine the whole case
- Need to move conversations from this.....



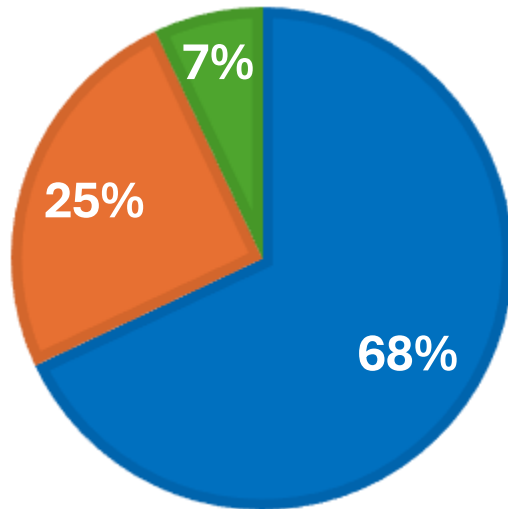
GPU Feasibility studies

- New activity proposed and led by the DiRAC RSE team with approval from DiRAC Board
 - 3-month RSE engagements to explore GPU
 - Reports will
 1. Identify blockers
 2. Determine what is required to overcome these
 3. Estimate effort required
 - Reports will be technically specific about algorithmic and/or implementation issues which need to be addressed to allow a move to GPU
 - Mix of targeted interventions by DiRAC and call for proposals
 - Will be used to quantify levels of effort required for code re-factoring across the community
 - Goal: significantly increase the fraction of code base that is GPU ready prior to any DiRAC-4 deployments
 - More information on this call will be available soon
- Question for today: if a GPU version of your code is possible in 5 years, what would make it possible in 2 years?

DiRAC User Survey 2024

SURVEY RESPONDENT'S CODE BASE

- CPU-Centric
- CPU+GPU enabled
- GPU only

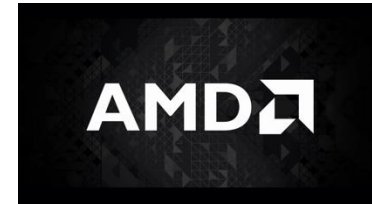


DiRAC Performance Portability Training Programme

- Currently under development
- Industry partners:



Broad overview of different GPU pathways available



Advanced OpenMP and introduction to HIP



Hardware agnostic approach to porting code with oneAPI

www.dirac.ac.uk/training

Today's workshop

- **Goal:** identify remaining gaps in evidence base and identify ways to gather the required information
- **NB:** Main focus today is ***not*** GPU/CPU
 - We need more quantitative information to progress this discussion
- Areas for discussion today:
 - Constructing the DiRAC-4 Technical Case
 - Scaling requirements
 - Requirements for RAM and storage
 - Post processing, user-centric services, and data management
 - Information gathering to evidence requirements
- Enjoy this opportunity to shape the future of DiRAC computing and the UKRI Digital Research Infrastructure....