

Funding and delivering novel hardware

Alastair Basden
DiRAC / Durham

Alastair Basden
DiRAC / Durham University



DiRAC
High Performance
Computing Facility



Changeable compute

- The compute landscape changes rapidly
 - Even during relatively static periods
- Cost-effective systems will use cost-effective hardware
 - Codes must be tested and adaptable
- Unmitigated risk is unacceptable for large capital purchases

Solution

- Code porting
- Profiling
- Benchmarking

The HPC Hardware Lab @Durham

- Providing researchers with access to new/novel technologies
- Experimental production environment
 - Rough edges
 - Mistakes are allowed
 - Some work may be required
- Aim to advise on purchases of new systems
 - And provide access to prepare for these

Aim

- Place hardware rapidly into production
 - New, novel, cutting edge, prototype and pre-production
- Easy access for users
 - Reducing entry barriers
 - Single account for multiple hardware components
- Flexibility of approach
 - Dedicated benchmarking/profiling time can be requested
- Aim for scalability testing
 - Beyond a single node
- Infrastructure is within scope:
 - “Users” can be RTPs

Funding sources

- No single funding source
 - Internal Durham
 - ExCALIBUR H&ES
 - ExCALIBUR-2 should focus on high TRL
 - DiRAC
 - EPSRC grants
 - SKA
 - Dell, AMD, Mellanox/NVIDIA
 - Others

Available testbeds

- CPU: Intel, AMD, ARM, NVIDIA
 - Emerald Rapids, Genoa/Bergamo, Grace (and others)
- GPU: AMD, NVIDIA, Intel
 - MI300X, Hopper, Ponte Vecchio (and others)
- Networking: InfiniBand, Ethernet, Rockport
 - 800G switch (and others)
- DPUs: BlueField1&2
- Cooling: Immersion, DLC, Rear door, reuse
- Storage: Lustre, Ceph, DAOS, VAST, StorJ, ...
- Composability
- Quantum

Hardware lab development

- New/novel component identified
 - Use case investigated
 - Will it be useful and of interest?
 - How much potential does it have?
- Route to funding identified
 - Worth being inventive
- Kit ordered and delivered
- Racking up and installation
 - Functionality testing
- Go-live and announcements

Key messages

- A funding stream would be good
- Relevant for future procurements:
 - Allowing codes to be tested and benchmarked
 - Maximising science per CO2 and/or per £
- Development of flexible UK skills base
 - Encouraging portability of codes

Watch out for...

- MI300A system
- Ultra Ethernet fabric
- AI accelerators:
 - Gaudi-3, UntetherAI
- Turin CPUs