Science and Technology Facilities Council

Conference Proceedings
DL-CONF-2020-001

Computing Insight UK 2019

Manchester Central Convention Complex, UK

5th-6th December, 2019

D Jones (editor)

June 2020

Enquiries concerning this report should be addressed to:

Chadwick Library
STFC Daresbury Laboratory
Sci-Tech Daresbury
Keckwick Lane
Warrington
WA4 4AD

Tel: +44(0)1925 603397
Fax: +44(0)1925 603779
email: librarydl@stfc.ac.uk

Science and Technology Facilities Council reports are available online at:
https://epubs.stfc.ac.uk

ISSN 1362-0223

COMPUTING INSIGHT UK 2019

Computing The Future

5 - 6 DECEMBER 2019
MANCHESTER, UK

Science & Technology Facilities Council | UK Research and Innovation

Conference Proceedings

www.stfc.ac.uk/ciuk

Computing Insight UK (CIUK) 2019 took place on 5-6 December 2019 at Manchester Central Convention Complex. These proceedings are a record of the presentations and posters from the Conference.

The CIUK Organising Committee would like to thank the exhibitors, sponsors, presenters and attendees who help to make the Conference a continued success.

Sponsors

GOLD SPONSORS	SILVER SPONSORS	BRONZE SPONSORS
CRAY a Hewlett Packard Enterprise company	DDN AI · BIG DATA · HPC	AMD
Atos	SUSE We adapt. You succeed.	BOSTON Servers I Storage I Solutions
VERNE GLOBAL LEADERS IN HIGH PERFORMANCE COMPUTING	vesper	HAPPYWARE IT's all you need!
Hewlett Packard Enterprise	NGD THE DATA CENTRE SUPER POWER	Orchestrating a brighter world NEC

Media Partner

SCIENTIFIC COMPUTING WORLD

Contents

CIUK 2019 Introduction

Now in its 29th year the event formerly known as the Machine Evaluation Workshop combines an exhibition of the latest High Performance Computing (HPC) hardware and software with a programme of presentations from users who provide delegates with an overview of their use of HPC in real life situations. Attendees are able to hear success stories and problems encountered as projects make use of the latest HPC tools and equipment, as well as the solutions implemented to make the projects successful.

Through its exhibition, Computing Insight UK (CIUK) allows attendees to communicate with a wide and varied selection of hardware and software vendors and resellers under one roof.

We take pride in the fact that CIUK is one of a very small number of HPC events in the UK that provides such a wide and varied programme of events within one conference. Now attracting around 400 attendees each year we are confident that CIUK is the UK's premier annual HPC conference.

Computing Insight UK 2019 took place at Manchester Central on 5-6 December. The theme for the conference was **"Computing the Future"** with the event divided into two sub-themes. The first day (Thursday 5 December) focussed on "Computing Today" with sessions on "Cloud Computing Today", "Software Development Today" and "Data Science - Convergence of AI and HPC". The second day (Friday 6 December) looked at "Computing Tomorrow and Into the Future" with sessions on "Hardware Towards Exascale", "Software Development Towards Exascale" and "Quantum Computing".

The Jacky Pallas Memorial Award

Earlier this year we lost an incredibly important member of the CIUK Scientific Advisory Committee with the sudden and unexpected passing of Jacky Pallas at the age of just 54. Jacky was head of e-Research at King's College London and for the last three years had been an active and vocal member of the CIUK SAC, helping to shape the direction our event has taken and pushing through many positive changes, whilst championing diversity and the inclusion of young researchers. In her memory, and in recognition of her passion for our conference, we decided to introduce an annual award that will highlight the work of an early career researcher and will allow the award winner a slot in the main programme at CIUK.

We received a number of nominations for this award and after lengthy deliberation; we were delighted to introduce the winner of the inaugural Jacky Pallas Memorial Award - Demi Pink from King's College London.

Demi is a PhD student in the Department of Physics at King's College London. Whilst she studied for her undergraduate degree in Chemistry at the University of Leicester, she received the OUP Achievement in Chemistry Prize before graduating with 1st Class Honours. Following this, she joined the BBSRC funded London Interdisciplinary Doctoral Training Program where she undertook rotation projects in cell biology and biophysics before beginning her PhD under the supervision of Dr Chris Lorenz and Prof. Jayne Lawrence. Her work uses molecular dynamics simulations and small angle neutron scattering to investigate the self-assembly of lipid-based drug delivery vehicles and their encapsulation of small hydrophobic drug molecules.

CIUK 2019 Programme

DAY 1 - Thursday 5 December 2019

TIME	MAIN PROGRAMME	BREAKOUT SESSIONS
From 08:30	**REGISTRATION OPEN (Main Foyer) EXHIBITION OPEN (Gallery) RESEARCH CENTRE ZONE OPEN (Main Foyer)**	
09:15 - 09:30	*Welcome* **Tom Griffin (Director, Scientific Computing, STFC)**	
09:30 - 10:00	*Movement without Disruption - or - How to Move Your Multi-Million Pound HPC Cluster without Your Users (Hardly) Knowing* **Cliff Addison (University of Liverpool)**	
10:00 - 10:30	*JASMIN and the Evolution of Cloud-Hosted Data Analytics Platforms for the Environmental Sciences* **Phil Kershaw (JASMIN)**	*Power AI User Group* *10:00 - 11:30* *Spectrum Scale User Group* *11:30 - 13:00*
10:30 - 11:00	*Cagliari Airport AI in Fog to Cloud HPC* **Jens Jensen (STFC)**	
11:00 - 11:30	**REFRESHMENTS**	
11:30 - 12:00	*Research Software Engineering Impact Showcase* **James Grant (University of Bath)** A session consisiting of a series of short talks from the RSE community using case studies to highlight impact followed by a panel discussion on how we can measure and promote the capability of RSEs to produce impact. **Presentations from James Grant (University of Bath), Dave Meredith (STFC) and Andy Turner (EPCC)**	
12:00 - 12:30		
12:30 - 13:00		
12:30 - 14:30	**LUNCH**	

14:30 - 15:00	*Machine Learning as a Cheaper Alternative to HPC Approaches in Science* **Jeyan Thiyagalingam (STFC)**	*Utilising the Open Source OpenFlightHPC Project for HPC Workflow Design and Implementation* *14:00 - 16:00*
15:00 - 15:30	*Linking National Imaging Facilities with HPC and Research Software Engineering centres – and their use on a Big Data problem* **Martin Turner (University of Manchester)**	
15:30 - 16:15	Research Centre Zone - Lightning Talks **Supercomputing Wales, DiRAC, ARCHER, Cirrus, ICHEC, N8CIR, the Materials Modelling Hub, Isambard, JASMIN**	
16:15 - 17:00	**REFRESHMENTS**	
17:00 - 18:00	*Application Performance on Multi-Core Processors: Performance Analysis of the AMD EPYC Rome Processors* **Martyn Guest (ARCCA, Cardiff University)**	
18:00 - 19:00	**Keynote Presentation** **Debora Sijacki (University of Cambridge, Institute of Astronomy) Winner of the 2019 PRACE Ada Lovelace Award for HPC** **"*Towards next generation computing in cosmological simulations: prospects and challenges*"**	

DAY 2 - Friday 6 December 2019

TIME	MAIN PROGRAMME	BREAKOUT SESSIONS
From 08:30	**REGISTRATION OPEN (Main Foyer) EXHIBITION OPEN (Gallery)** **RESEARCH CENTRE ZONE OPEN (Main Foyer)**	
09:30 - 10:00	*NVMe over PCIe Fabrics Using Device Lending* **Jonas Markussen (Dolphin Interconnect Solutions)**	
10:00 - 10:30	*On the Road to ExaScale - New Storage Technologies to Support ExaData* **Torben Kling Petersen (Cray)**	*NVIDIA Deep Learning Institute "Fundamentals of Accelerated Computing with CUDA C/C++"* *The CUDA computing platform enables the acceleration of CPU-only applications to run on the world's fastest massively parallel GPUs.* *Upon completion of this day-long workshop, attendees will be able to accelerate and optimize existing C/C++ CPU-only applications using the most essential CUDA tools and techniques whilst achieving an industry-recognised certification (subject to passing the assesment)*
10:30 - 11:00	**The Jacky Pallas Memorial Presentation** ***On the Structure of Lipid-Based Nanoparticles for Drug Delivery*** **Demi Pink (King's College London)**	
11:00 - 11:30	**REFRESHMENTS**	
11:30 - 12:00	*Readying an Industrial CFD Code for Pre-Exascale* **Yvan Fournier (EDF)**	
12:00 - 12:30	*Improving Application Performance: Mellanox's Collaboration with UK HPC* **Richard Graham (Mellanox Technologies)**	
12:30 - 13:00	*Towards Understanding Exascale IO needs – insights from LASSi on ARCHER* **Karthee Sivalingam (CRAY)**	
13:00 - 14:30	**LUNCH**	
14:30 - 15:00	***UKRI E-Infrastructure Roadmap: Next Steps*** **Mark Thomson (STFC)**	
15:00 - 15:30	*Quantum Computing for the 21st Century* **Kate Marshall (IBM)**	

15:30 - 16:00	*Quantum Computing Ambitions from University of Liverpool* **Michael Bane (University of Liverpool) and Shane Rigby (Atos)**	
16:00	**CIUK 2019 CLOSES**	

CIUK 2019 Presentations

Cliff Addison

University of Liverpool

Movement without Disruption - or - How to Move Your Multi-Million Pound HPC Cluster without Your Users (Hardly) Knowing

Cliff Addison was dragged into parallel computing in 1986 whilst at the Christian Michelsen Institute in Bergen Norway. His initial parallel computing efforts there targeted the oil and gas industry, including a parallel pre-stack migration demonstrator and the first steps towards a commercial oil reservoir simulator. He moved to Liverpool in 1989 as part of what became the Institute of Advanced Scientific Computing (IASC). He was involved in several Esprit and Eureka projects, including GENESIS, Supernode-2, PULSAR and PARSIM. When the University decided that there was no future in HPC in 1996, he moved to the Fujitsu European Centre for Information Technology (FECIT).

Abstract

One of the challenges in moving an HPC system from one data centre to another is the extended downtime that users have to experience. A solution is to provide some form of HPC system in the cloud.

Over a trial period in June and a move of two weeks in October 2019, cloud Barkla mirrored the mission-critical components that are the lifeblood of current Liverpool research computing services.

This talk will cover off four key outcomes of utilizing public cloud in what, essentially, is a planned Disaster Recovery scenario. This talk is ideal for those looking at public cloud for Disaster Recovery, are wanting to know how public cloud supports on-premises expansion, and those interested in multi-cloud solution architecture.

Movement without Disruption - or - How to Move Your Multi-million Pound HPC Cluster without Your Users (Hardly) Knowing

Cliff Addison, Wil Mayers, Cristin Merritt and Manhui Wang

UNIVERSITY OF LIVERPOOL

Overview

- Liverpool in 2018 demonstrated strategic benefits of cloud for research.
- Deployment was ad hoc, but worked.
- We want to embed cloud for research and graduate teaching.

UNIVERSITY OF LIVERPOOL

Successful cloud scenarios

- Cloud bursting – more cycles needed for a short period
  - typically for papers or presentations
- High throughput workflows
  - Current Windows Condor pool limited to circa 8 hr jobs
- Scoping studies
  - I think I need X cores and Y GB of memory for my research
- GPU nodes for Deep Learning
- **Avoiding** large data transfers in this first instance

UNIVERSITY OF LIVERPOOL

State of play end of 2018

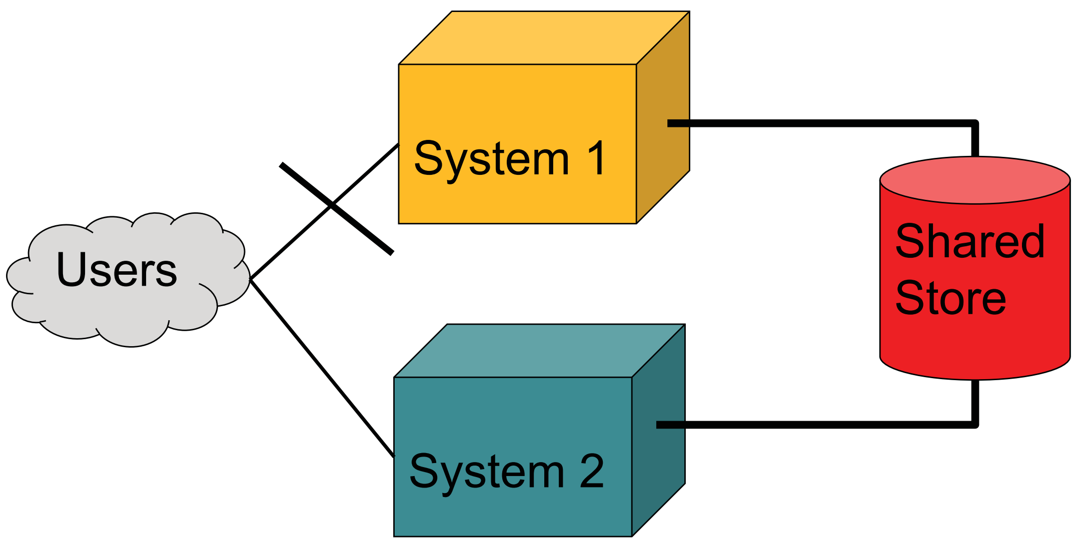
- Working on AWS an eye-opening experience.
  - Some learning curve with the EC2 and with S3 storage
- Started with Alces Flight clusters and spinning specialised instances on EC2 (e.g. Condor in the cloud).
  - Just creating instances with keys for particular groups is great for small numbers of groups, but it does not scale.
- Can get major additional benefits to an on premise HPC.
- Planned use cases for 2019:
  - Seamless cloud access from local cluster
  - Replication of some cluster functionality in the cloud

UNIVERSITY OF LIVERPOOL

Moving forward on a narrow front - HPC

- Have some basics in both AWS and Azure tenancies
  - Have Active Directory authentication for Azure
  - Direct Connect (faster network connections) still coming
- Liverpool HPC has:
  - Defined set of users (circa 70 active every month)
  - Stable software offering (slowly growing)
  - Alces Flight provide system and software framework
- Liverpool HPC needs:
  - Better resiliency – no failover component
  - More flexible environment for new users
  - Better development / experimentation support

Classical Active-Passive Failover



Issues with HPC resiliency

- Hard to support two on-campus HPC systems with an active-active failover mode (active-passive is silly)
  - HPC systems often sited in a single data centre
  - HPC systems bought at different times, maybe from different vendors
- HPC usage composed of many jobs running for hours, possibly days. [Not transactional]
- HPC storage geared towards performance and supporting a large number of simultaneous accesses
  - Hard (impractical?) to mirror all storage
- Not possible to migrate running or queued jobs.
- Cannot failover for brief outages.

What is mission critical for resilient HPC?

- Basic login and compute node environments.
- User authentication and authorisation.
- Mechanisms to load application environments and to submit jobs.
- Non-volatile user storage?
- Some compute nodes.
  - Replication of important node families, e.g. some GPUs
  - Interconnect for capability jobs (e.g. InfiniBand)?
- Budget for all of this??
  - Likely need to keep under control!

Scenarios where HPC resiliency relevant

- Power cuts to part / all of a data centre.
  - Power blips need to treated by other means
- Cooling infrastructure failure.
- Storage issues.
- Planned system maintenance.
- Relocation or “swapping” HPC systems
- Need mechanisms that can kick-in automatically.

HPC resiliency in the cloud

- Want to have an on-demand clone available
- Compute can be brought on-line relatively quickly.
- Front-end / login node and storage need to be there through the life-time on the cluster.
- Compute node costs can be controlled via autoscaling options and by exploiting the spot-market (on AWS) – how many nodes are needed?
- Pricing of cloud compute for resiliency is an issue.
  - Most cloud platforms want a year of always on use before offering major discounts over their on-demand price.
- How deal with storage??

HPC cloud resiliency – storage issues

- Three types of data storage to consider
  - System – node images and applications
    - Persistent, relatively stable, modestly sized
    - Probably current to within a week is fine
  - User home directories
    - Many systems keep to a small number of TBytes for local backup
    - Daily incremental back-up to the cloud with occasional full back-up should be possible.
  - User volatile / work areas
    - These can be huge.
    - If shutdown is planned, can get relevant users to pre-stage important data; typically during the local rundown before shutdown..
    - Cloud as a primary and permanent site for volatile data?
      - Will be slow and might be very expensive…

My understanding of AWS storage

- There are the traditional 3 storage layers:
  - Elastic Block Storage – fast access for active storage tied to hardware instances. Always need some of this on cluster. [100 GB costs about $8.10 per month]
  - Standard S3 Object Storage – slower but accessible from anywhere in the AWS cloud (and elsewhere with S3 supported logical devices) [100 GB costs about $2.30 per month]
  - S3 Intelligent Tiering, S3 Standard Infrequent access – slowly changing; not often accessed [100 GB/month $2.40, $1.31 resp.]
- Also there are the archival options, not for HPC(?)
  - S3 Glacier and S3 Glacier Deep Archive – 6 month no-change?
- Other cloud vendors have similar arrangements.

UNIVERSITY OF LIVERPOOL

2019 Motivation

- New data centre with better cooling and generator-backed power for all systems finally finished.
- Needed to move Dell / Alces system to new home.
- Old Bull (SandyBridge) cluster available during this time, but lose 4000 cores for circa 10 days.
- Idea – augment SandyBridge cluster with some cloud-based Cascade Lake (AVX-512 support) and AMD nodes – great general purpose + GPU

UNIVERSITY OF LIVERPOOL

September 2019 – 6 racks air-cooled

Racks generally look like this – cabling and storage challenge

October - moved to these water cooled racks

Trial run – June 2019

- Data centre 24 hr outage to swap power supplies
- Used outage to test cloud cluster.
- Huge advantage with Alces Flight.
  - Environment largely cloud ready
- Cloud prices vary with provider and time needed etc.
  - AWS better this test period.
  - Sacrificed faster interconnect for more nodes
- Plan – basic system login node, storage, small test node available several days before and after outage.

UNIVERSITY OF LIVERPOOL

System configuration

- Login node – Skylake 24 cores, 350 GB memory
- 10 TB shared storage for all nodes
- 2 x 2C/16GB small compute nodes for testing
- 1 x Single Nvidia V100 GPU compute node
- 20 x 36C/128GB Skylake compute nodes
  - Only available just before poweroff and for 3 days after
- 100 GB of data / day down load from cluster
- Whilst local system up, easy to copy files.
- Used existing usernames with ssh keys for access.
- Home filestore **NOT** copied over.

UNIVERSITY OF LIVERPOOL

Important considerations

- Our tailored Alces Flight Gridware preinstalled, we needed to copy over local application files.
  - Local module files completely replicated on cloud system
- Emphasis was on SMP parallel or coarse grain parallel plus Deep Learning on GPU.
- Nodes were hyperthreaded - users told to ask for exclusive access to avoid overloading.
- ssh key access – users told where to grab this from and the name of the cluster to ssh to.
- Cluster was only accessible from on-campus or via VPN to local system and then to cloud.

UNIVERSITY OF LIVERPOOL

Lessons learned

- Once access obtained, people had no problem editing slurm scripts to run jobs.
- ssh keys, off-campus access slight niggles.
- Fewer people than expected used the system (only over a weekend).
  - Next time – check on how many likely users there will be.
  - Had more compute nodes than necessary; gpu node was used
- Cost of main compute nodes ~ 75% of overall cost
  - GPU node ~ 20%
- Similar look and feel to local system big help.
- Test nodes not helpful because lacked AVX-512

UNIVERSITY OF LIVERPOOL

The major October outage

- Cloud cluster front end available from Friday 18/10
  - Easy connection / copy from on premise system
- Full cloud from 21st – powered off local compute
- Local storage / front-end power off 23rd.
- Local service restored on 30th October.
- Dropped cloud compute as jobs finished.
- Cloud front-end and storage kept available for several days so files could be transferred back.

Cloud cluster configuration - hardware

- Wanted 100 gbps InfiniBand
  - Got 100 gbps Ethernet
  - Performance was a bit erratic – need to checkout why
- Started with 16 Cascade Lake (36 cores) with 4 AMD nodes for codes without AVX-512 builds plus v100 GPU.
  - AMD nodes not being used so went with 22 Cascade Lake nodes and just one AMD node.
- Could power off/on nodes to match demand
- 10 TB shared storage across cloud cluster

Cloud cluster configuration - environment

- Users connected using standard username and password (serviced by Active Directory on campus)
- Main login system appeared to be on Alces network
- Node and login images as per local system
- User home storage copied over in advance
- Module files largely worked as normal
- Tweaks to slurm scripts needed
- Had preliminary period for file-upload
- Used appliance on campus to channel AD requests

UNIVERSITY OF LIVERPOOL

Use over the period

- Cascade Lake nodes very heavily used.
  - AMD node not used much at all
- GPU node constantly used after first couple of days
- Resources used
  - 278 user sessions to login node (19 individuals)
  - 275 GB new data generated
  - 560 slurmjobs processed (including 23 x GPU jobs)
  - 411GB data in / 275GB data out
- Mixture of SMP and small MPI jobs
- Cloud cost circa £20,000 (plus Alces logistical cost)
  - Roughly £2000 per day

UNIVERSITY OF LIVERPOOL

Plans after data centre move

- Local appliance functionality can be expanded.
  - Backup copy of node images, user data and orchestrate failover
- Integrate cloud cluster with university network
  - VPC so cluster appears as on the University network
- Bring up cloud cluster alongside Barkla to test “easy access” and cloud bursting potential.
- Experiment with spot market on AWS
  - Massive savings but small number of nodes
- Get firm University budget to sustain resiliency – local appliance plus occasional compute
  - Compute costs for full cluster mount very quickly!!

Summary – general issues

- Need cloud cluster to have a similar look and feel to the local cluster.
- Integrate the cloud cluster into your local environment
  - Local appliance helps a lot
  - Active Directory / VPC so appears on campus network
- What storage is put where?
  - Local storage that is pushed to the cloud avoids lock-in – flexibility is good!
- Compute and login nodes created on-demand
  - How many compute nodes makes sense?? Interconnect??
  - Spot-market for some / all nodes
  - Ability to power on / off nodes is a must

UNIVERSITY OF LIVERPOOL

Phil Kershaw

JASMIN

JASMIN and the Evolution of Cloud-Hosted Data Analytics Platforms for the Environmental Sciences

Philip leads the development of software and services for the multi-petabyte CEDA data archive of earth observation, atmospheric science and climate data and for JASMIN, an innovative and globally unique data intensive analysis and computational infrastructure. He has played a key role in the technical development of JASMIN since its inception in 2012, instigating the development of its cloud computing system. He leads the UKRI Working Group on Cloud Computing, which supports the interests of the UK research community in the application of cloud computing technology.

Abstract

JASMIN is a large-scale computing platform for data-intensive science dedicated to the needs of the UK environmental sciences community and its European partners. Funded by the Natural Environment Research Council, it is hosted at STFC Rutherford Appleton Laboratory in Oxfordshire where it is operated by the Centre for Environmental Data Analysis in partnership with the Scientific Computing Department.

In this presentation I will explore the role of cloud computing for JASMIN and the application of cloud for the wider climate and earth observation communities. Now in its eighth year of operation, JASMIN was originally conceived around the paradigm of bringing the compute to the data in response to the challenge of analysing and managing increasing data volumes facing many in the user community. JASMIN provides a data commons: a shared community resource which acts as host to a curated data archive, shared user-managed workspaces together with colocated computing capacity for data analysis.

Cloud provides a natural fit to this model with its characteristics of remote network access to pooling of computing resources. Together with other key technologies – high performance networking, shared file system and batch computing environment – cloud has been applied from the outset to deliver JASMIN's goals.

JASMIN and the evolution of cloud-hosted data analytics platforms for the environmental sciences

Philip Kershaw
CEDA Technical Manager
Centre for Environmental Data Analysis, RAL Space, STFC

Bryan Lawrence, Jonathan Churchill, Matt Pritchard, Victoria Bennett

Overview

- What is JASMIN?
- Evolution of cloud and data analysis platforms
- JASMIN in the wider context of the environmental science community with focus on
  - Climate
  - Earth Observation
- Big Data, Federated Systems Cloud Computing

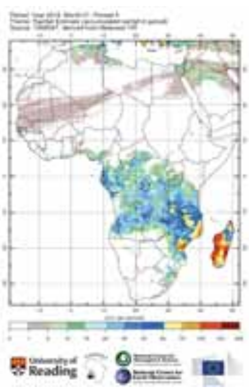
What is JASMIN?

- Large-scale computing platform for data-intensive science for NERC environmental science community
- Operated by STFC on behalf of NERC
  - Architecture: **CEDA** & **Scientific Computing**
  - Physical infrastructure: **Scientific Computing**
  - User services: **CEDA**

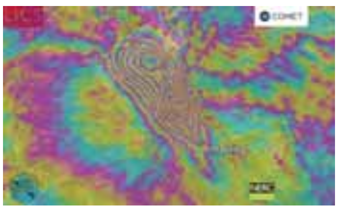
Science and Technology Facilities Council
Natural Environment Research Council

What is JASMIN used for?

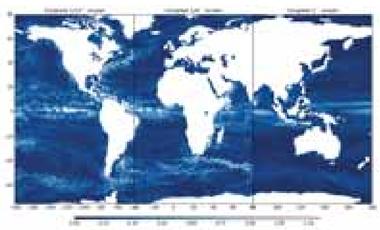
Enabling climate services – allows insurance for >1million African farmers



Earthquake monitoring



Improving resolution of climate models



Analysing biases in biodiversity data

Access to over 13PB of environmental data held in the CEDA Archive

… and lots more!!

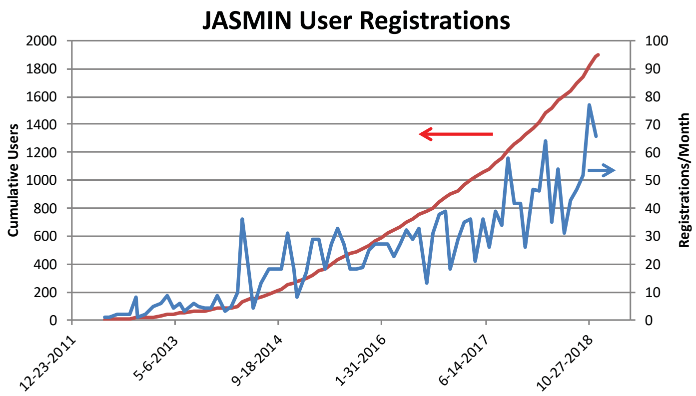
Science and Technology Facilities Council
Natural Environment Research Council

International Impact

User Growth

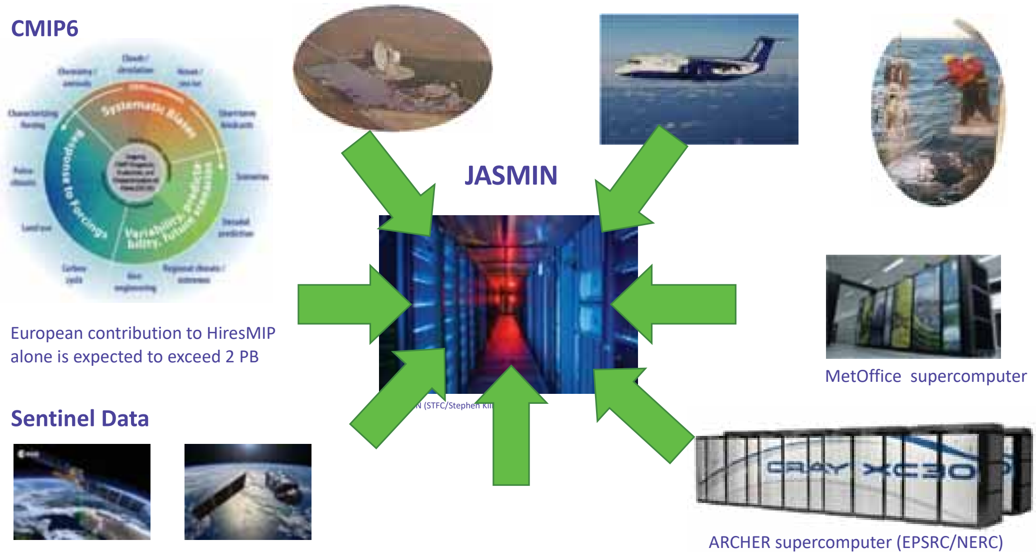
>2,000 (Sept 2019)

JASMIN user workshop Sept 2019



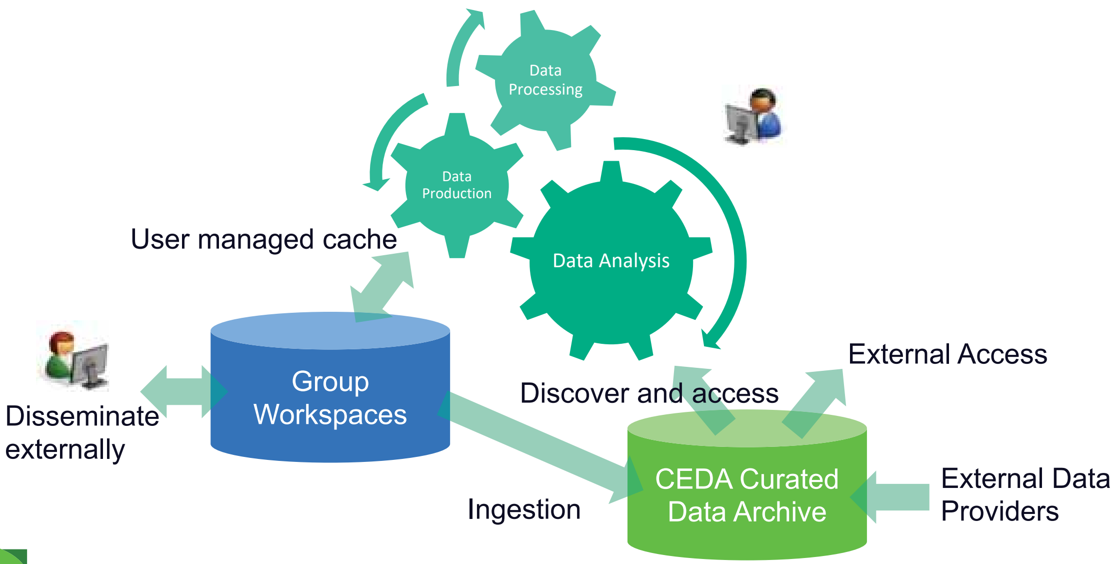
Note: Excludes 17,000+ CEDA Download users

JASMIN: the missing piece



Science and Technology Facilities Council
Natural Environment Research Council

JASMIN as a Data Commons



Science and Technology Facilities Council
Natural Environment Research Council

Cloud and JASMIN

- Cloud provides a natural fit to JASMIN’s model
- Characteristics* of -
  - remote network access to
  - pooling of computing resources

*The NIST Definition of Cloud Computing, Special Publication 800-145

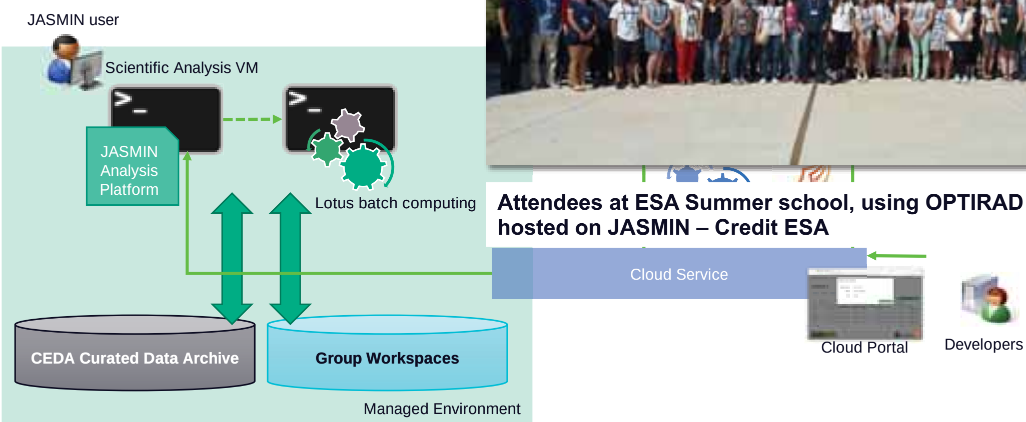
Science and Technology Facilities Council
Natural Environment Research Council

JASMIN Cloud – facts and figures

- Started out with VMware vCloud Director
- Migrated to VIO (VMware Integrated OpenStack)
  - 90 tenancies
  - 229 VMs spread across all tenancies
  - Current utilisation: 824 vCPUs, 1.6TB of RAM, 55TB storage
- Augmenting VIO with new Mirantis OpenStack from new year
  - vCPUs: 2400, RAM: 18.4TB
  - SSD storage for some hypervisors
  - SR-IOV for higher performance applications

Science and Technology Facilities Council
Natural Environment Research Council

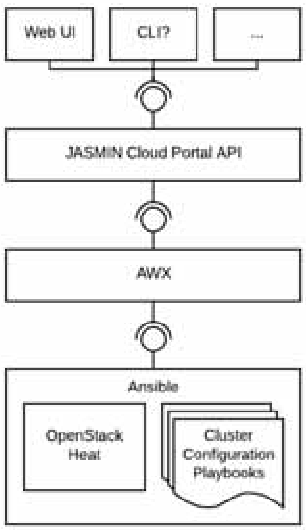
JASMIN Cloud - Evolution



Attendees at ESA Summer school, using OPTIRAD environment hosted on JASMIN – Credit ESA

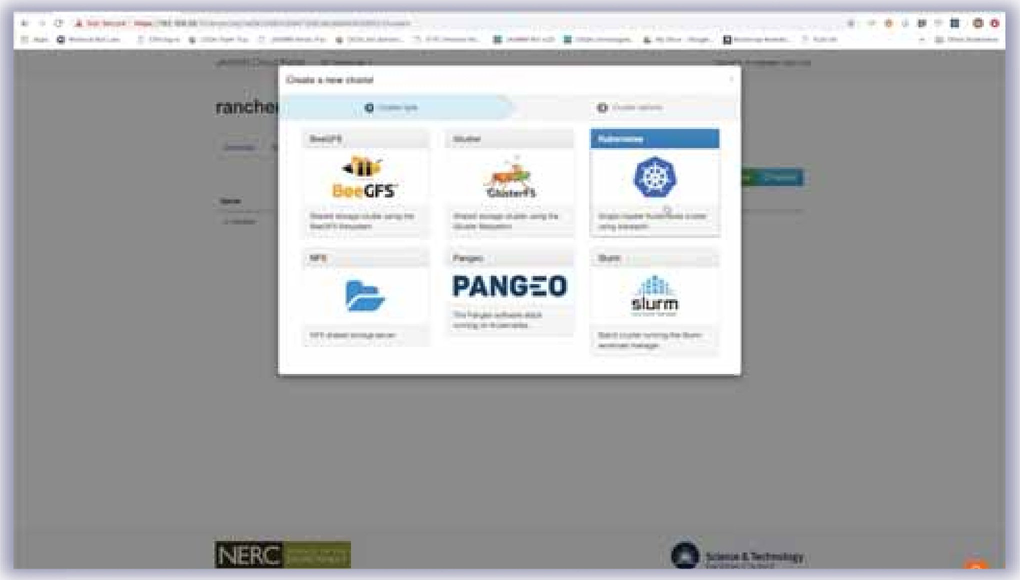
JASMIN Cluster-as-a-Service

- 3 month project to create ready made appliances for users to deploy from JASMIN's cloud
  - Pangeo
  - Kubernetes
  - Storage (BeeGFS, NFS and Gluster)
  - Batch cluster (Slurm)
- Driven with Ansible Playbooks and OpenStack Heat templates
- Playbooks managed by AWX (Ansible Tower)
- Supports updates and patching to existing clusters

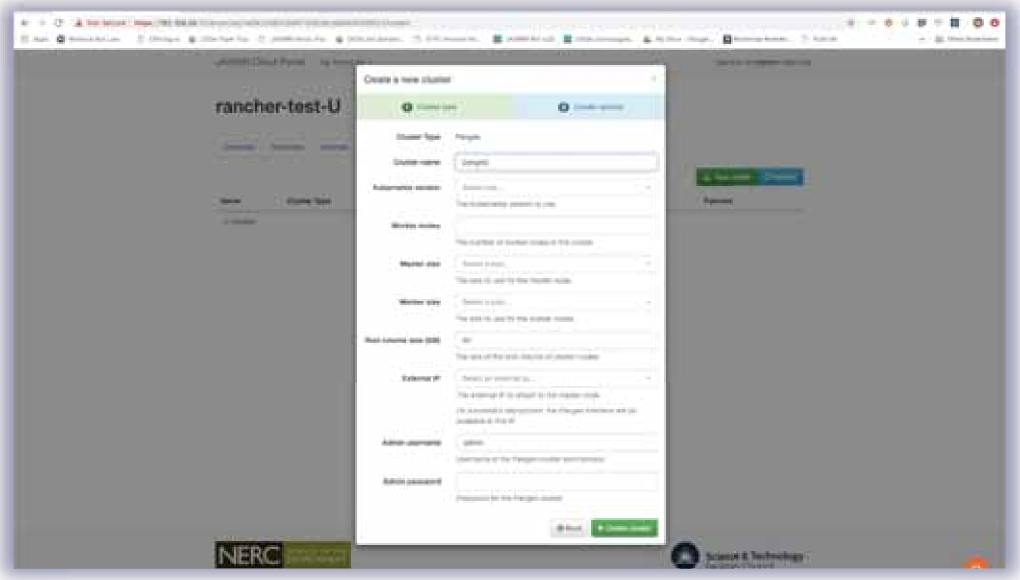


JASMIN Cluster-as-a-Service

JASMIN Cluster-as-a-Service



JASMIN Cluster-as-a-Service

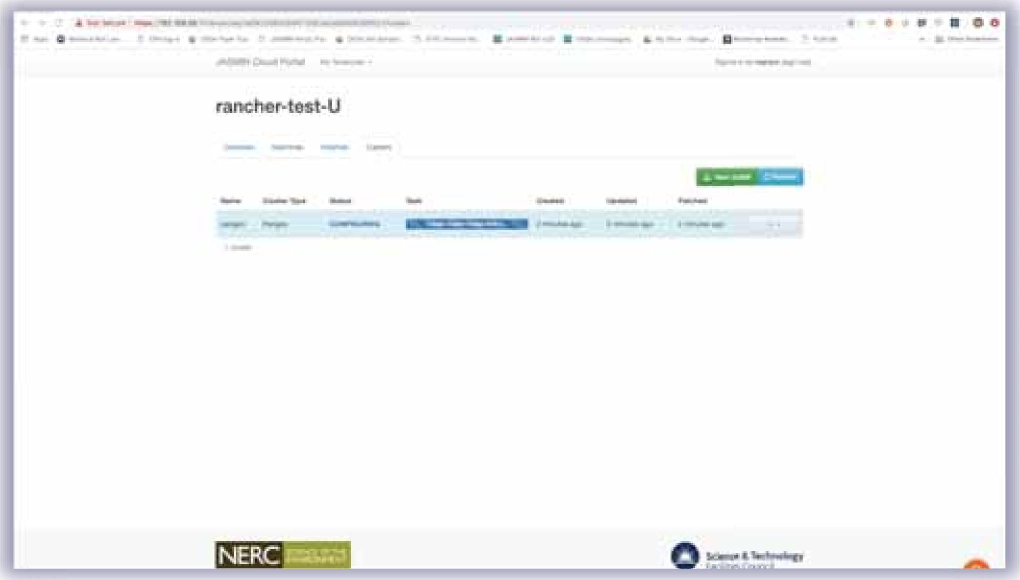


Science and Technology Facilities Council
Natural Environment Research Council

JASMIN Cluster-as-a-Service

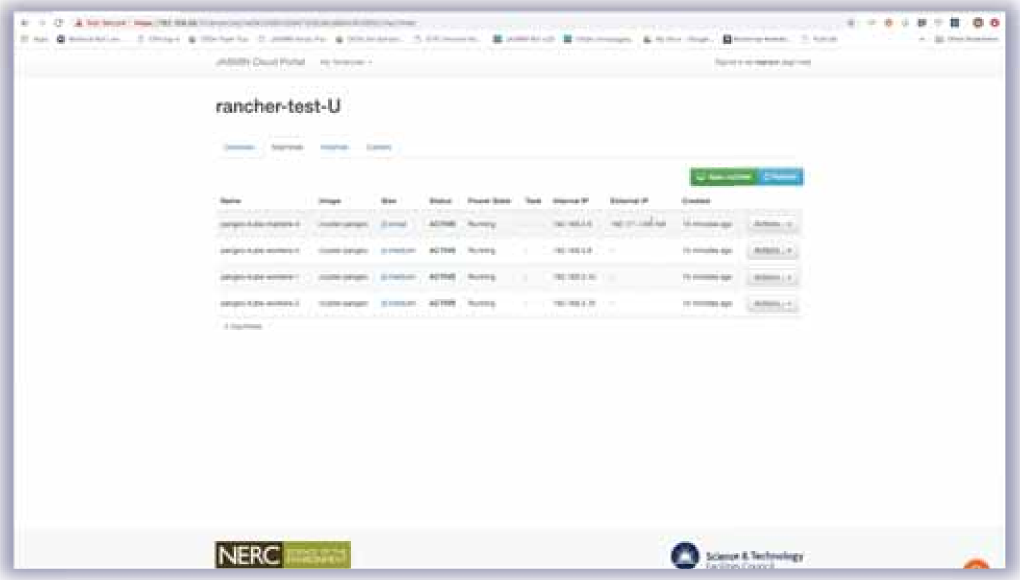
Science and Technology Facilities Council
Natural Environment Research Council

JASMIN Cluster-as-a-Service

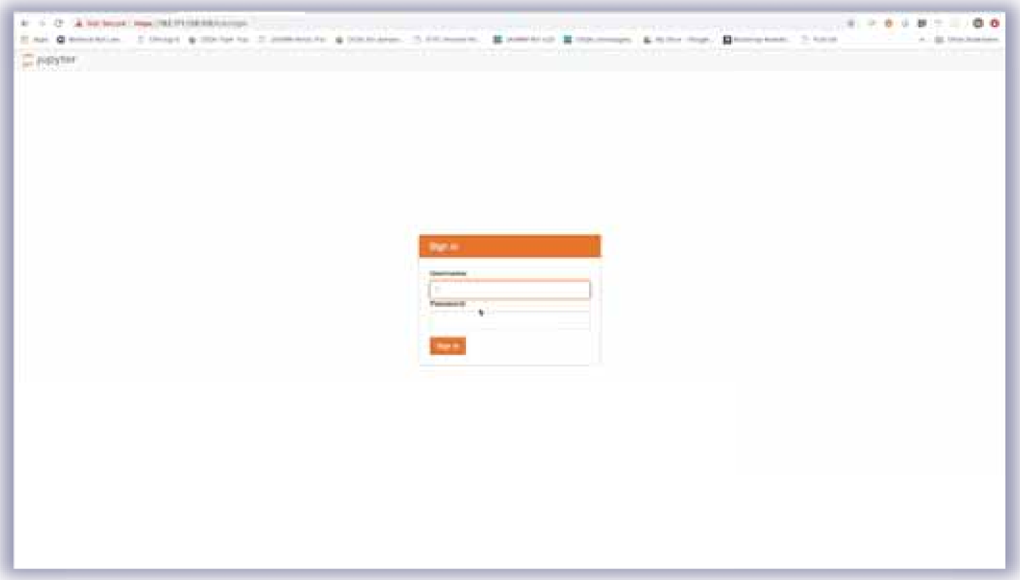


JASMIN Cluster-as-a-Service

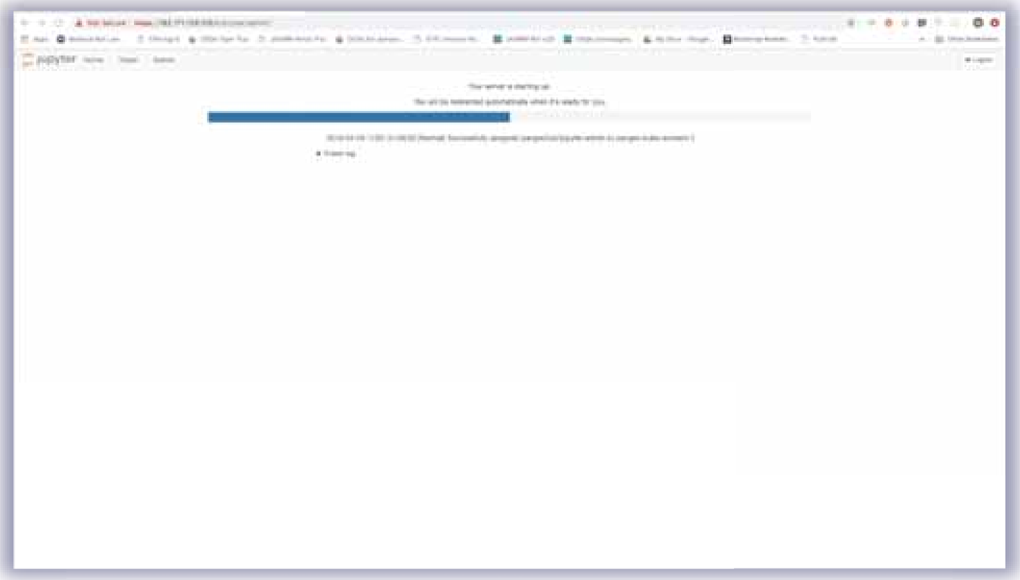
JASMIN Cluster-as-a-Service



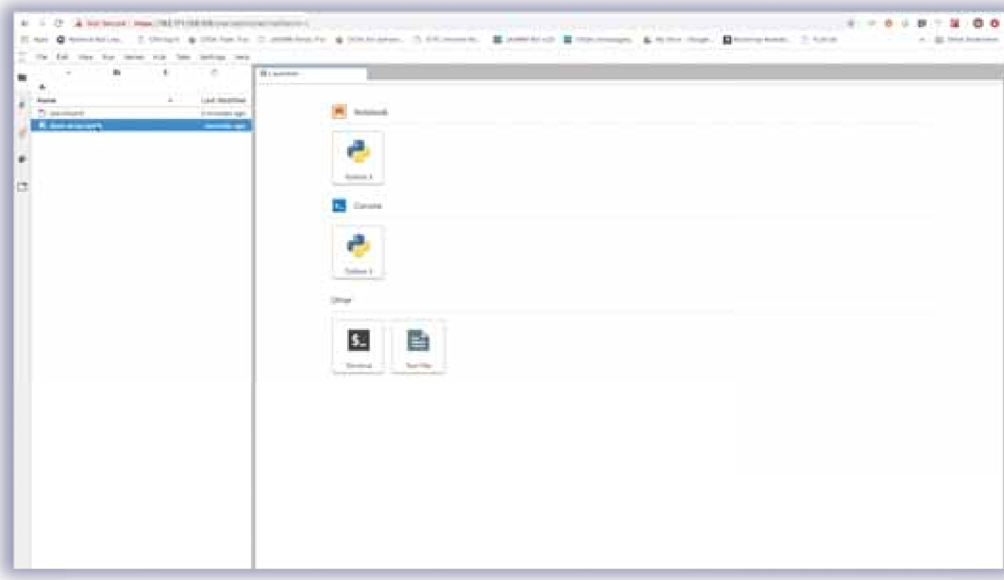
JASMIN Cluster-as-a-Service



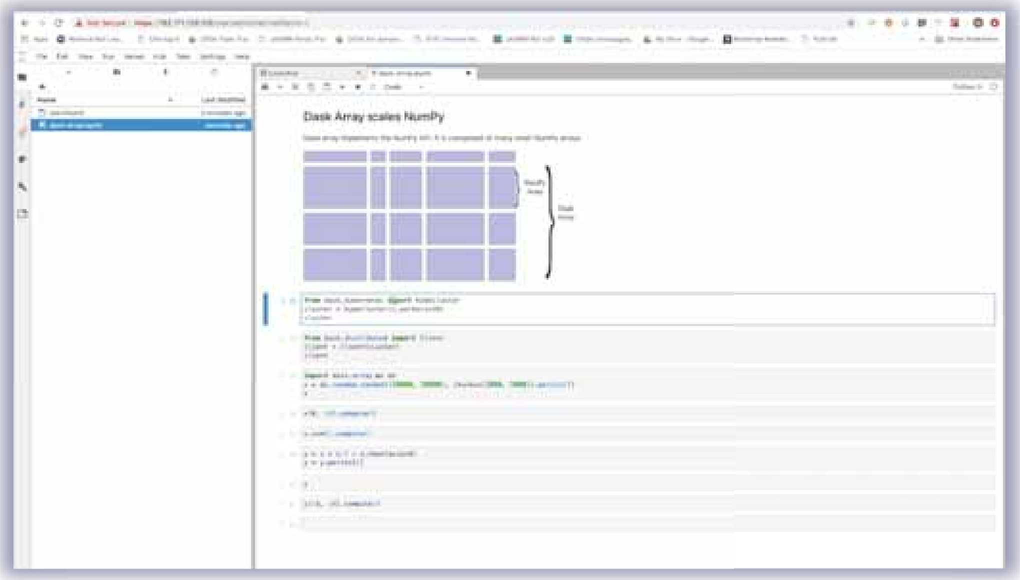
JASMIN Cluster-as-a-Service



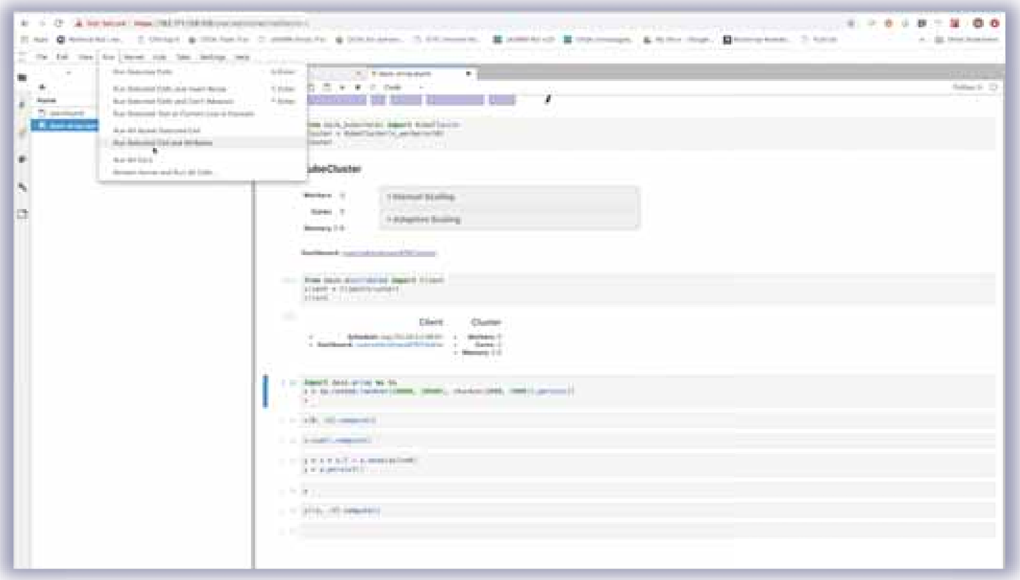
JASMIN Cluster-as-a-Service



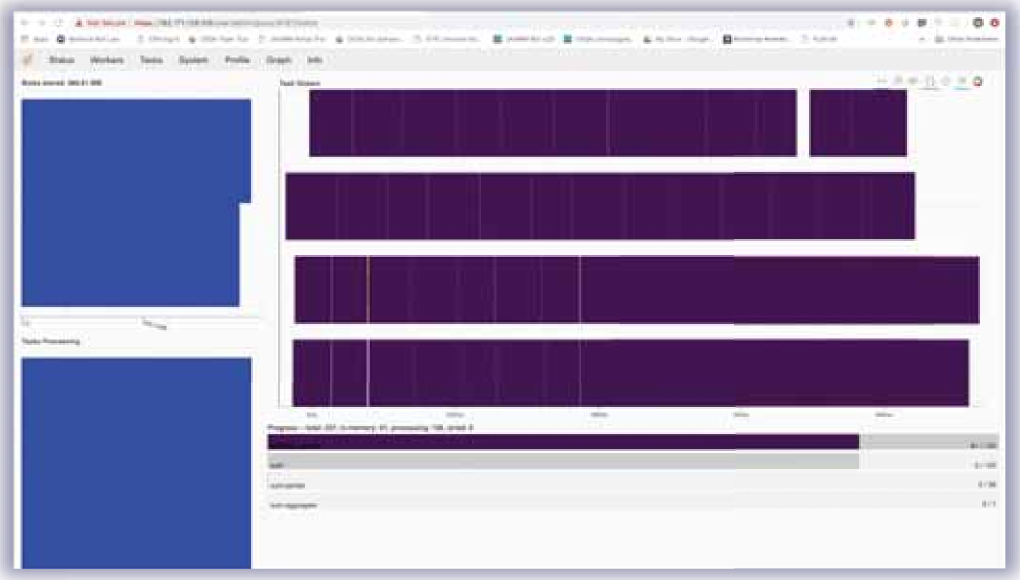
JASMIN Cluster-as-a-Service



JASMIN Cluster-as-a-Service



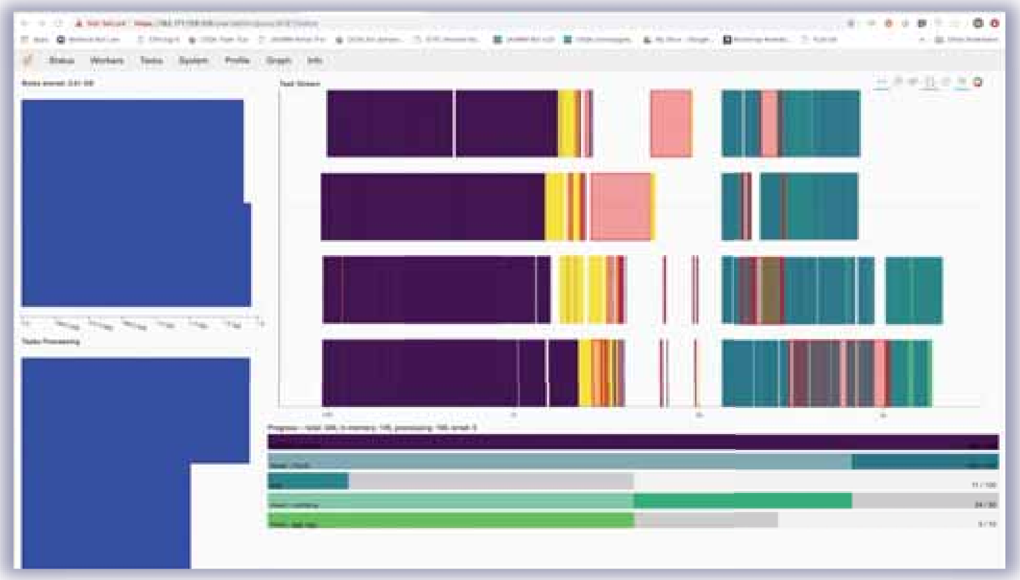
JASMIN Cluster-as-a-Service



Science and Technology Facilities Council

Natural Environment Research Council

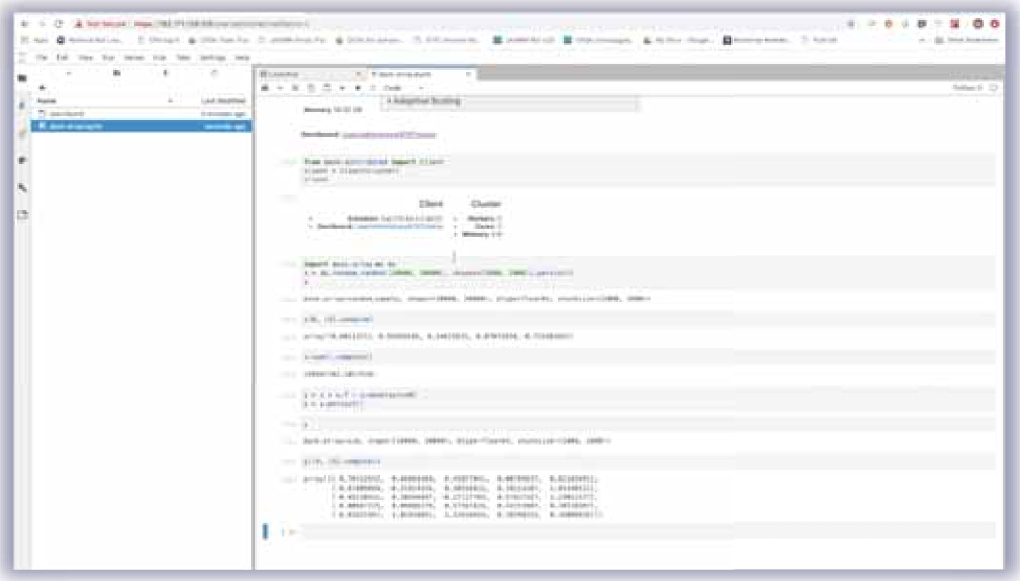
JASMIN Cluster-as-a-Service



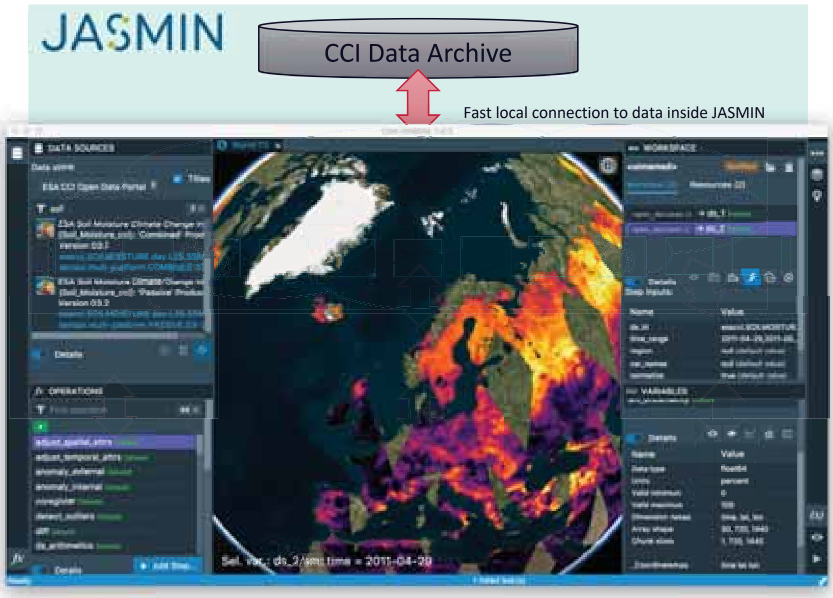
Science and Technology Facilities Council

Natural Environment Research Council

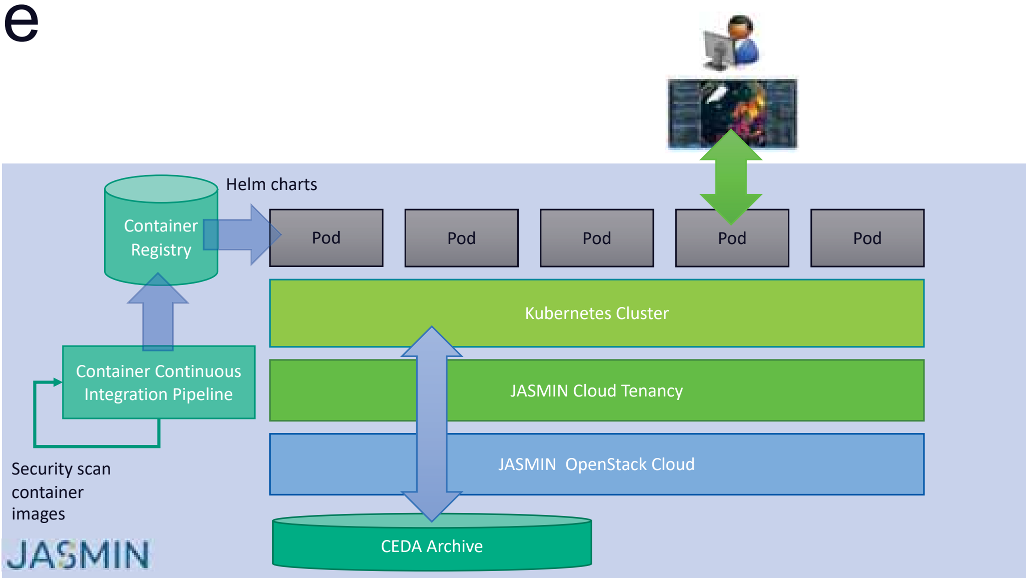
JASMIN Cluster-as-a-Service



ESA Climate Change Initiative - Toolbox



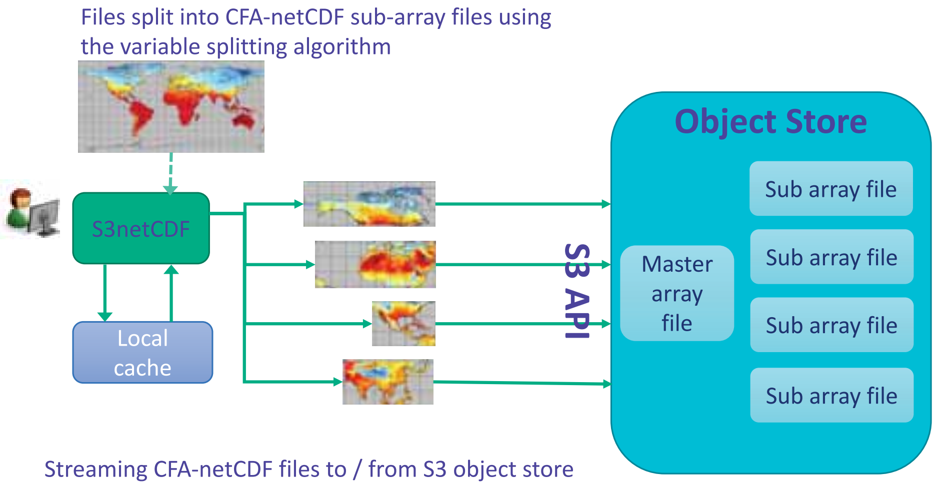
Kubernetes, Cloud and Infrastructure-as-Code



But what about storage and cloud?

- POSIX mount semantics and cloud sit together uneasily
- JASMIN has been experiencing the tension of parallel file system vs scale
- Object storage attractive option to address these challenges
- … But most scientific applications still use POSIX
- Most recent JASMIN procurement has purchased scale-out file system and object storage
- Initiatives in the community such as Pangeo have used Xarray with Zarr as an interface to store and access data efficiently with object stores

Integration of object storage with scientific data format



Big Data, cloud and the evolution of systems for data distribution and analysis

Big Data driving changes in architecture

Single data centre
- Discover and download user model
- CEDA (< 2008: pre-ESGF and pre-JASMIN)

Federated data centres
- Multiple organisations
- geographically distributed download capability
- **Earth System Grid Federation** initially for **CMIP5 Global Climate Projections** from 2008

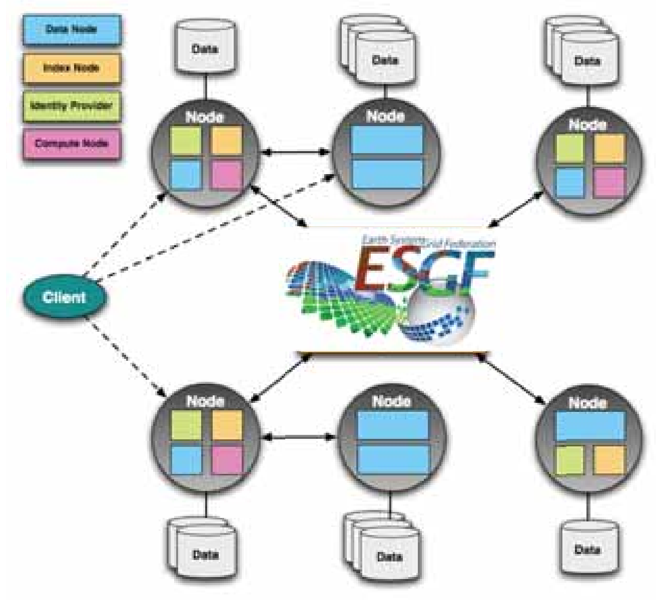
Data analysis facility
- Bring the compute to the data paradigm
- **JASMIN** (from 2012)

Public Cloud
- Content Delivery Network

Data Analysis Platforms
- Analysis ready data
- Community Resources
- ESA Thematic Exploitation Platforms

Earth System Grid Federation (ESGF)

- Globally distributed infrastructure for dissemination of earth sciences data
  - Including DoE, EU IS-ENES collaboration, NASA, NOAA, NCI Australia …
  - ~20 nodes
  - ~17000 users
- Federation inherently supports redundancy and replication capabilities
- Existing community, operational procedures and governance



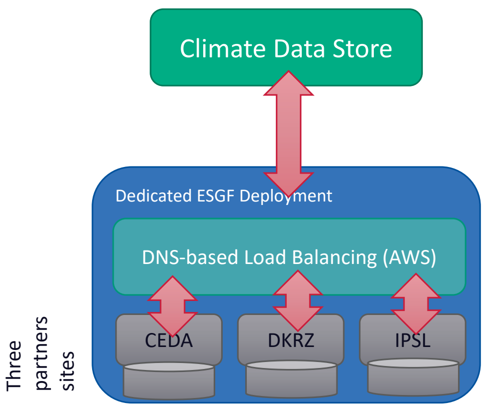
Delivering CMIP5 data for Copernicus using ESGF

- Climate Data Store (CDS) is part of the Copernicus Climate Change Service (C3S) operated by ECMWF on behalf of the EU
- CDS is a single, freely available interface to a range of climate-related observations and simulations
- To provide key indicators of climate change drivers, supporting all sectors
- Wide range of data sources from many participating organisations
  - In-situ observations, **models**, reanalyses, satellite products
- **CEDA have led a project to provide climate model data from CMIP5 to the CDS using ESGF …**

Copernicus requirements and Cloud

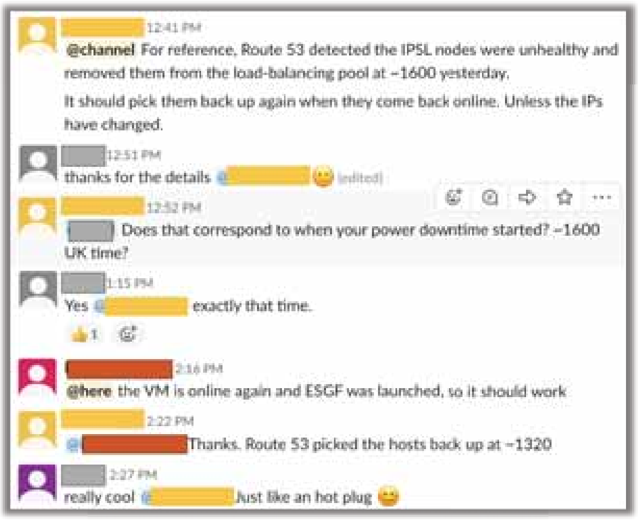
- Requirement for >= 98% uptime
- Greater uptime figure than typical for research infrastructures: ~95%
- Public cloud hosting
  - Can provide the necessary resilience
  - But storage costs are high for the volume required (100s TB)
- Solution: Load balance between partners sites running replicas of the data ➔ gives aggregate uptime meeting the requirements

Load balancing across three site replicas



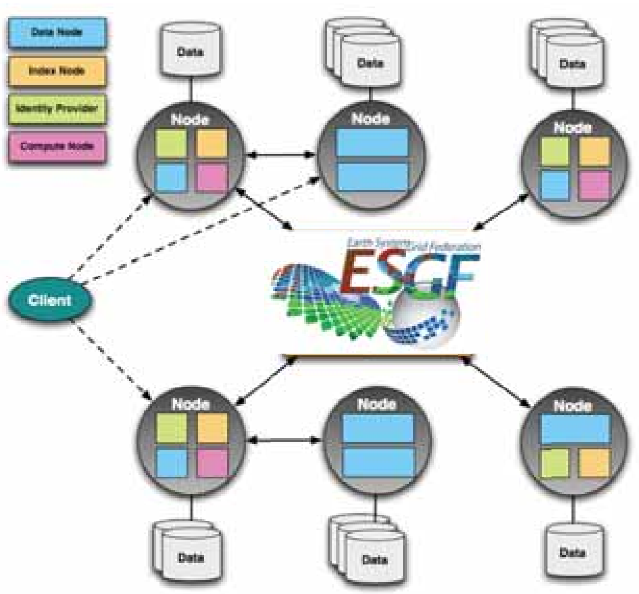
- Each site run operates with three identical copies of the data
  - Using ESGF publication and replication capabilities
- DNS-based load balancing between the three using AWS Route 53
- Health checks monitor for and remove any one site's service if inoperative
- Data search services have a smaller footprint and so could be public cloud hosted
  - Successfully piloted with Google Kubernetes Engine

Load Balancing for Resilience in Action



What does Copernicus experience with Cloud mean for ESGF?

- ESGF federated services for resilience
- This model can be rethought using public cloud
- Some services can be centralised and run on public cloud
  - Search and identity services
- Hosting of large data volumes is still not cost effective for public cloud



Conclusions and thoughts for the future

- Cloud model fits JASMIN's model but there are challenges
  - Storage interfaces and performant access
- Cloud has provided an impetus for data analytics platforms
  - Set-up and tear down model for projects on public cloud works
  - But data centres need steady-state long term hosting of large data volumes
  - Infrastructure-as-code is powerful and is being transformative

Science and Technology Facilities Council
Natural Environment Research Council

Conclusions and thoughts for the future (2)

- Where next with 'traditional' hypervisor virtualisation and containers and Kubernetes?
- JASMIN is in pre-operations for a bare-metal Kubernetes cluster
- Removes much of the overhead of complexity and performance
- Kubernetes multi-tenancy features not as developed

Science and Technology Facilities Council
Natural Environment Research Council

Recap – Evolution of systems for data distribution and analysis

Big Data driving changes in architecture

Single data centre
- Discover and download user model
- CEDA (< 2008: pre-ESGF and pre-JASMIN)

Federated data centres
- Multiple organisations
- geographically distributed download capability
- **Earth System Grid Federation** initially for **CMIP5 Global Climate Projections** from 2008

Public Cloud
- Content Delivery Network

Data analysis facility
- Bring the compute to the data paradigm
- **JASMIN** (from 2012)

Data Analysis Platforms
- Analysis ready data
- Community Resources
- ESA Thematic Exploitation Platforms

ESA Network of Platforms
- Federating analysis platforms

Thank you!

Website: www.jasmin.ac.uk

Twitter: @cedanews

Email: support@ceda.ac.uk

To get access:

help.jasmin.ac.uk/article/189-get-started-with-jasmin

JASMIN CaaS Jupyter Notebook demo:

https://www.youtube.com/watch?v=pUQp3ZCWVH4

Jens Jensen

Science and Technology Facilities Council

Cagliari Airport AI in Fog to Cloud HPC

Dr Jens Jensen is a scientist in STFC's Scientific Computing Department. His interests include managing hundreds of petabytes of data globally, architecting data security for research, and defining best practices for trustworthy identity management and scalable authorisation to enable researchers to collaboratively share and analyse this data. He manages and/or contributes to projects that range in scale from IoT to international research infrastructures. With a background in mathematics, he also likes to promote scientific software engineering and deployment, and mathematical methods for data analysis, statistics, and machine learning.

Abstract

Airports can be confusing environments even at the best of times.

Recent advances in IoT have made it possible to develop real-time improved traveller assistance tools for mobile phones, assisted by cloud-based machine learning. This assistance will offer value to all travellers, but will be particularly valuable for the elderly or disabled travelers, or others who need assistance.

The app covers the essential path through the airport onto the flight, from the least busy security queue through to the time to walk to gate, gate changes, and other obstacles that airports tend to entertain travellers with.

While waiting for boarding, travellers are given the opportunity to discover the facilities of the airport, aided by a recommender system using collaborative filtering. Whether they are looking for gifts, a meal, duty free, a book or magazine, the system knows the layout of the airport and the location of the traveller, and can, based on the traveller's preferences or on similarities to other travellers, make recommendations and offer vouchers. Users choose how much data to share: at the lowest level, only their position is revealed; at the higher level they have shared their flight information (so get updates only for that flight) and preferences.

At the same time the system provides obvious benefits to the airport operator, not just potentially increased footfall in the shops, but also a user "heat map" which can highlight congestion and other anonymised data to highlight situations that require intervention, such as emergencies.

Smart Airports

ENGINEERING
ENGINEERING SARDEGNA

Science and Technology Facilities Council

Antonio Salis, Roberto Bulla, Glauco Mancini
Jens Jensen

CIUK December 2019

Project Number 730929
Start Date 01/01/2017
Duration 36 months
Topic ICT-06-2016 Cloud Computing

Why smart?

And what is a "smart airport" anyway?

Navigating airports

Help people navigate airport “features”

- Which terminal?
- How to get to the terminal?
- How to get to gate?
- Time to clear security

Innsbruck, Österreich; Ralf Roletschek https://commons.wikimedia.org/wiki/File:12-06-05-innsbruck-by-ralfr-151.jpg

Navigating airports

Which gate is my flight supposed to depart from?

Which gate will it actually depart from?

What time is it supposed to depart?

What time will it actually depart?

Milan Nykodym https://commons.wikimedia.org/wiki/File:Saab_JAS-39_Gripen_of_the_Czech_Air_Force_taking_off_from_AFB_%C4%8C%C3%A1slav.jpg

Navigating airports

Disability

“needing extra time to board”

Travelling together

By George F. Sargent - http://ihm.nlm.nih.gov/images/A13292, Public Domain, https://commons.wikimedia.org/w/index.php?curid=7607428

By Claus Ableiter, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=2445063

Spending ~~Time~~ Money in Airports

While you wait for your flight, would you like to buy
... things you’ve forgotten?
… replacements for things confiscated in security?
... some presents?
… some souvenirs?
… something to read?
… something to eat?
… something more “medicinal”?

Missouri History Museum http://collections.mohistory.org/resource/86794

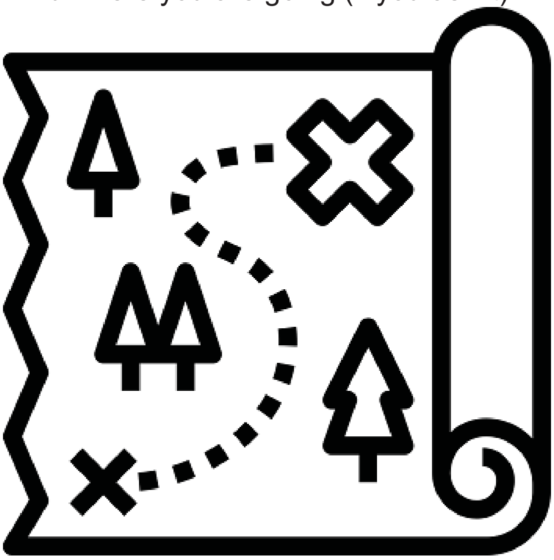
Aim of Project

Develop a phone app to support the traveller

Initially, Android only

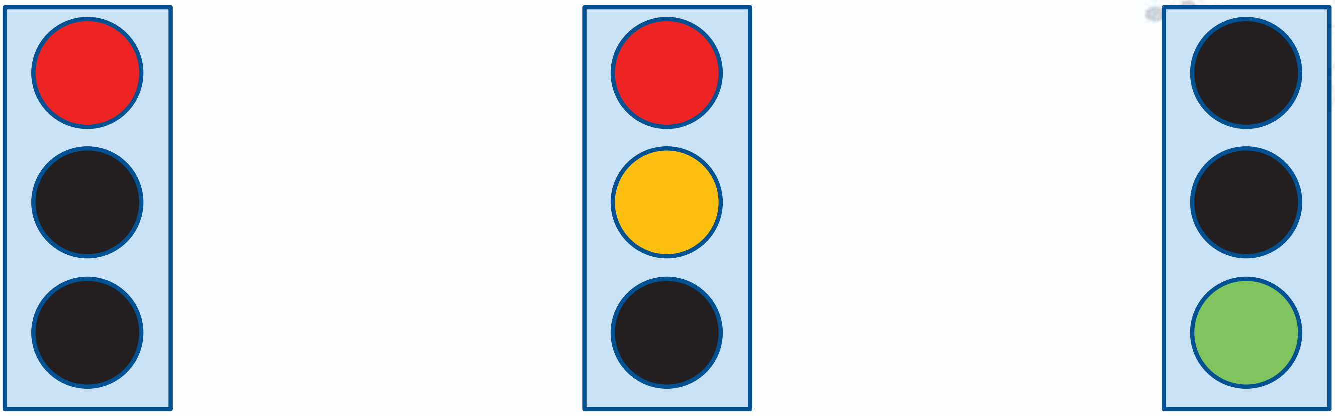
It knows where you are

And where you are going (if you ask it)



… in the airport, by checking your wifi signal strengths

Three Levels of Data Sharing



None: app infers your interests by comparing you to others (recommender system)

Some: traveller gives hints to app about their interests/need (but still get all dept. announced)

Full – traveller is still anonymous, but the app knows the flight number

Implementation

Functional requirements

What does the airport need to do?

Know its own layout:

- Gates, security, shops and other points of interest
- How to route a user from any point to any other point
  - Optionally step-free
  - Time it takes to walk this distance at given walking pace

Departures:

- Know planned and live departure schedule
- (Optionally) know duty free rules for all destinations

What does the airport need to do?

For each user:

- Register user
- Compare WiFi endpoint signal strengths and calculate position
  - Based on relative signal strengths, not TDOA (Time Difference of Arrival)
- Track user’s flight (if known)
  - Otherwise will have to give user a timely alert to all departures
- Stop messaging them when they’ve left!

What does the airport need to do?

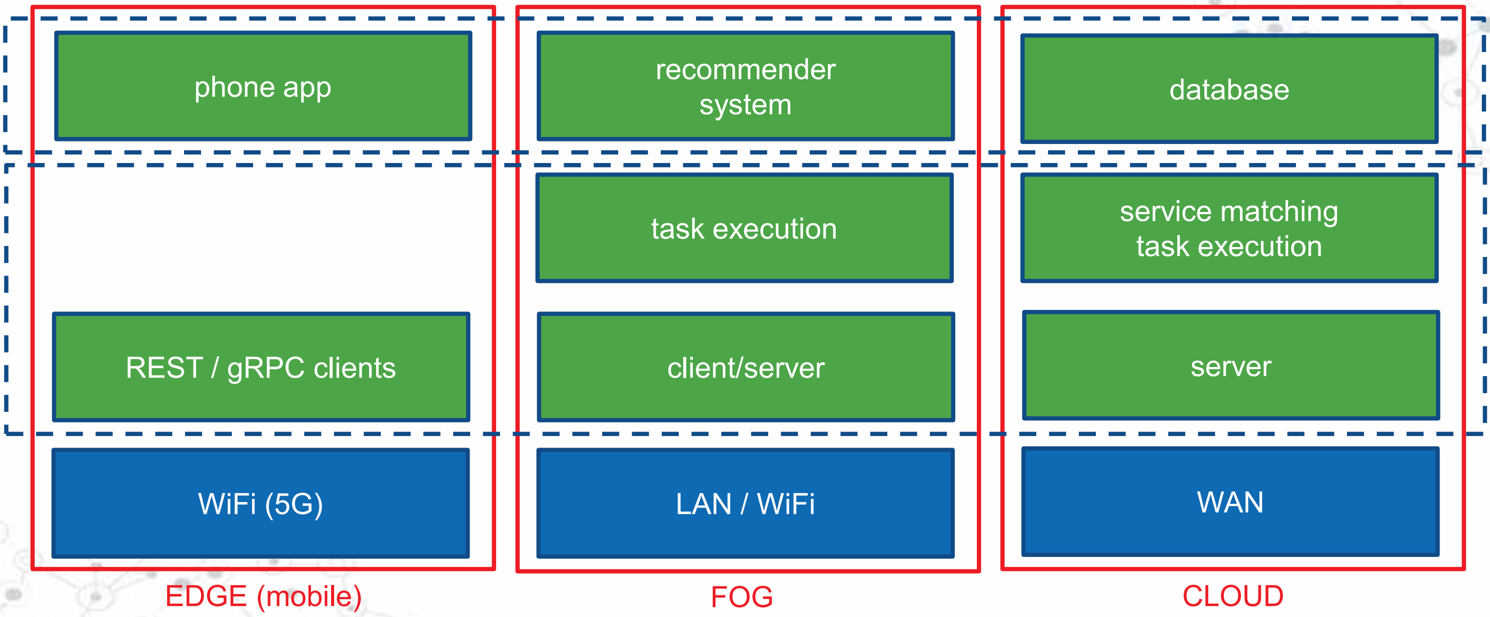
For all users:

- Make (and update) recommendations for points of interest (if they have time)
- Like “users who bought X also often buy Y or Z”

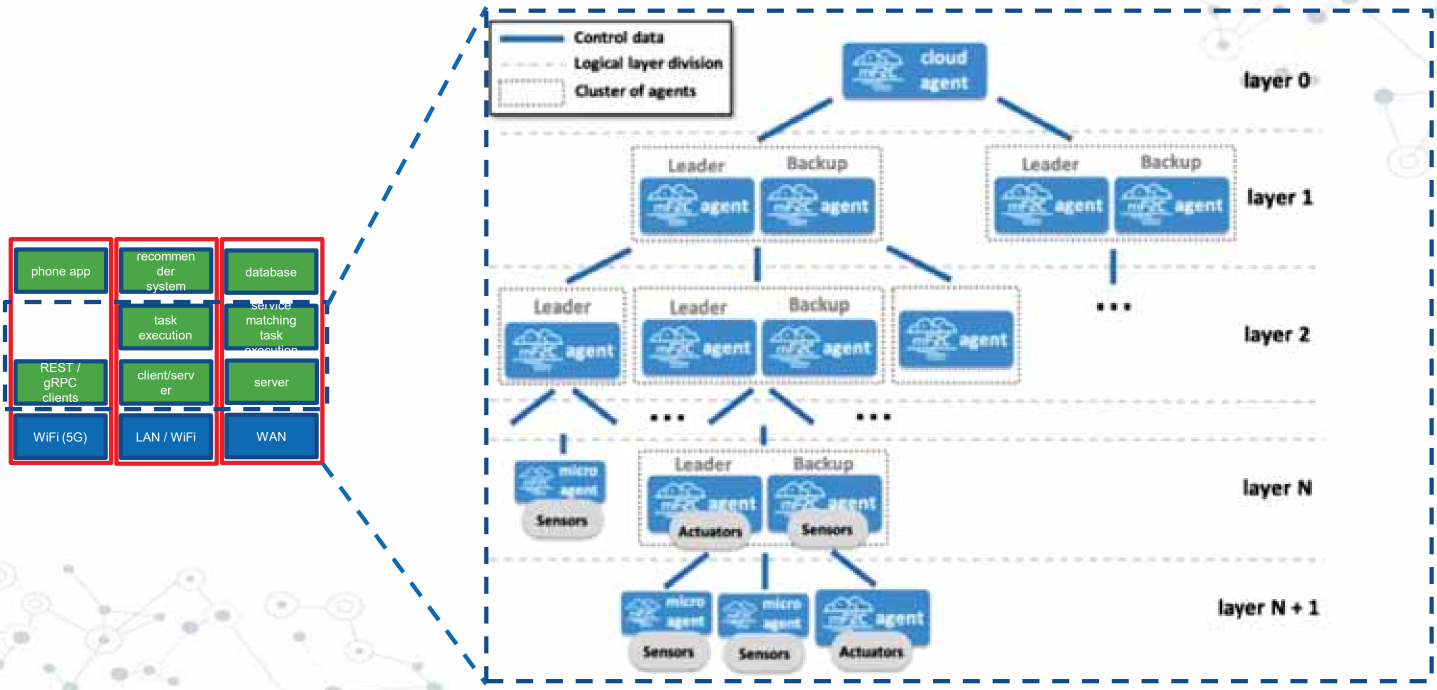
Implementation

Edge-to-Cloud platform

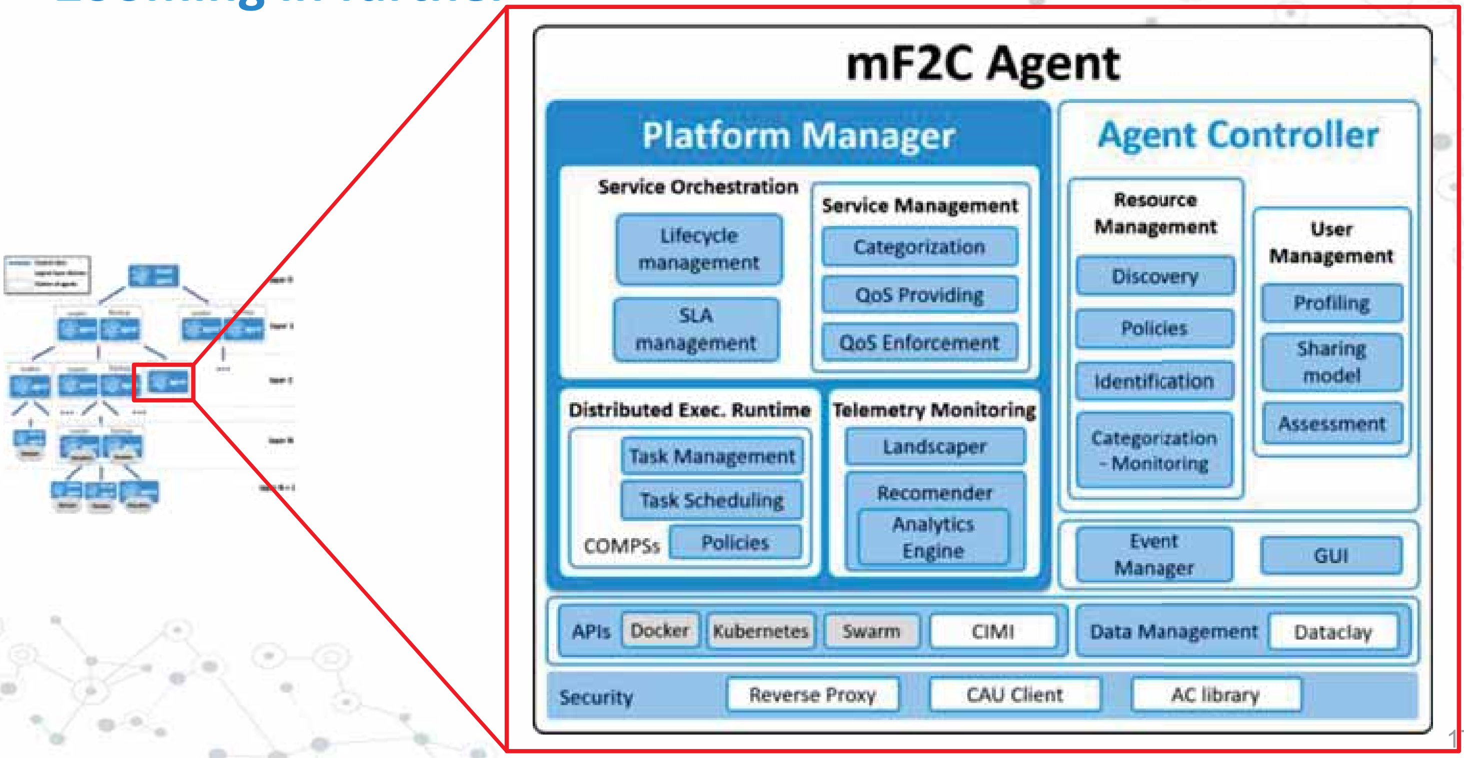
OSI (near enough) stack view



Platform Architecture

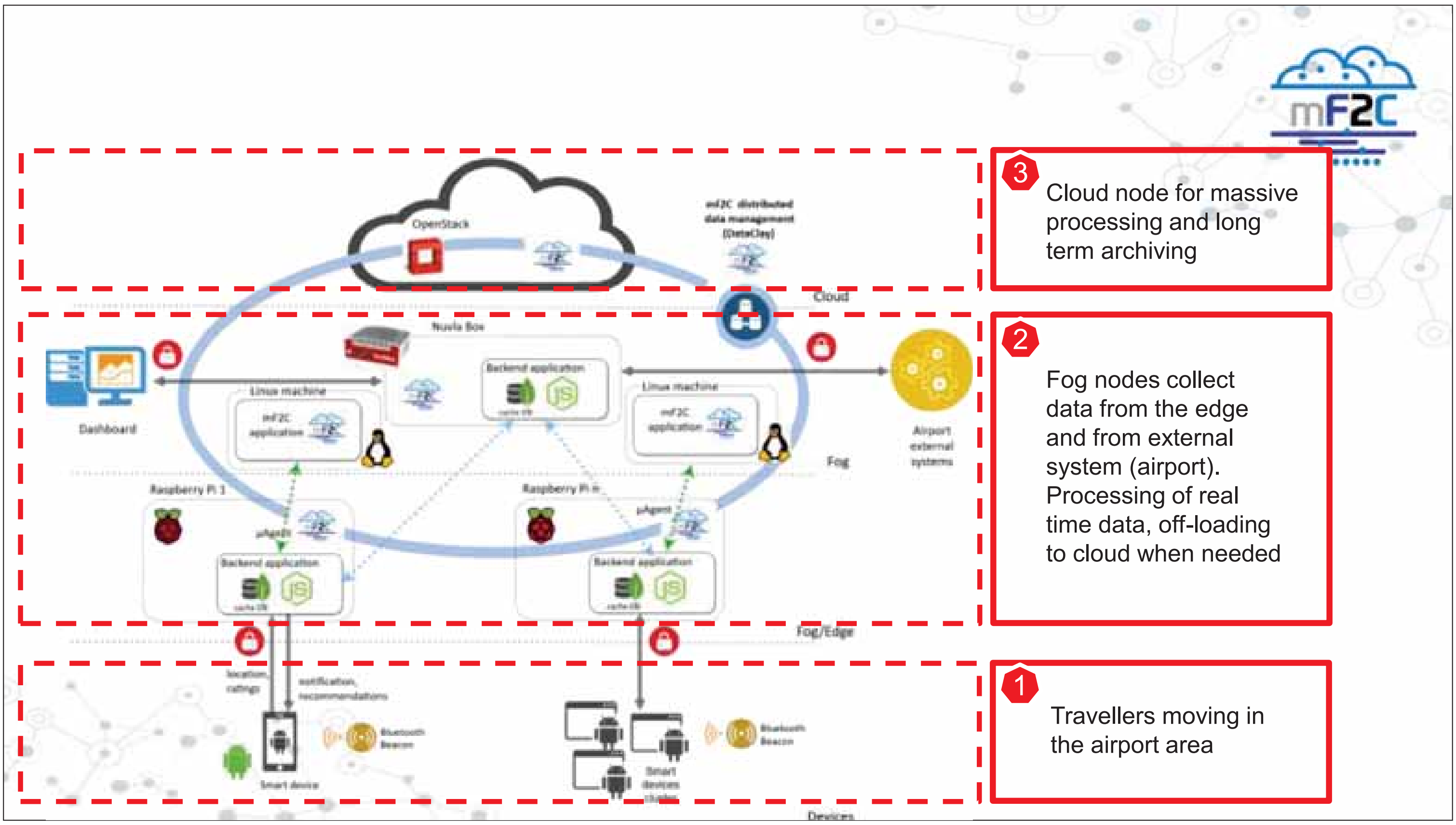


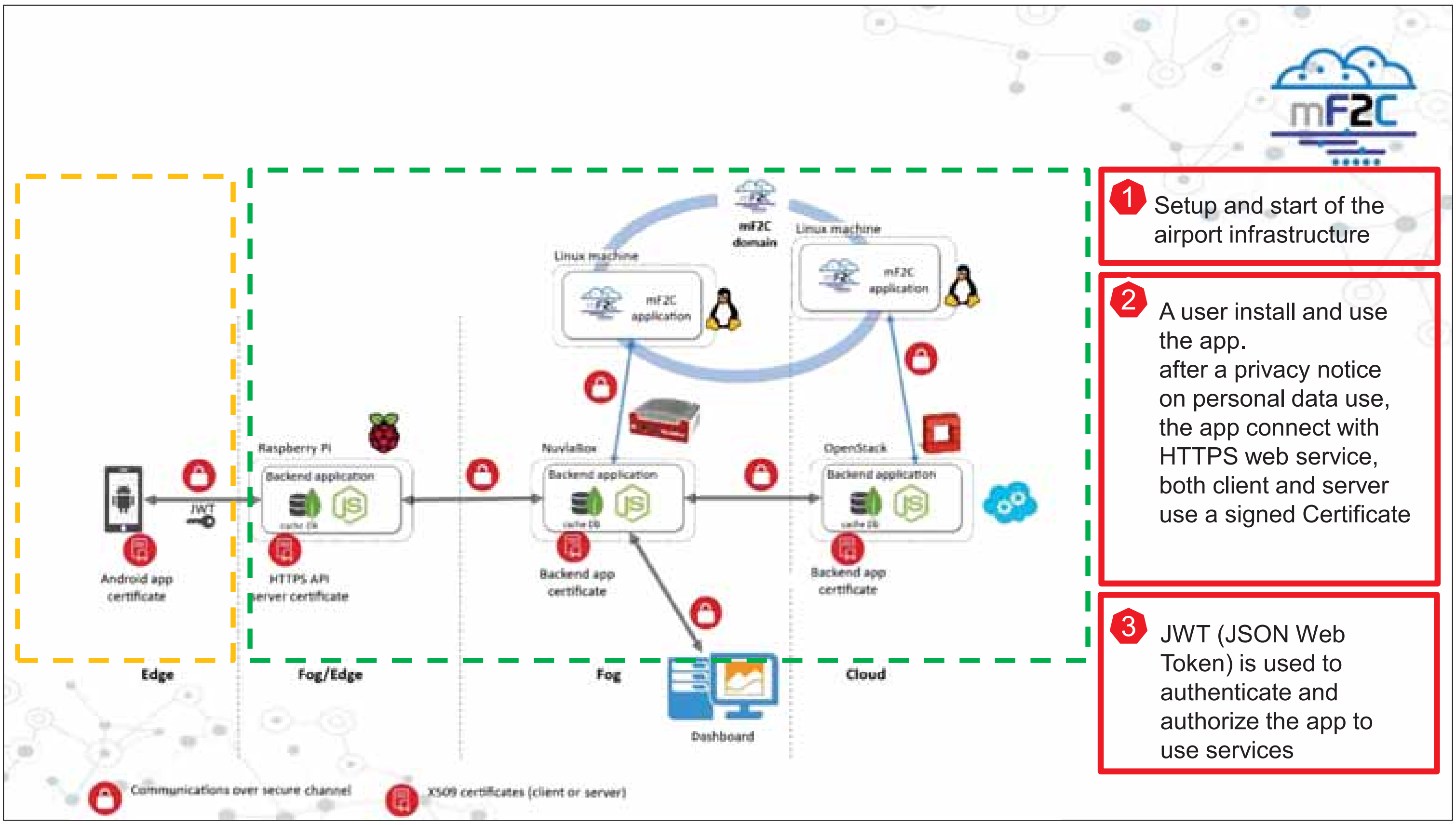
Zooming in further



Platform Features for Airport App

- Built-in security
- PKI: agents obtain credentials from a cloud-based CA, through a fog-cloud gateway
- Libraries to manage message confidentiality, integrity
- Task execution matching & monitoring
- Agents are asked to run tasks and find suitable locations
- Execution is monitored


mF2C
3
Cloud node for massive processing and long term archiving
2
Fog nodes collect data from the edge and from external system (airport). Processing of real time data, off-loading to cloud when needed
1
Travellers moving in the airport area
OpenStack
mF2C distributed data management (DataClay)
Cloud
Nuvla Box
Backend application
Linux machine
mF2C application
Dashboard
Airport external systems
Fog
Raspberry Pi 1
Raspberry Pi n
µAgent
Fog/Edge
location, ratings
notification, recommendations
Bluetooth Beacon
Smart device
Smart devices
Devices


mF2C
1
Setup and start of the airport infrastructure
2
A user install and use the app. after a privacy notice on personal data use, the app connect with HTTPS web service, both client and server use a signed Certificate
3
JWT (JSON Web Token) is used to authenticate and authorize the app to use services
mF2C domain
Linux machine
mF2C application
Raspberry Pi
Backend application
NuvlaBox
OpenStack
JWT
Android app certificate
HTTPS API server certificate
Backend app certificate
Dashboard
Edge
Fog/Edge
Fog
Cloud
Communications over secure channel
X509 certificates (client or server)

Experiences

Reusing Stuff From A Research Project? ☺😐☹?

What's ☺

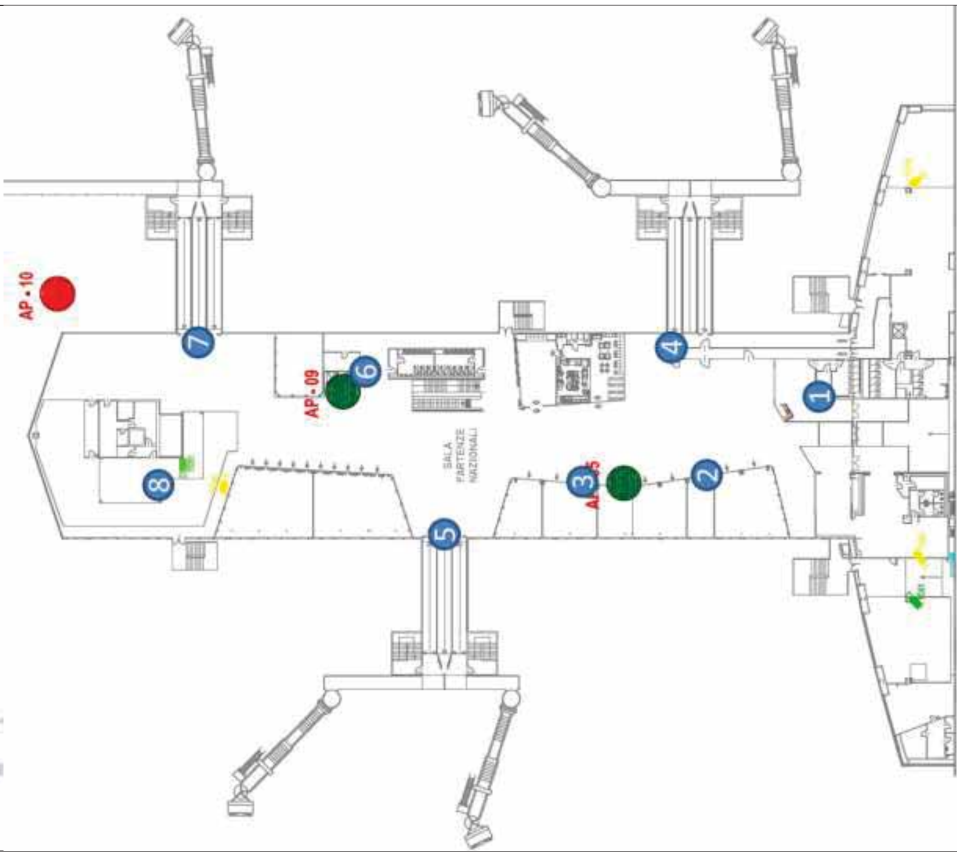
- Lots of clever people putting stuff together
- Software is open source
- https://github.com/mF2C/
- Fully CI/CD/devopsed
- https://hub.docker.com/search?q=mf2c&type=image
- Featureful platform
- Easy to implement now, add features later
- Some components have high TRL
- Recommender system uses Apache Mahout
- COMPSs from Barcelona Supercomputing Center (http://compss.bsc.es/) – HPC parallel execution with support for Java and python

What's ☹?

- Research software
  - Being written in parallel with app
  - Some components fully supported only during project lifetime
  - Lowish TRL – student programmers are not always RSEs
  - devops but not devsecops
- Featureful platform
  - Fairly high memory/CPU requirements
  - May deploy features we don't need

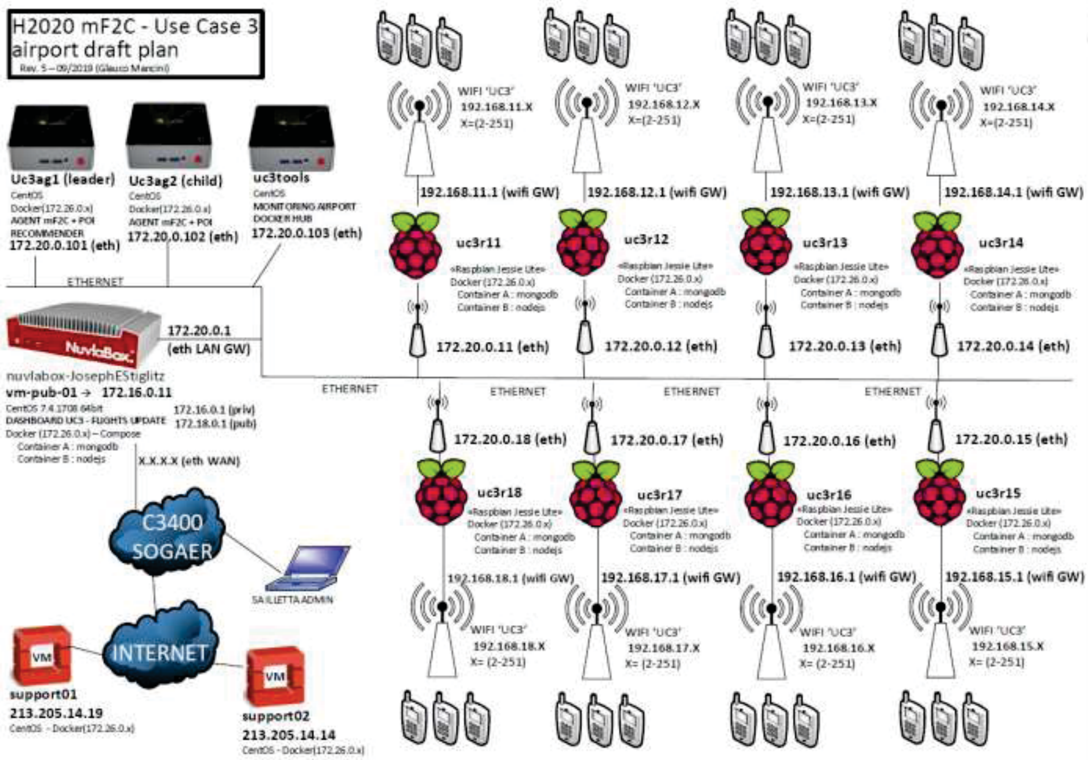
Go to gate!

Deployment in Cagliari, Sardinia



BLUE spots are RaspberryPI

RED and GREEN spots are Airport Wifi Access Points (to be used for better precision)



Security – testing & demo

A wireless security testing procedure has been defined to guarantee the end-to-end security

KALI distribution has been chosen as it offers a wide range of tools (FERN, NMAP, KISMET, etc.) for the most common attacks and is also available on both VM and container

The security procedure is based on the following steps:

- Planning – Gather information (detected APs, with hw, OS, sw, related version)
- Execution – Post Authentication (check security aspects as regular user)
- Execution – Unauthorized Access Attempt (check if an unauthorized person can gain access relying on one weakness)
- Post Execution – Reporting (with vulnerabilities found, details to reproduce it)

General Indicators for App

feature	detail	Benefits	Indicator
indoor application	Object positioning based on signal strength measures	• Precise and fast (local) calculation (no need to move data to cloud) • better data privacy management (GDPR compliance)	It should be Privacy oriented
Interactive application	real-time response needed for proximity marketing	• Fast & optimized response time leveraging the mF2C orchestration, load balancing/distribution, resource mgmt	• Low Latency • Fast Response time
Connectivity	Application based on continuous data communication (tight engagement)	• The use case take advantage of the redundant links at the edge (data connection is guaranteed, better resiliency)	It should be QoS oriented
Data intensive	It works with a huge amount of user's data (movements, choices, rates, preferences, ...)	• Load balancing, offloading, scalability, coming from the mF2C capabilities	• Response time

ENGINEERING
ENGINEERING SARDEGNA

Satisfaction guaranteed? – KPIs

- **Latency:** measured from the smartphone to fog and cloud devices
- **Response Time:** measured in the following scenarios
  - Smartphone to Fog (Raspberry)
  - Smartphone to Cloud (OpenStack)
  - Smartphone to Fog-Cloud (Raspberry+OpenStack)

  A battery of tests with increasing number of proximity requests are run, collecting data on response time under different loads.

  In case of fog-to-cloud scenario the percentage of requests served by fog is tracked as well (requests served by fog / total requests)

Business Case

What's in it for the airport?

Why should the airport want this?

Where are the travellers congregating?

Chamara CC-BY-SA https://commons.wikimedia.org/wiki/File:Herd_of_Elephant_in_Minneriya_National_Park.jpg

Why should the airport want this?

Travellers spend more money in shops (potentially)

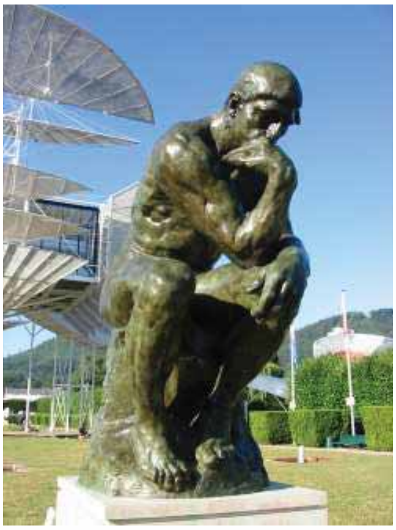
PD https://commons.wikimedia.org/wiki/File:100_million_front.JPG

Why should the airport want this?

Users can report emergencies

David Monniaux CC-BY-SA https://commons.wikimedia.org/wiki/Ambulance#/media/File:French_medical_VAB_dsc06842.jpg

Why should the airport want this?



Better user support

Le Penseur (Saint Dié des Vosges) – Christian Amet CC-BY
https://commons.wikimedia.org/wiki/Category:Le_Penseur#/media/File:Le_Penseur_de_Rodin_%C3%A0_Saint-Di%C3%A9.JPG

Why should the airport support this?

More relaxed travellers

Ken Wieland CC-BY-SA https://commons.wikimedia.org/wiki/File:1_Sannyasi_in_yoga_meditation_on_the_Ganges,_Rishikesh_cropped.jpg

Next Steps

and conclusion

Yes, it can be done

- If you've tried to navigate a large/unknown airport, this makes sense
- Strike balance between usefulness/privacy
- User selects choice
- IoT platform makes app development easier
- Building on research software is challenging, though
- Eventually the dust (components) will settle

Thanks!

=> Antonio.Salis@eng.it
=> jens.jensen@stfc.ac.uk

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730929. Any dissemination of results here presented reflects only the consortium view. The Research Executive Agency is not responsible for any use that may be made of the information it contains.

James Grant

University of Bath

Research Software Engineering Impact Showcase

A session consisting of a series of short talks from the RSE community using case studies to highlight impact followed by a panel discussion on how we can measure and promote the capability of RSEs to produce impact.

Abstract

Since the term was first coined in 2012, Research Software Engineers have been recognised as a distinct role in the Scientific computing ecosystem and today thousands of people worldwide self-identify as RSEs. With knowledge about applications, hardware and software, a focus on a good software engineering practices and collaborative working, RSEs often form the key link in the chain between end-users and compute providers (either onprem HPC, or cloud), realising the potential of these platforms to create real-world impact. This session will consist of short talks from the RSE community giving case studies and evidence of how RSEs have impacted on several communities including users of ARCHER, and commercial and industrial collaborators of the Hartree Centre, followed by a panel discussion on how we can measure and promote the capability of RSEs to produce impact. It will give an opportunity for CIUK stakeholders who are unaware of what RSEs have to offer a chance to see what RSEs can do for them, and find out how to access RSEs in their own context.

Research Software Engineering Impact Showcase

CIUK

CIUK 2019
Manchester
5th December

Talks:

- Dave Meredith
- Andy Turner
- James Grant

Panel:

- Christine Kitchen
- Barbara Montanari
- Kirsty Pringle

#CIUK
#researchsofteng
#rseimpact

Join at
slido.com
#RSEImpact

RSE
SOCIETY OF RESEARCH SOFTWARE ENGINEERING

UKRI
Science and Technology Facilities Council

Research Software Engineering Impact Showcase

UNIVERSITY OF BATH

- RSE a very brief history
- Why is impact an issue for RSE
- Session overview

RSE a very brief history

- Term coined at Software Sustainability Institute Collaborations Workshop in 2012 [software.ac.uk]

[SSI promote: Better Software, Better Research]

- Software [and its developers] is fundamental to research but developers lacked much, including a name.
- Led to creation of UKRSE Association and Society of Research Software Engineering (2019)
- RSE Conference started in 2016, 2019 saw >400 RSEs attend
- National organisations in Germany, Netherlands, Nordic, US

UKRSEA
Society of Research Software Engineering

- We will create a community to represent the UK's Research Software Engineers.
- We will raise awareness of Research Software Engineers and their fundamental role in research.
- We will campaign for the recognition and reward of Research Software Engineers.
- We will campaign for the Research Software Engineer to be adopted as a formal role within academia.
- We will organise regular events to allow Research Software Engineers to meet, exchange knowledge and collaborate.
- http://rse.ac.uk/about/
- https://society-rse.org/

Why is impact an issue for RSE?

- UKRSEA has been successful in campaigning for RSE:
  - 2 Calls for EPSRC RSE Fellowships
  - RSE support for Tier-2 HPC proposals
- Many universities now have central RSE Groups
  - Training
  - Project support
  - Specific expertise

Why is impact an issue for RSE?

- Without impact RSE will continue to struggle:
  - Funding
  - Recognition
  - Recruitment
  - Progression

https://cosden.github.io/RSE-career-path

Session overview

- How to measure impact of RSE?
- How to demonstrate impact of RSE?
- How does this fit in with all other challenges in HE?
- Cost, value, science, culture

Session overview

Talks (45 minutes):

- Dave Meredith, STFC
- Andy Turner, EPCC
- James Grant, University of Bath

Panel (45 minutes):

- Christine Kitchen, University of Cardiff
- Barbara Montanari, STFC
- Kirsty Pringle, University of Leeds

Audience Participation

Present slido

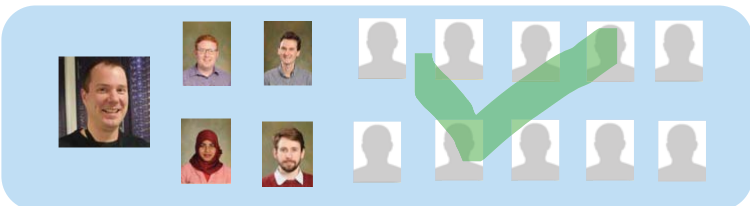
RSE @ Hartree Centre

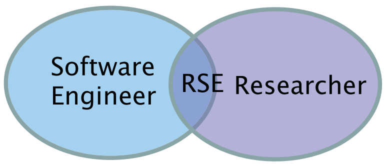
david.meredith@stfc.ac.uk
Research Software Engineering Group

Hartree: Industry facing:

- Remit is to promote the use of HPC, AI, Big Data & other goodies to give UK PLC a competitive advantage.

Hartree RSE Team





(Re)Engineering code to:

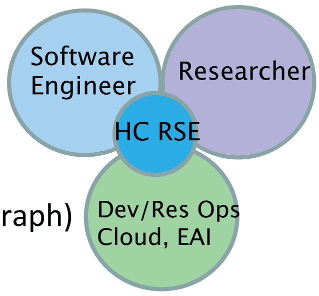
- Increase usability
- Maintainability
- Re-use

Apply best practices:

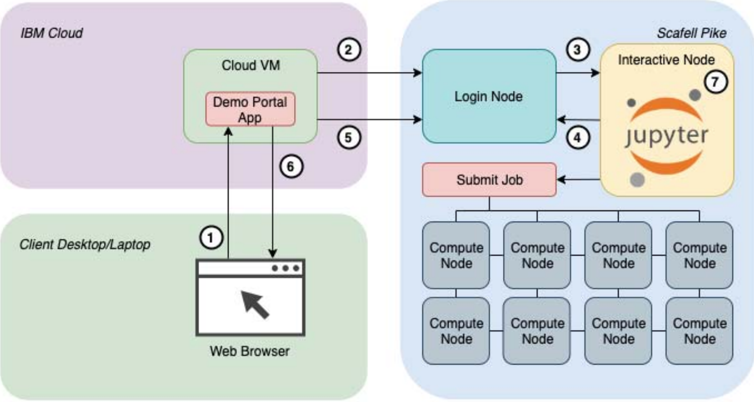
- Testing and CI
- Version Control
- Supportive Code Review
- Refactoring
- Design Patterns
- Documentation
- Design Thinking

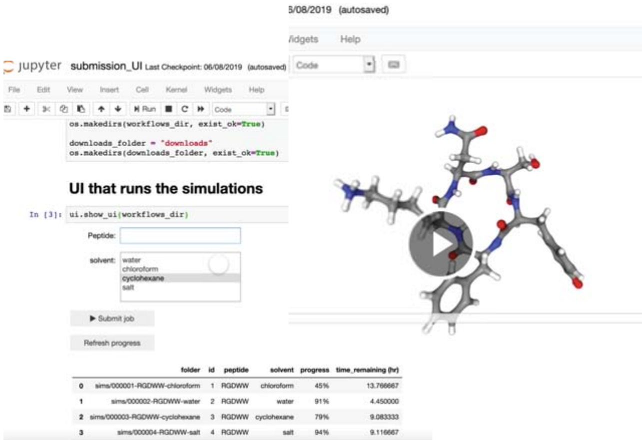
Tech:

- Visualization
- UIs
- Security
- HPC Workflows
- DBs (SQL, NoSQL, Graph)
- AI & Virtual Agents
- Cloud & Containers
- Client-server, microservices
- Mobile iOS/Android
- Webapps, SPAs, GraalVM, WebAssembly



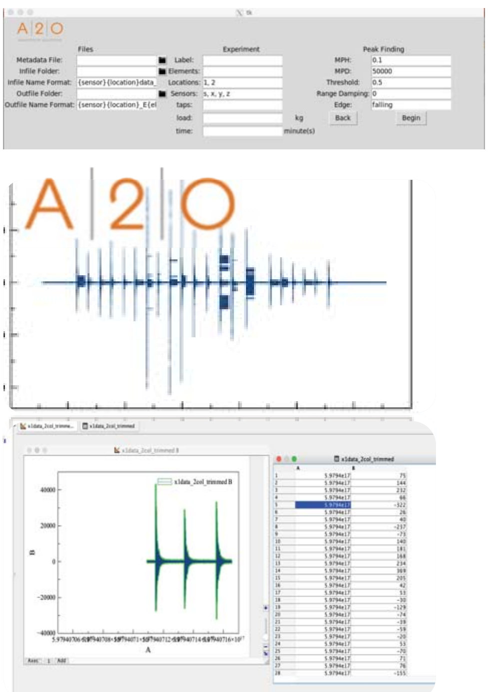
Supporting different styles of workload - Portal access to Interactive Notebooks on Scafell Pike

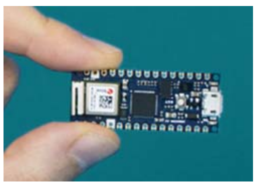


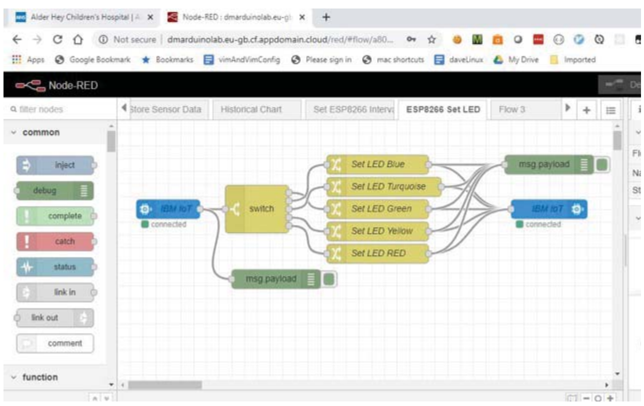


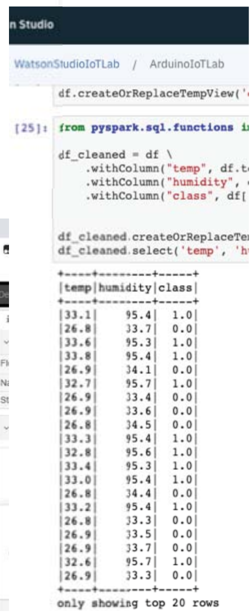
- Supports interactive working
- Registry of pre-canned apps

IoT / Edge

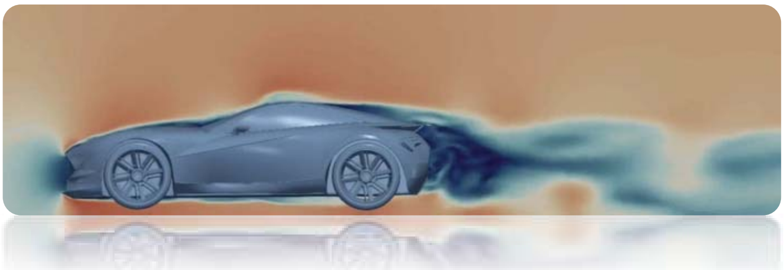


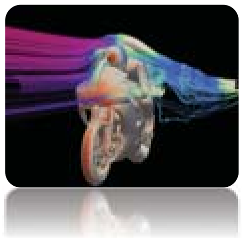






Virtual Wind Tunnel

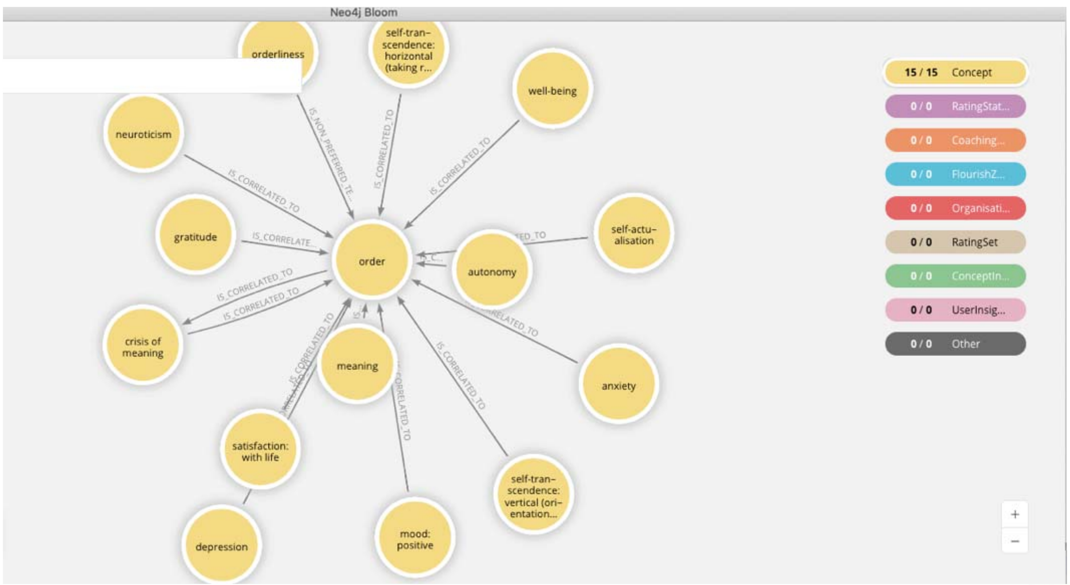




"Bringing the power of High Performance Computing to non-HPC experts."

BAC case study ! IMPACT

Flourish Zone



- IUK project
- Graph of interconnected concepts that affect flourishing in the workplace
- Recommender
- Rating Statements

AI Based Virtual Agents

1. AskOli (Alder Hey, Warrington)
2. Ufonia
3. Birmingham City Council Bot

NHS Alder Hey Children's NHS Foundation Trust

ufonia autonomous speech-based monitoring of health

Birmingham City Council

IBM WATSON

Hartree Centre
Science & Technology Facilities Council

Science & Technology Facilities Council

Typical OPD appointment

- Red: Explain / Trust / Engage / Q&A -an area that is expanding
- Blue: Diagnosis – not the time consuming part, also the part the clinicians like doing

Intro

Trust

Gather info

Examine

Diagnose

Explain/trust — Red is the expanding part

Admin/non value

The Problem:
- Clinicians spend 7mins of a 15min consultation on low value Q&A equating to £14.8M annually.
- In 2016/17, patient DNAs cost the trust an estimated £5M.

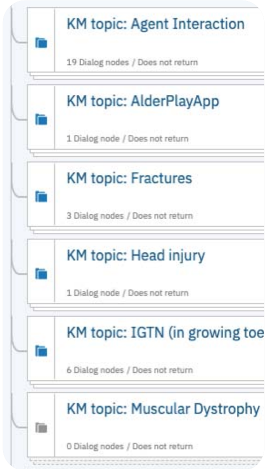
Courtesy: Iain Hennessey (Paediatric Surgeon, Director of Innovation, Alder Hey)

Inspired by Children

Science & Technology Facilities Council

- First real-world use of AI in a chatbot to improve patient experience in a hospital.
- Reduce patient anxiety through informative Q&A
- Reduce cancellations?
- Reduce consult time? (improve quality of consult)
- Reinvestment of 1min/consult = £2.1M/yr @AH
- Address CQC for proactive care & engagement

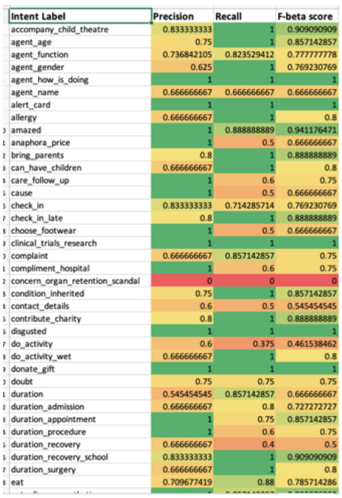
Working with hospital staff on building transferable KMs/skills



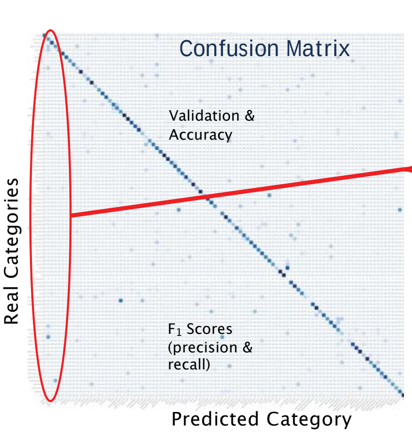
https://alderhey.nhs.uk

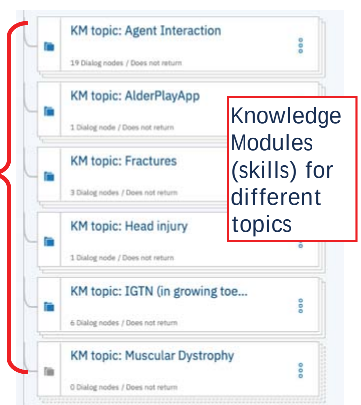
Inspired by Children

How do we know the AI is performing well ?

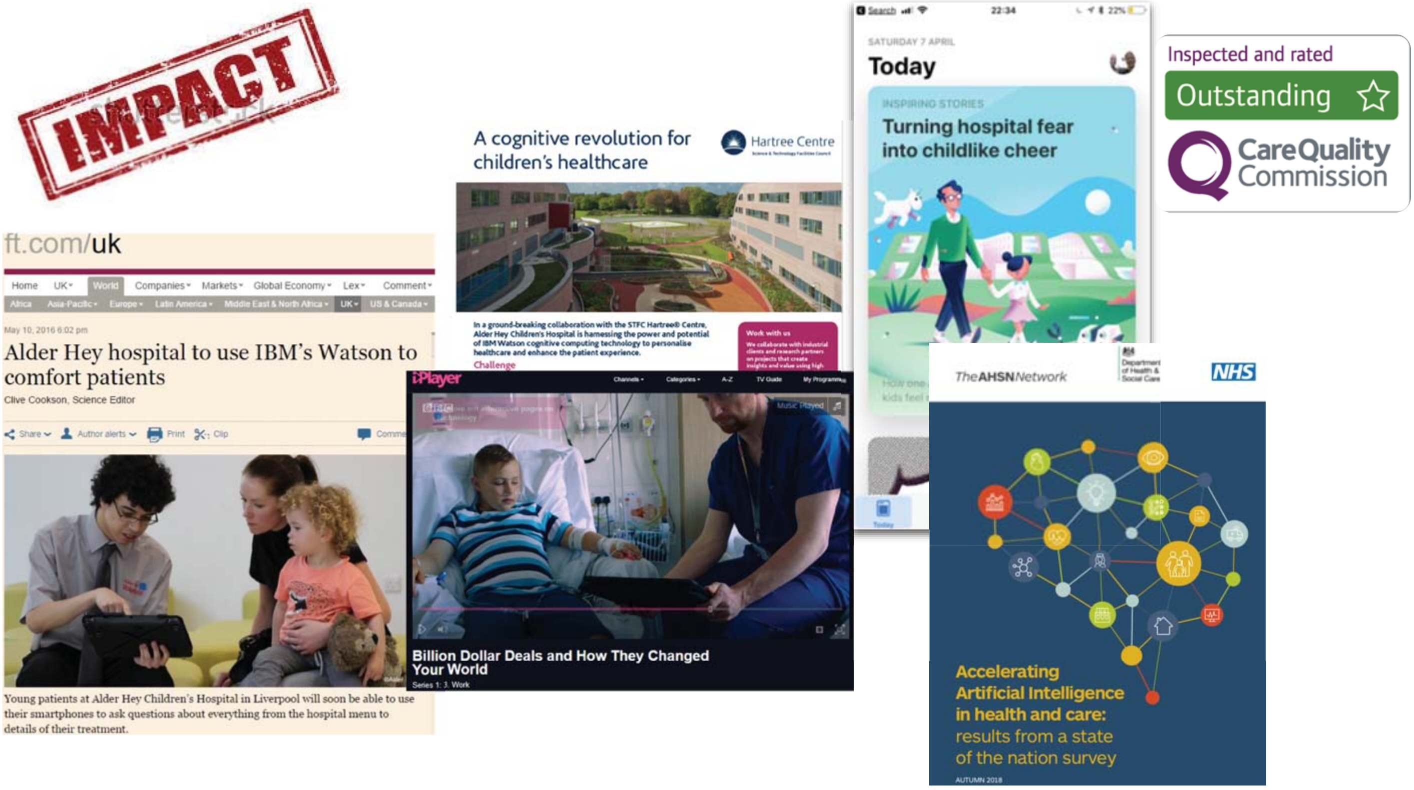


Intent Label	Precision	Recall	F-beta score
accompany_child_theatre	0.833333333	1	0.909090909
agent_age	0.75	1	0.857142857
agent_function	0.736842105	0.823529412	0.777777778
agent_gender	0.625	1	0.769230769
agent_how_is_doing	1	1	1
agent_name	0.666666667	0.666666667	0.666666667
alert_card	1	1	1
allergy	0.666666667	1	0.8
amazed	1	0.888888889	0.941176471
anaphora_price	1	0.5	0.666666667
bring_parents	0.8	1	0.888888889
can_have_children	0.666666667	1	0.8
care_follow_up	1	0.6	0.75
cause	1	0.5	0.666666667
check_in	0.833333333	0.714285714	0.769230769
check_in_late	0.8	1	0.888888889
choose_footwear	1	0.5	0.666666667
clinical_trials_research	1	1	1
complaint	0.666666667	0.857142857	0.75
compliment_hospital	1	0.6	0.75
concern_organ_retention_scandal	0	0	0
condition_inherited	0.75	1	0.857142857
contact_details	0.6	0.5	0.545454545
contribute_charity	0.8	1	0.888888889
disgusted	1	1	1
do_activity	0.6	0.375	0.461538462
do_activity_wet	0.666666667	1	0.8
donate_gift	1	1	1
doubt	0.75	0.75	0.75
duration	0.545454545	0.857142857	0.666666667
duration_admission	0.666666667	0.8	0.727272727
duration_appointment	1	0.75	0.857142857
duration_procedure	1	0.6	0.75
duration_recovery	0.666666667	0.4	0.5
duration_recovery_school	0.833333333	1	0.909090909
duration_surgery	0.666666667	1	0.8
eat	0.709677419	0.88	0.785714286





Hartree Centre
Science & Technology Facilities Council



Birmingham City Council

- BCC call center >2.5m calls/yr
- Largest city council in Europe
- Many simple questions about tax

- Cost savings IMPACT
- 0.5m calls: $13k based on 10 API conv
- Less mentioned: Job displacement ?
- No, frees us from mundane to handle complex calls

10 API calls (~2p)

...

Consider your website integration – iframe, SPA, full-screen

---

ufonia

Autonomous speech-based monitoring of health

The Problem:
With an ageing population, there are growing concerns about the sustainability of healthcare & monitoring costs.

Ufonia: Autonomous health monitoring over the phone by a bot.

"Everyone knows how to have a conversation. Voice is an engaging platform that mirrors the natural doctor to patient consultation and we can deliver it using a telephone – without the potential barriers of internet connectivity, downloads or training. It's a natural choice for the ageing population while having widespread applicability."

– Nick de Pennington, CEO, Ufonia

IMPACT

Greater reach via ubiquitous phone.

Reduce monitoring costs through automation.

Increase the volume of monitoring calls to help catch problems early.

Buckinghamshire Healthcare NHS Trust

Oxford University Innovation | University of Oxford

My Clinical Outcomes

Oxford Academic Health Science Network

Hartree Centre

OXFORD Knee Score (OKS)

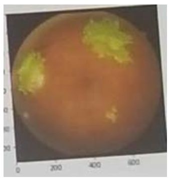
	User Utterance	Classification	Score
oks_q1	"it's been very severe "	Severe	0
oks_q2	"I would say moderate difficulty "	Moderate trouble	2
oks_q3	"I had extreme difficulty with that "	Extreme difficulty	1
oks_q4	"but half an hour "	16 to 30 minutes	3
oks_q5	"well that's been very painful "	Very painful	1
oks_q6	"all the time "	All of the time	0
oks_q7	"I find that to be moderately differ "	With moderate difficulty	2
oks_q8	"I would say most nights "	Most nights	1
oks_q9	"do a little bit "	A little bit	3
oks_q10	"most of the time "	Most of the time	1
oks_q11	"well I have moderate difficulty with that "	With moderate difficulty	2
oks_q12	"I couldn't it's impossible "	No, impossible	0
Total Score:			*16*

U "it is a little bit red "

B It's normal for the eye to be quite red for the first few days after surgery, especially near the nose where the anaesthetic is injected. Compared to the day after surgery, do you think the eye is more red now or less red?

U "%HESITATION it's much less red "

B And how is your vision is it improving?



U "yes it is "

B Okay. And how are you managing with the eye drops?

U "they sting my eyes "

B It's quite normal for the eyedrops to hurt a bit when you use them. Try to keep using them as they will help with the healing of your eye. Was there anything else you wanted to talk about with your eye drops?

U "no that's fine "

Main challenge is Speech to Text:

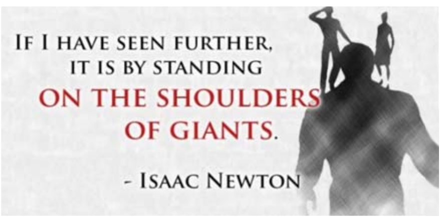
- Regional accents
- Hesitations & rambling answers
- Clashing user utterances - ear-muff' the bot while it speaks ....

How's the knee doing?

- "When talking with a human, I'll talk like a human, If I know I'm talking to a bot, then it's like talking to a five-year old - I will be precise."
- Too human can be creepy (opinion only warning)

RSE Impact (Opinion Only Warning!)

- Job creation
- Cost savings
- Revenue generation
- Abstracting the complex
- Enabling new ways of working
- Supporting new types of workload
- Improving engagement and customer experience
- Increase levels of monitoring
- Automation and speedup



Standing on the shoulders of giants !

Hartree Centre
Science & Technology Facilities Council

Thank you

david.meredith@stfc.ac.uk

Hartree Centre
Science & Technology Facilities Council

Find out more:

- hartree@stfc.ac.uk
- hartree.stfc.ac.uk
- /company/stfc-hartree-centre
- @hartreecentre

Hartree Centre
Science & Technology Facilities Council

Benefits of the eCSE Programme

Andy Turner, Lorna Smith, EPCC
With thanks to Chris Johnson, Neelofer Banglawala, Xu Guo, Jo Beech-Brandt and Alan Simpson

Background to eCSE Programme

- Allocated funding to the UK computational science community for software development through a series of funding calls over a period of 6 years
- eCSE is a significant source of funding for RSEs across the UK
- All HEIs are able to apply for projects
- It is important to be able to demonstrate the benefit of the programme to different funding bodies, to help secure future funding of this type
- This talk gives more details of the programme and includes data on how the money was spent

eCSE Programme

- Goal: to deliver a funding programme that is fair, transparent, objective and consistent. Aims:
- Aims
  - To enhance the quality, quantity and range of science produced on the ARCHER service through improved software;
  - To develop the computational science skills base, and provide expert assistance embedded within research communities, across the UK;
  - To provide an enhanced and sustainable set of HPC software for UK science.
- Scope
  - Any HEI may apply, technical staff members may be located in the HEI or at a third party institution, or may be an ARCHER team member
- Due to an extension, the final call is currently underway
  - Proposals submitted, currently under review
- Many projects complete (90%)

archer

epcc

THE UNIVERSITY OF EDINBURGH

Benefits

- Measuring benefits is an on-going process while projects are still active
  - A high quality, fair and objective eCSE selection process, delivering maximum value to the community;
  - Increased science productivity;
    - Including financial saving reinvested to allow scientists to achieve more science from the same resource allocation
  - Increased novelty and range of science on the system, both traditional and new;
  - Enhanced computational science skills base across the UK.
- Programme outputs and metrics link to these benefits

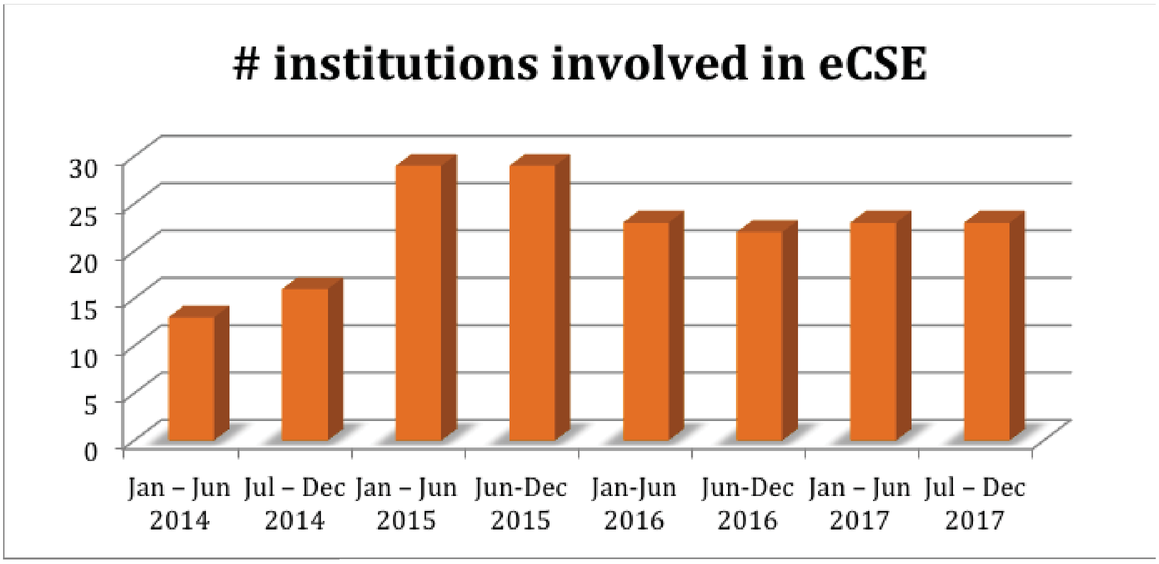
archer

epcc

THE UNIVERSITY OF EDINBURGH

A high quality, fair and objective eCSE

- Regular calls and independent panel members
- Not for profit, FEC costing model
- Open to all, not just organisations using FEC

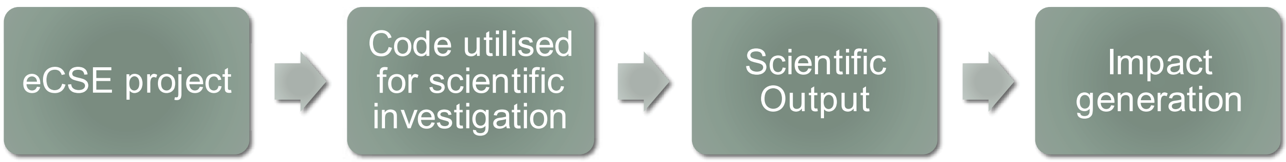


archer

epcc

THE UNIVERSITY OF EDINBURGH

Science productivity, novelty and range



- eCSE projects are early in the timeline
  - Some benefits may not be seen for years after the project is complete
- The eCSE involves a set of separate projects, but looking to demonstrate benefit across the whole programme
- Solution is to measure a range of benefits
  - One size doesn't fit all

archer

epcc

THE UNIVERSITY OF EDINBURGH

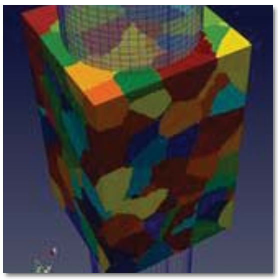
Science productivity, novelty and range

eCSE project → Code utilised for scientific investigation → Scientific Output → Impact generation

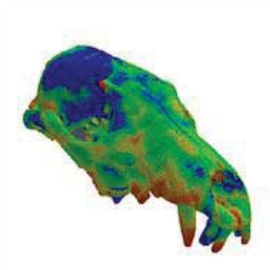
Computational achievements e.g code availability, publications; future science benefits; enhanced skills → Enhanced scientific productivity; Increased size of user community → Scientific achievement e.g. publications → Impact e.g. case studies

Computational achievements, future science benefit

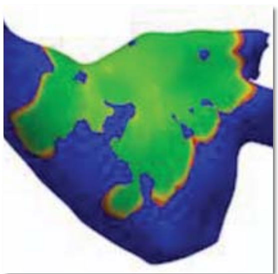
Shterenlikht, Margetts, Emerson



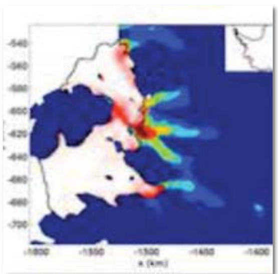
Fagan, Bethune



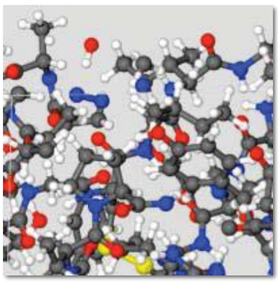
Sherwin, Cantwell, Moxey



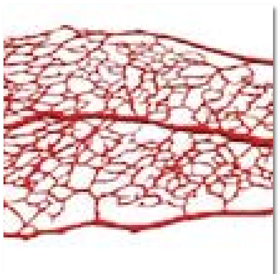
Jones, Goldberg, Holland, Ferreira



Probert, Hasnip, Refson, Bush



Bernabeu, Krüger, Coveney, Hetherington, Silva



Develop the computational science skills base

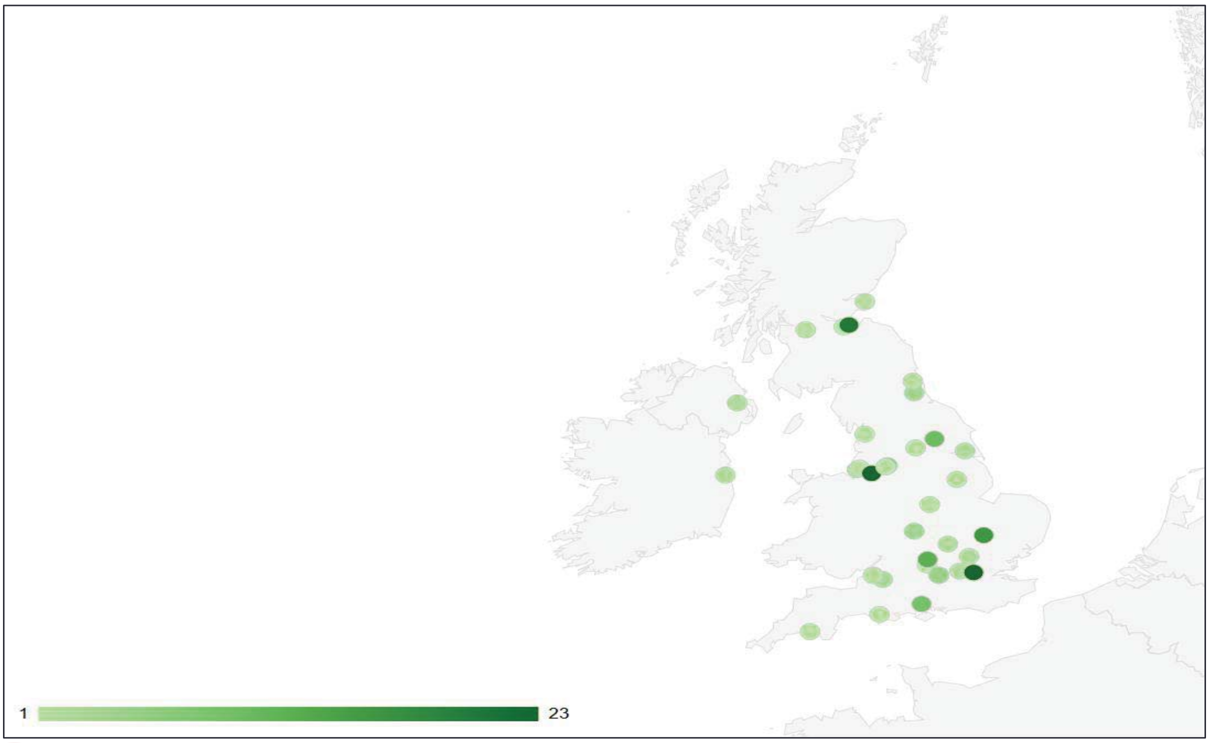
- A key outcome from the eCSE programme relates to people
- Aim is to develop the computational science skills base
- And provide expert assistance *embedded* within research communities, across the UK
- Track location of PIs/Co-Is/technical members of staff

archer

epcc

THE UNIVERSITY OF EDINBURGH

Skilled embedded workforce



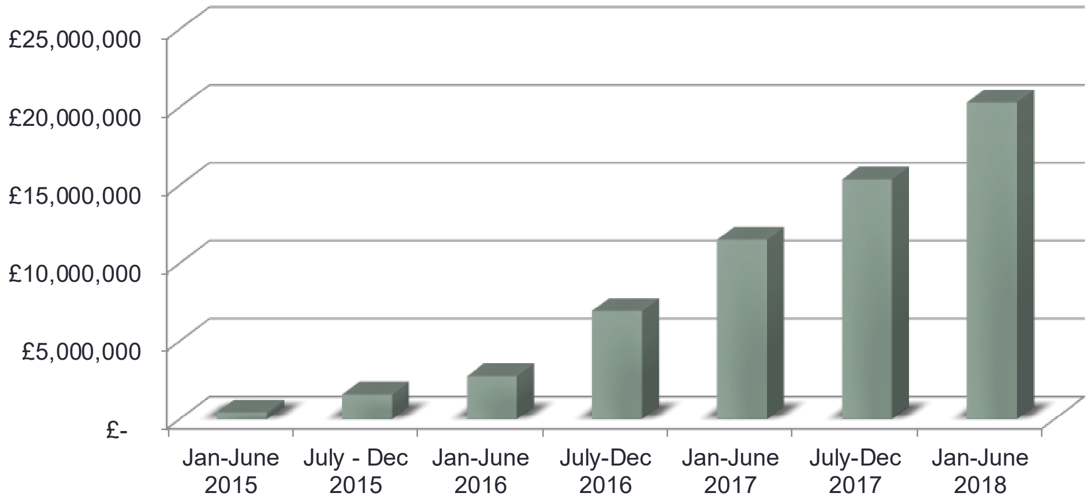
archer

epcc

THE UNIVERSITY OF EDINBURGH

Increased science productivity

- Financial saving reinvested to allow scientists to achieve more science from the same resource allocation
- Not all projects contribute to this particular benefit, depends on the nature of the work
- Overall cost of the eCSE programme £6M, reported benefits to date £20.3M

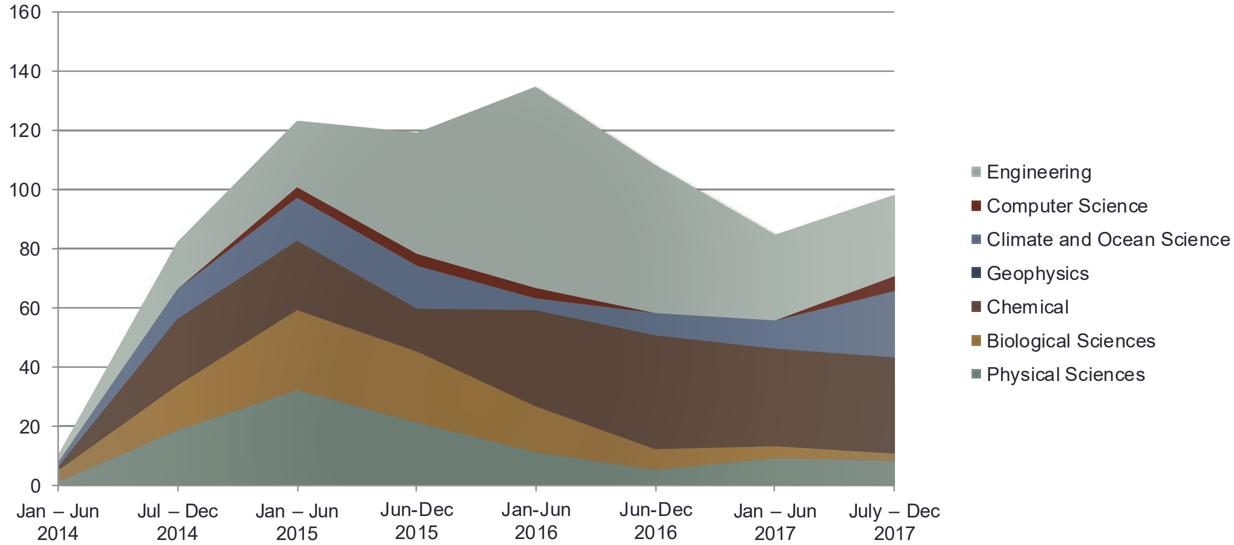


archer

epcc

Range of science

- Since the 4th eCSE call, we actively encouraged proposals from "New Communities”
- 10 such proposals were funded – 11% of all projects, 18% average across relevant calls

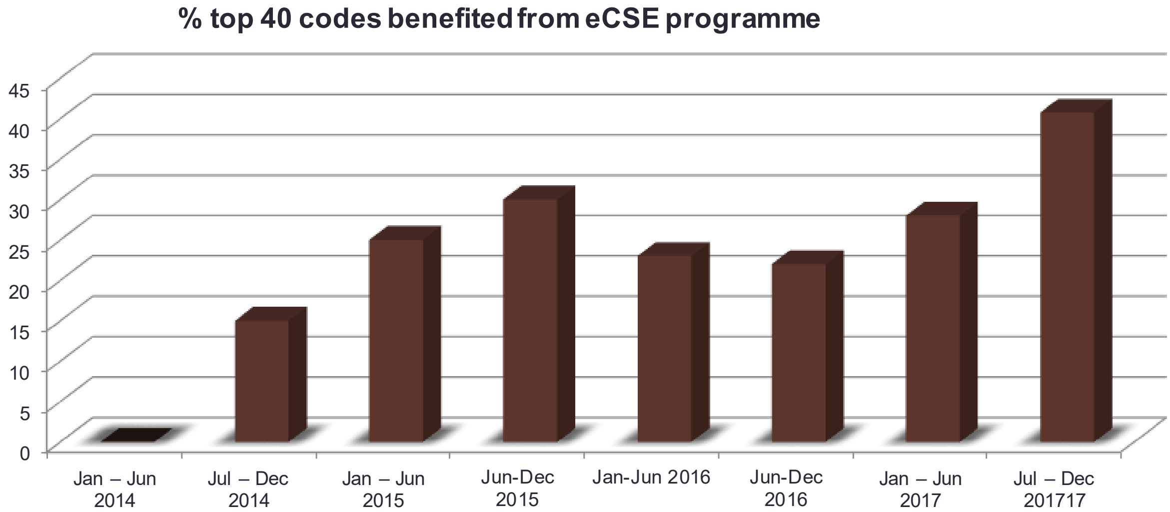


archer

epcc

Increased range of science

- In the last 6-month period, over 40% of the top 40 codes had benefitted from some form of eCSE support



archer

epcc

THE UNIVERSITY OF EDINBURGH

Conclusions

- Graphs demonstrate that eCSE programme has funded RSEs:
  - In a broad range of scientific areas
  - And at many different HEIs
- Measuring financial benefits is tricky but helps make the case that investment in RSE support is essential to extract maximum benefit from the hardware

- While some projects are still running, it is clear that the programme has already:
  - provided a consistent, fair and not-for-profit funding programme
  - Funded a wide variety of projects
  - Enhanced the skills base of the UK computational community across the UP
  - Generated considerable financial benefits (more than 3x return on investment)
- As the codes continue to be used we anticipate even more high quality science will be performed

archer

epcc

THE UNIVERSITY OF EDINBURGH

Looking Forward

- How could we improve eCSE?
- What metrics are most important to demonstrate that investment in RSE support is valuable?
- Integration between Tier-1 and Tier-2 is essential for UK
  - Should eCSE also fund projects on Tier-2?
- How can we increase the range of HEIs further?
- Are there any barriers that we could erode?

RSE Communities: The Research Software Reactor

James Grant, RSE, University of Bath

rjg20@bath.ac.uk

CIUK 2019

5th December 2019

HPC/RSE Communities

- HPC-SIG
- HPC Champions
- Tier-2 HPC RSE Community

HPC/RSE Communities

- HPC-SIG
- HPC Champions
- Tier-2 HPC RSE Community
- Research Software Reactor
- Regional Isambard Community
- Local RSE collective

HPC/RSE Communities

- HPC-SIG
- HPC Champions
- Tier-2 HPC RSE Community
- **Research Software Reactor**
- Regional Isambard Community
- Local RSE collective

Research Software Reactor

- What prompted it?
- What is it?
- What has happened?
- What is planned?
- What impact will it have?

What prompted it?

- 2 Day Summit MS Executive Briefing Centre
- ~20 RSEs across EMEA:
  - Netherlands, Switzerland, Romania, South Africa
  - UK: Bath, Imperial, King's, Leeds, Manchester
- Co-located with summit for 'grown-ups'
- Aimed to discuss issues facing community
- {Why isn't/What is preventing} research adopting Cloud?

What did we do?

- RSE Contributed talks
- MS: Azure tools/services; Learn
- Round table discussions
- Group discussion

Diversity

- The lack of diversity was embarrassing
- Tania Allard, Microsoft/SocRSE Committee, only female RSE.

UNIVERSITY OF BATH

Why RSEs?

Central IT

RSEs

Researchers

UNIVERSITY OF BATH

Why RSEs?

Central IT

Researchers

UNIVERSITY OF BATH

Why doesn't research use Cloud?

Data security | Budgeting | Authentication | Funding | Culture | Deployment

UNIVERSITY OF BATH

Why doesn't research use Cloud?

Data security | Budgeting | Authentication | Funding | Culture | Deployment

What is the Research Software Reactor?

Maintain the enthusiasm of the meeting
Developing the community #cloudcomputing on RSE slack

Outcomes/Deliverables:

Skills and training: Research Software Reactor

What is the Research Software Reactor?

First sprint 20th-22nd May Imperial/Reactor

- 3day play on Azure with MS staff on hand
- First hand experience deploying cloud resources
- Developing resources for the community:
  - Training materials
  - 'blueprints' for research workloads.
  - FAQ

What is the Research Software Reactor?

- Cloud in itself is not the objective
- RSE/RCA need to be able to help researchers deliver
  - as quickly
  - cost effectively
  - on most appropriate resource

as possible

What has happened?

- CycleCloud (+Dask)
- Binderhub – Sarah Gibson workshop @RSECon19!
  - Single button deployment for binder of Azure

What is planned?

- DevOps Sprint: 9-10th January 2020
- RSR - AWS Sprint: 6-7th April 2020

- https://research-software-reactor.github.io/

What is planned? I have a dream.

Researcher preparing proposal

Research Compute Analyst
- Spins up resources
- Explore compute setup
- Esimate cost for proposal

Successful proposal

RSE/RCA
- Optimise performance for production compute
- Optimise code

Efficient time and cost to research

Impact of RSE/RSR (Community)

- Creating resources and learning materials:
  - Individuals can learn and upskill
  - Community benefits from resources and individuals experience
  - Wider UK-EMEA benefit from resources beyond HE
- Ultimately enabling better research: quicker, effective, reproducible
- Cultural change:
  - Research doesn't value compute resources (cycles, memory, data, software)
  - Understanding the cost of compute -> value resources -> value the software

Thanks to

- Gerard Gorman, Imperial College
- Brad Tipp, Tania Allard, Lee Stott, Microsoft
- Lucy Antysz, Francis Dauncey, AWS
- Iain Bethune, STFC
- Damian Jones, Organisers, CIUK 2019

Thank you for listening ...

Jeyan Thiyagalingam

Science and Technology Facilities Council

Machine Learning as a Cheaper Alternative to HPC Approaches in Science

Jeyan Thiyagalingam heads the Scientific Machine Learning (SciML) research group at the Rutherford Appleton Laboratory, Science and Technology Facilities Council (STFC-RAL), Harwell. The SciML group focuses on the development and application of machine learning and signal processing techniques for addressing fundamental scientific problems. Prior to joining STFC-RAL, he was an assistant professor in the school of Electrical Engineering, Electronics and Computer Sciences at the University of Liverpool, and prior to that at the University of Oxford both as a post-doctoral researcher and later as a James Martin Fellow. He has also worked in industry, including MathWorks UK. His research interests and expertise are on machine learning models, data processing algorithms, and signal processing. He is a Fellow of the British Computer Society and part of the Alan Turing Institute. He also serves as an associate editor for the Patterns, and the Concurrency & Computation: Practice and Experience Journals.

Abstract

High performance computing-based approaches, such as long-running simulations, play a crucial role in science – from theoretical computational physics to identifying cloud in satellite imagery. This talk will look into how machine learning has started changing this landscape, as a computationally cheaper alternative to conventional HPC approaches. The talk will particularly focus on the influence of machine learning on different domains of sciences with relevant examples.

Martin Turner

University of Manchester

Linking National Imaging Facilities with HPC and Research Software Engineering centres – and their use on a Big Data problem

Dr Martin Turner emphasises on research and services for Video, Computation and Visualization. Currently he is a research Relationship Manager in the University of Manchester and is a Visiting Scientist within the Scientific Computing Division in STFC; after overlapping secondments being Visualisation Director for the Harwell Imaging Partnership (HIP) at STFC/RAL and as a Visualisation Group Leader at STFC/DL. Related to e-Science and Grid infrastructure he has worked as project manager for numerous RCUK and JISC funded Virtual Research Environment, video and data projects.

Abstract

The Science and Technology Facilities Council (STFC, part of the government UKRI) fund and manage some of the largest imaging capture Facilities in the UK; including the Diamond Light Source (the national x-ray synchrotron) and ISIS (the neutron and muon spallation source). We make a strong comparison with these multi-million pound national imaging facilities to HPC services; both aim to run as continuously as possible 24/7, producing streams of data needing analysis for science and business needs.

For example the DLS beamlines i12/i13/DIAD can daily produce terabytes of imaging data, and the ISIS IMAT (Imaging and Materials Science & Engineering) beamline requires semi-realtime interactivity visualisation streaming for 100+GB datasets; and related storage, archiving and post processing.

In this session we show how a planned Python pipeline infrastructure with HPVisualisation architectures are combining data streaming with HPC. The Core Imaging Library (https://www.CCPi.ac.uk), is a python framework for image processing including data loading, preprocessing, reconstruction, postprocessing and visualisation. CIL is designed for a wide range of tomographic data, including parallel- and cone-beam, 2D and 3D cases as well as 4D dynamic and spectral. The modular design of CIL allows a variety of customised algorithms to be constructed by the user, in addition to pre-defined commonly used algorithms. This is a software glue to integrate facilities with HPC where we wish to minimise data movement.

Martyn Guest

ARCCA, Cardiff University

Application Performance on Multi-Core Processors: Performance Analysis of the AMD EPYC Rome Processors

Professor Martyn Guest has led a variety of high performance and distributed computing initiatives in the UK. He spent three years as Senior Chief Scientist and Group Leader of the HPC Chemistry Group at PNNL, before returning to the UK as Associate Director of Daresbury's Computational Science and Engineering Department. He joined Cardiff University in April 2007 and is their Director of Advanced Research Computing as well as Technical Director of the Supercomputing Wales programme. Martyn has provided HPC consultancy to organisations in both the UK and abroad. His research interests cover the development and application of computational chemistry methods. He is lead author of the GAMESS-UK electronic structure program, and has written or contributed to more than 250 articles.

Abstract

This session will overview application performance on a variety of clusters, primarily focusing on the AMD EPYC Rome family of processors. Using the Intel Skylake Gold 6148 as the baseline, an assessment is made across a variety of Rome SKUs (e.g., the 7702, 7742, 7452 and 7502), with system interconnects from both Mellanox and Intel. Our analysis will involve the familiar parallel benchmark performance using community codes from Molecular Dynamics (DL_POLY, LAMMPS, Gromacs, NAMD), Quantum Chemistry (GAMESSUS and GAMESS-UK), Materials Science (VASP and Quantum Espresso), Computational Engineering (OpenFOAM) together with the NEMO code (Ocean General Circulation Model). The challenge now is how to present a 'like for like' comparison given the vast array of core densities and whether this should now be on a "node-by-node" basis rather than the traditional "core-by-core" consideration.

COMPUTING INSIGHT UK 2019
Computing The Future
5 - 6 DECEMBER 2019
MANCHESTER, UK

UK Research and Innovation

Performance Analysis of the AMD EPYC Rome Processors

Jose Munoz, Christine Kitchen & Martyn Guest

Advanced Research Computing @ Cardiff (ARCCA) & Supercomputing Wales

Introduction and Overview

CARDIFF UNIVERSITY PRIFYSGOL CAERDYD

- Presentation part of our ongoing assessment of the performance of parallel application codes in materials & chemistry on high-end cluster systems.
- Focus here on systems featuring the **current high-end processors from AMD** (EPYC Rome SKUs – the 7502, 7452, 7702, 7742 etc.).
  - Baseline clusters: the SNB e5-2670 system and the recent Skylake (SKL) system, the **Gold 6148/2.4 GHz** cluster – “Hawk” – at Cardiff University.
  - Major focus on two AMD EPYC Rome clusters featuring the 32-core **7502 2.5GHz** and **7452 2.35 GHz**.
- Consider performance of both synthetic and **end-user applications**. Latter include molecular simulation (**DL_POLY, LAMMPS, NAMD, Gromacs**), electronic structure (**GAMESS-UK & GAMESS-US**), materials modelling (**VASP**, **Quantum Espresso**), computational engineering (**OpenFOAM**) plus the **NEMO** code (Ocean General Circulation Model).
  - Seven in Archer Top-30 Ranking list: **https://www.archer.ac.uk/status/codes/**
- Scalability analysis by **processing elements (cores)** and by **nodes** (guided by ARM Performance Reports).

AMD EPYC Rome multi-chip package

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

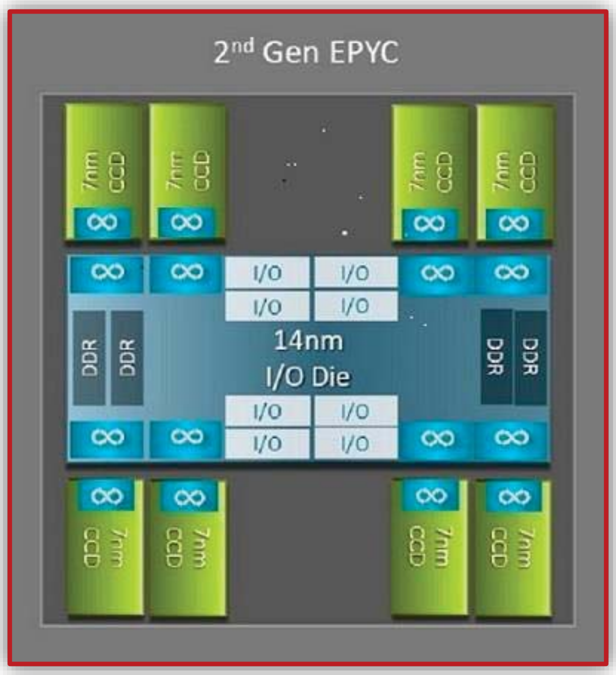


Figure. Rome multi-chip package with one central IO die and up to eight-core dies.

- In Rome, each processor is a multi-chip package comprised of up to 9 **chiplets** as shown in the Figure.
- There is **one central 14nm I/O die** that contains all the I/O and memory functions – memory controllers, Infinity fabric links within the socket and inter-socket connectivity, and PCI-e.
- There are **eight memory controllers per socket** that support eight memory channels running DDR4 at 3200 MT/s. A single-socket server can support up to 130 PCIe Gen4 lanes. A dual-socket system can support up to **160 PCIe Gen4 lanes**.

AMD EPYC Rome multi-chip package

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

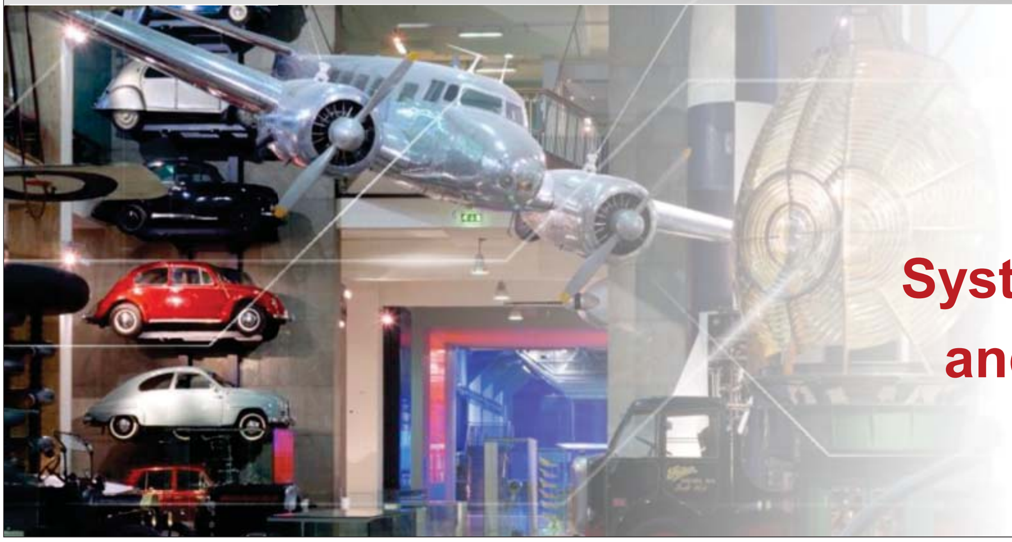
Figure A CCX with four cores and shared 16MB L3 cache

- Surrounding the central IO die are up to **eight 7nm core chiplets**. The core chiplet is called a **Core Cache die** or CCD.
- Each CCD has CPU cores based on the **Zen2 micro-architecture**, L2 cache and 32MB L3 cache. The CCD itself has two Core Cache Complexes (CCX), each CCX has up to four cores and 16MB of L3 cache.
- The figure shows a CCX.
- The different Rome CPU models have different numbers of cores, but all have one central IO die.

Rome CPU models evaluated in this study

CPU	Cores per Socket	Config	Base Clock	TDP
7742	**64c**	**4c per CCX**	**2.2 GHz**	**225W**
7502	**32c**	**4c per CCX**	**2.5 GHz**	**180W**
7452	**32c**	**4c per CCX**	**2.35 GHz**	**155W**
7402	24c	3c per CCX	2.8 GHz	180W

Performance Analysis of the AMD EPYC Rome Processors



Systems, Software and Installation

Baseline Cluster Systems

Cluster	Configuration
	Intel Sandy Bridge Cluster
"Raven"	128 x Bull|ATOS b510 EP-nodes each with 2 Intel Sandy Bridge E5-2670 (2.6 GHz), with Mellanox QDR infiniband.
	Intel Skylake Cluster
Supercomputing Wales "Hawk"	"Hawk" – Supercomputing Wales cluster at Cardiff comprising 201 nodes, totalling 8,040 cores, 46.080 TB total memory. • CPU: 2 x Intel(R) Xeon(R) Skylake Gold 6148 CPU @ 2.40GHz with 20 cores each; RAM: 192 GB, 384GB on high memory and GPU nodes; GPU: 26 x nVidia P100 GPUs with 16GB of RAM on 13 nodes. • Mellanox IB/EDR infiniband interconnect.

Partition Name	# Nodes	Purpose
compute	134	Parallel and MPI jobs (192 GB)
highmem	26	Large memory jobs (384 GB)
GPU	13	GPU and Cuda jobs
HTC	26	High Throughput Serial jobs

The available compute hardware is managed by the **Slurm job scheduler** and organised into 'partitions' of similar type/purpose.

AMD EPYC Rome Clusters

Cluster / Configuration
AMD Minerva cluster at the Dell EMC HPC Innovation Lab – Number of AMD EPYC Rome sub-systems with Mellanox EDR and HDR interconnect fabrics
10 x Dell EMC PowerEdge C6525 nodes with EPYC Rome CPUs running SLURM; • **AMD EPYC 7502 / 2.5 GHz**; # of CPU Cores: 32; # of Threads: 64; Max Boost Clock: 3.35 GHz Base Clock: **2.5 GHz; L3** Cache 128 MB**;** Default TDP / TDP: 180W; Mellanox ConnectX-4 EDR **100Gb/s** • System reduced from **ten to four cluster nodes** during the evaluation period.
64 x Dell EMC PowerEdge C6525 nodes with EPYC Rome CPUs running SLURM; • **AMD EPYC 7452 / 2.35 GHz**; # of CPU Cores: 32; # of Threads: 64; Max Boost Clock: 3.35 GHz Base Clock: **2.35 GHz; L3** Cache 128 MB**;** Default TDP / TDP: 155W; Mellanox ConnectX-6 HDR100 **200Gb/s**
• Number of smaller cluster nodes available – 7302, 7402, 7702 – these do not feature in the present study

AMD EPYC Rome Clusters

Cluster / Configuration
AMD Daytona cluster at the AMD HPC Benchmarking Centre – AMD EPYC Rome sub-systems with Mellanox EDR interconnect fabric
32 nodes with EPYC Rome CPUs running SLURM; • **AMD EPYC 7742 / 2.25 GHz**; # of CPU Cores: 64; # of Threads: 128; Max Boost Clock: 3.35 GHz Base Clock: **2.25 GHz; L3** Cache 256 MB**;** Default TDP / TDP: 225W; Mellanox EDR **100Gb/s**
AMD Daytona_X cluster at the HPC Advisory Council HPC Centre – AMD EPYC Rome system with Mellanox ConnectX-6 HDR100 interconnect fabric
8 nodes with EPYC Rome CPUs running SLURM; • **AMD EPYC 7742 / 2.25 GHz;** # of CPU Cores: 64; # of Threads: 128; Max Boost Clock: 3.35 GHz Base Clock**: 2.25 GHz; L3** Cache 256 MB**;** Default TDP / TDP: 225W; • **Mellanox ConnectX-6 HDR 200Gb/s InfiniBand/Ethernet** • Mellanox HDR Quantum Switch QM7800 40-Port 200Gb/s HDR InfiniBand • Memory: 256GB DDR4 2677MHz RDIMMs per node • **Lustre Storage, NFS**

The Performance Benchmarks

- The ***Test suite*** comprises both **synthetics & end-user applications**. Synthetics limited to **IMB** benchmarks (*http://software.intel.com/en-us/articles/intel-mpi-benchmarks*) and **STREAM**
- Variety of "open source" & commercial end-user application codes:

DL_POLY classic, DL_POLY-4 , LAMMPS, GROMACS and NAMD (molecular dynamics)

Quantum Espresso and **VASP** (ab initio Materials properties)

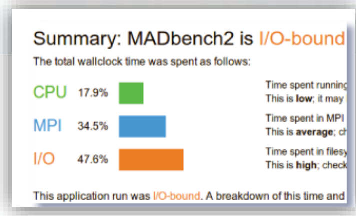
GAMESS-UK and GAMESS-US (molecular electronic structure)

OpenFOAM (engineering) and **NEMO** (ocean modelling code)

- These stress various aspects of the architectures under consideration and should provide a level of insight into why particular levels of performance are observed e.g., ***memory bandwidth and latency, node floating point performance and interconnect performance (both latency and B/W) and sustained I/O performance***.

Analysis Software - Allinea|ARM Performance Reports

Provides a mechanism to characterize and understand the performance of HPC application runs through a single-page HTML report.



- Based on Allinea MAP's adaptive sampling technology that keeps data volumes collected and **application overhead low**.
- Modest application slowdown (ca. 5%) even with 1000's of MPI processes.
- **Runs on existing codes: a single command added to execution scripts.**
- If submitted through a batch queuing system, then the submission script is modified to load the Allinea module and add the 'perf-report' command in front of the required mpirun command.

perf-report mpirun $code

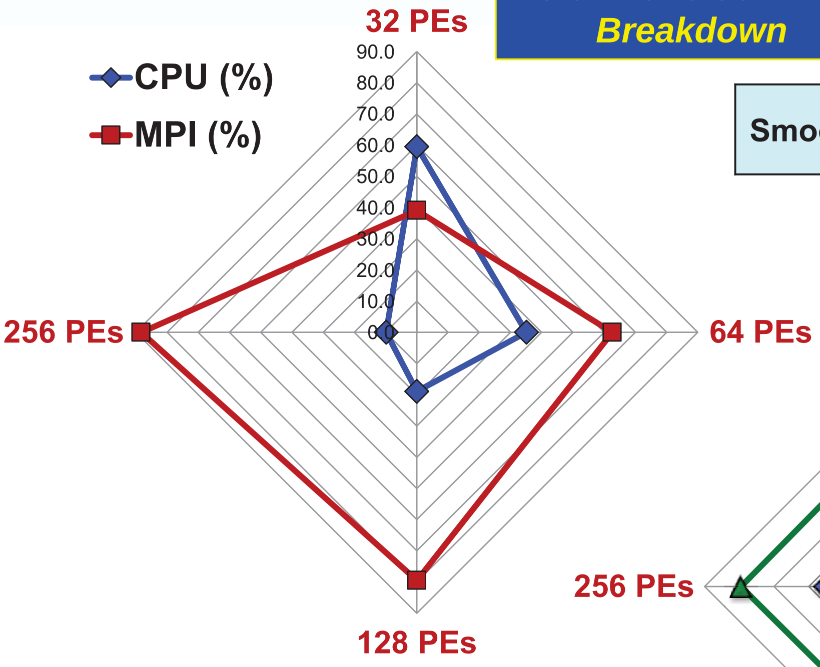
- ***A Report Summary:*** This characterizes how the application's wallclock time was spent, broken down into CPU, MPI and I/O
- All examples from the **Hawk Cluster (SKL Gold 6148 / 2.4GHz)**

DLPOLY4 – Performance Report

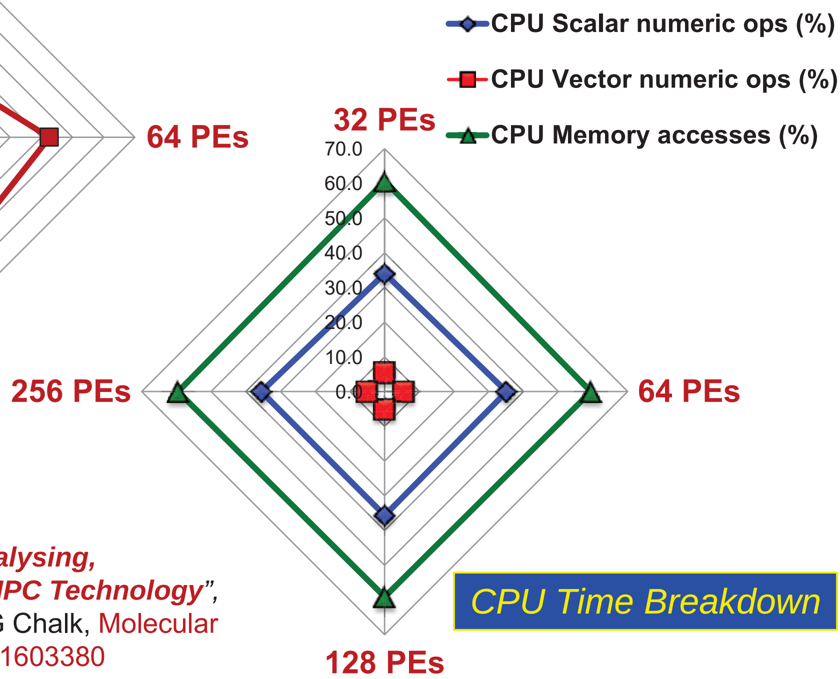
CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

Total Wallclock Time Breakdown

Performance Data (32-256 PEs)



Smooth Particle Mesh Ewald Scheme



CPU Time Breakdown

"DL_POLY - A Performance Overview. Analysing, Understanding and Exploiting available HPC Technology", Martyn F Guest, Alin M Elena and Aidan B G Chalk, Molecular Simulation, (2019) 10.1080/08927022.2019.1603380

EPYC - Compiler and Run-time Options

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

STREAM (AMD Daytona Cluster):

```
icc stream.c -DSTATIC -Ofast -march=core-avx2 -DSTREAM_ARRAY_SIZE=2500000000 -DNTIMES=10 -mcmodel=large -shared-intel -restrict -qopt-streaming-stores always -o streamc.Rome
icc stream.c -DSTATIC -Ofast -march=core-avx2 -qopenmp -DSTREAM_ARRAY_SIZE=2500000000 -DNTIMES=10 -mcmodel=large -shared-intel -restrict -qopt-streaming-stores always -o streamcp.Rome
```

```
# Preload the amd-cputype library to navigate
# the Intel Genuine cpu test
module use /opt/amd/modulefiles
module load AMD/amd-cputype/1.0
export LD_PRELOAD=$AMD_CPUTYPE_LIB

export OMP_DISPLAY_ENV=true
export OMP_PLACES="cores"
export OMP_PROC_BIND="spread"
export MKL_DEBUG_CPU_TYPE=5
```

STREAM (Dell|EMC EPYC):

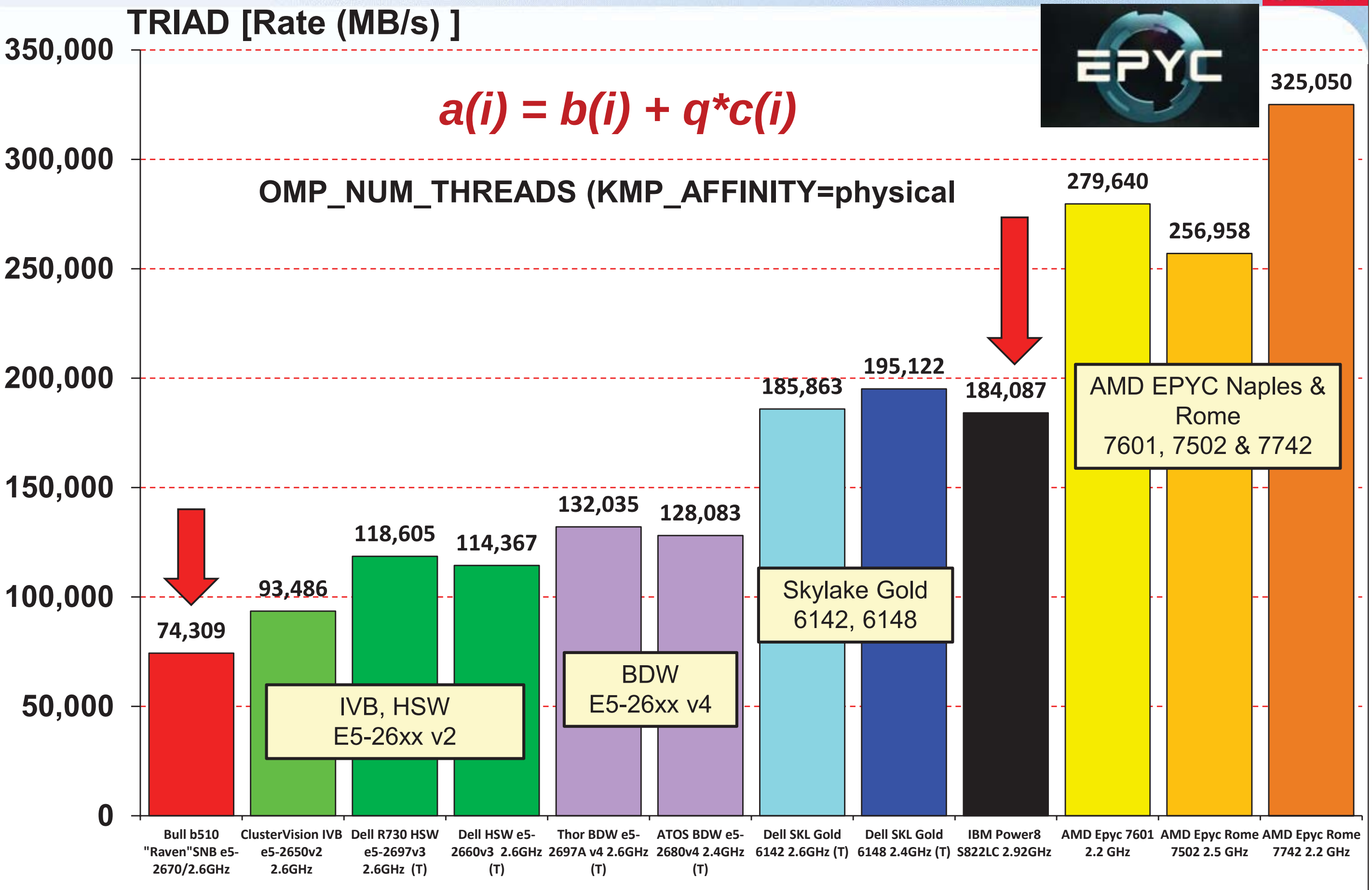
```
export OMP_SCHEDULE=static
export OMP_DYNAMIC=false
export OMP_THREAD_LIMIT=128
export OMP_NESTED=FALSE
export OMP_STACKSIZE=192M

for h in $(scontrol show hostnames); do
echo hostname: $h
# 64 cores
ssh $h "OMP_NUM_THREADS=64
GOMP_CPU_AFFINITY=0-63
OMP_DISPLAY_ENV=true $code
```

Compilation:

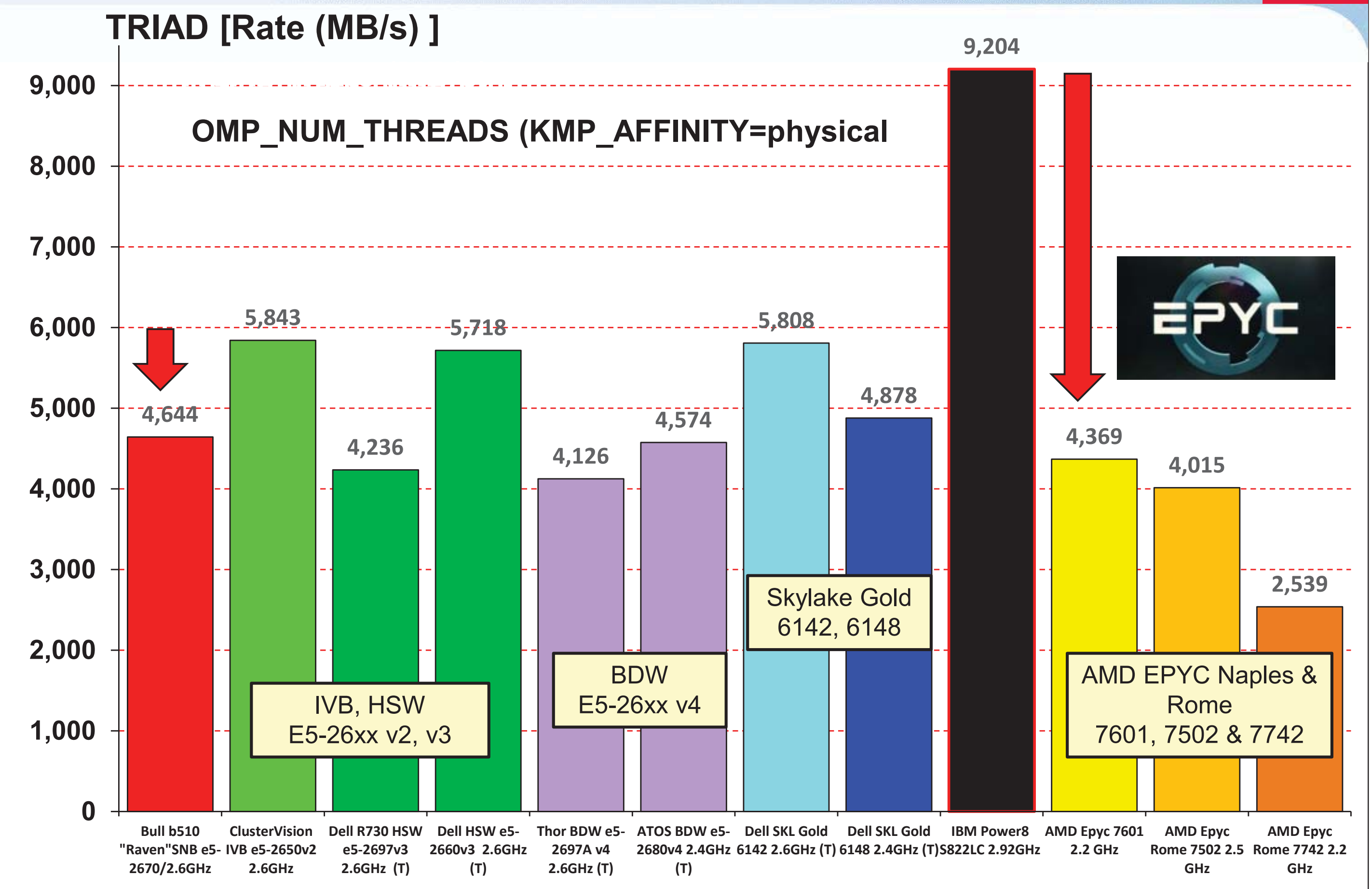
INTEL COMPILERS 2018u4, IntelMPI 2017 Update 5, FFTW-3.3.5

INTEL SKL: –O3 –xCORE-AVX512
AMD EPYC: –O3 –march=core-avx2 -align array64byte -fma -ftz -fomit-frame-pointer

Memory B/W – STREAM performance



Memory B/W – STREAM / core performance

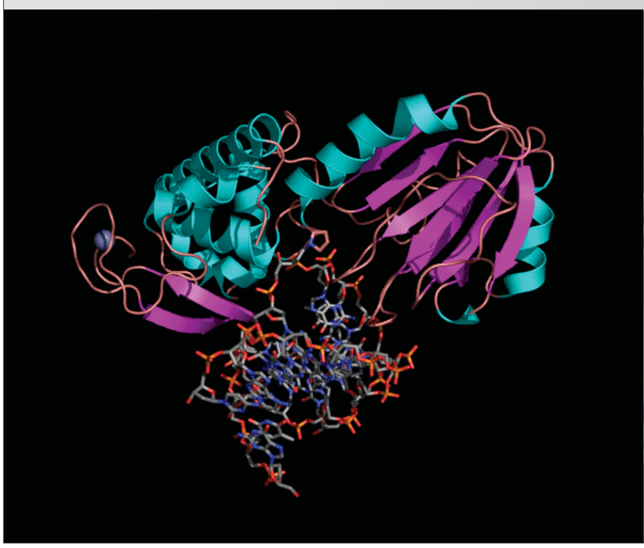


Performance Metrics – "Core to Core" & "Node to Node"

- Analysis of performance Metrics across a variety of data sets
  - "**Core to core**" and "**node to node**" workload comparisons
    - ***Core to core*** comparison i.e. performance for jobs with a fixed number of cores
    - ***Node to Node*** comparison typical of the performance when running a workload (real life production). Expected to reveal the major benefits of increasing core count per socket
  - Focus on two distinct "**node to node**" comparisons of the following:

1	***Hawk - Dell |EMC Skylake Gold 6148 2.4GHz (T) EDR with 40 cores / node***	***AMD EPYC 7452 nodes with 64 cores per node. [1-7 nodes]***
2	***Hawk - Dell |EMC Skylake Gold 6148 2.4GHz (T) EDR with 40 cores / node***	***AMD EPYC 7502 nodes with 64 cores per node. [1-7 nodes]***

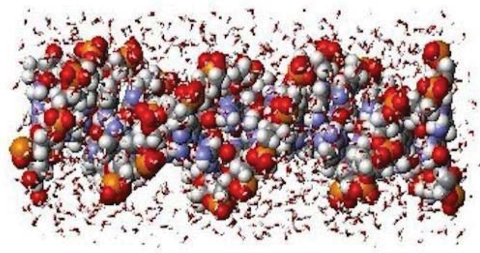
Performance Analysis of the AMD EPYC Rome Processors



Molecular Simulation; DL_POLY (Classic & DL_POLY 4), LAMMPS, NAMD, Gromacs

Molecular Simulation I. DL_POLY

Molecular Dynamics Codes: AMBER, DL_POLY, CHARMM, NAMD, LAMMPS, GROMACS etc



DL_POLY

- Developed as CCP5 parallel MD code by W. Smith, T.R. Forester and I. Todorov
  - UK CCP5 + International user community
  - DLPOLY_classic (replicated data) and DLPOLY_3 & _4 (distributed data – domain decomposition)
- Areas of application:
  - liquids, solutions, spectroscopy, ionic solids, molecular crystals, polymers, glasses, membranes, proteins, metals, solid and liquid interfaces, catalysis, clathrates, liquid crystals, biopolymers, polymer electrolytes.

The DLPOLY Benchmarks

DL_POLY Classic

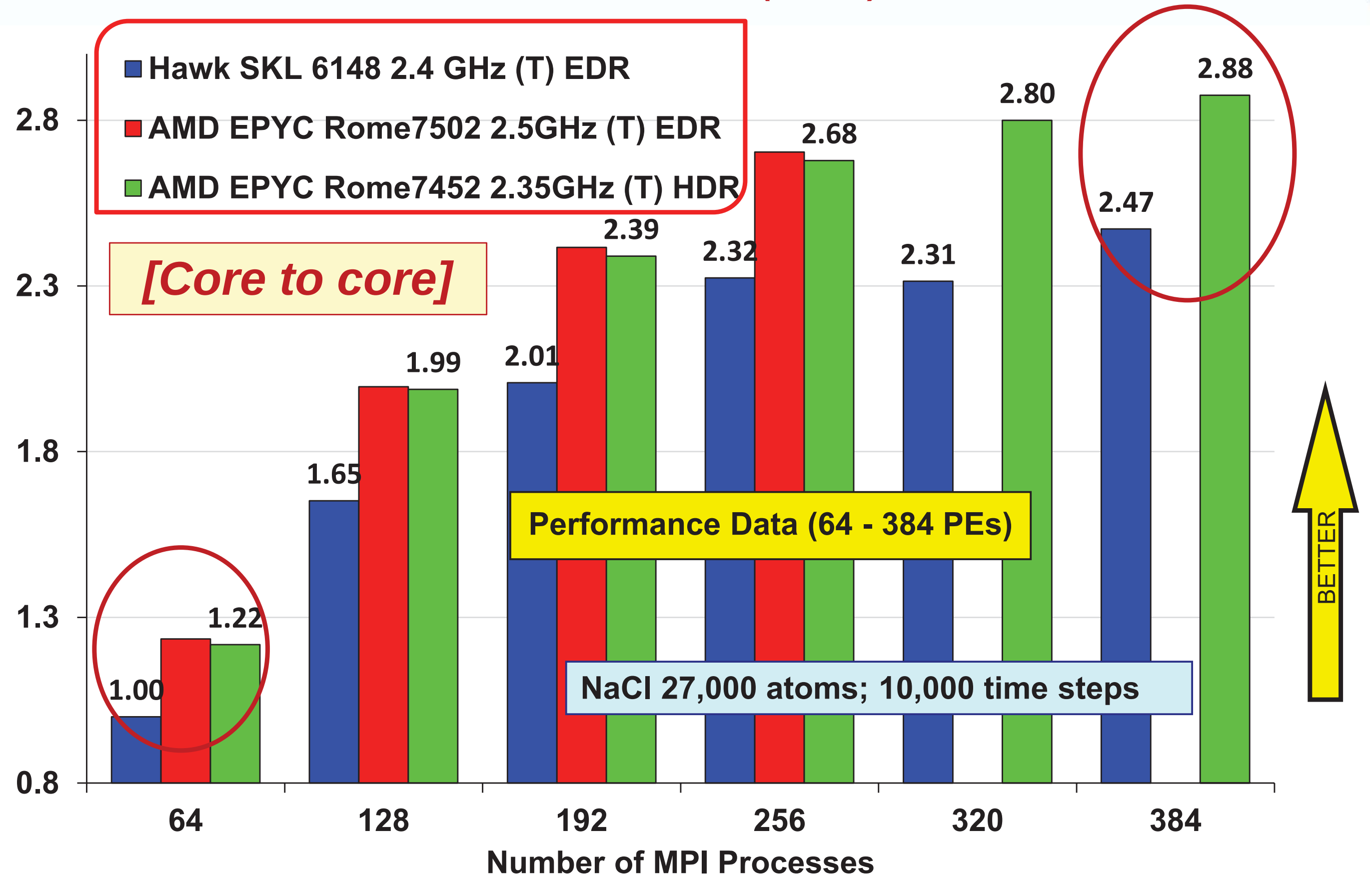
- **Bench4**
  - NaCl Melt Simulation with Ewald sum electrostatics & a MTS algorithm. 27,000 atoms; 10,000 time steps.
- **Bench5**
  - Potassium disilicate glass (with 3-body forces). 8,640 atoms: 60,000 time steps
- **Bench7**
  - *Simulation of gramicidin A molecule in 4012 water molecules using neutral group electrostatics. 12,390 atoms: 100,000 time steps*

DL_POLY 4

- **Test2 Benchmark**
  - NaCl Simulation; 216,000 ions, 200 time steps, Cutoff=12Å
- **Test8 Benchmark**
  - *Gramicidin in water; rigid bonds + SHAKE: 792,960 ions, 50 time steps*

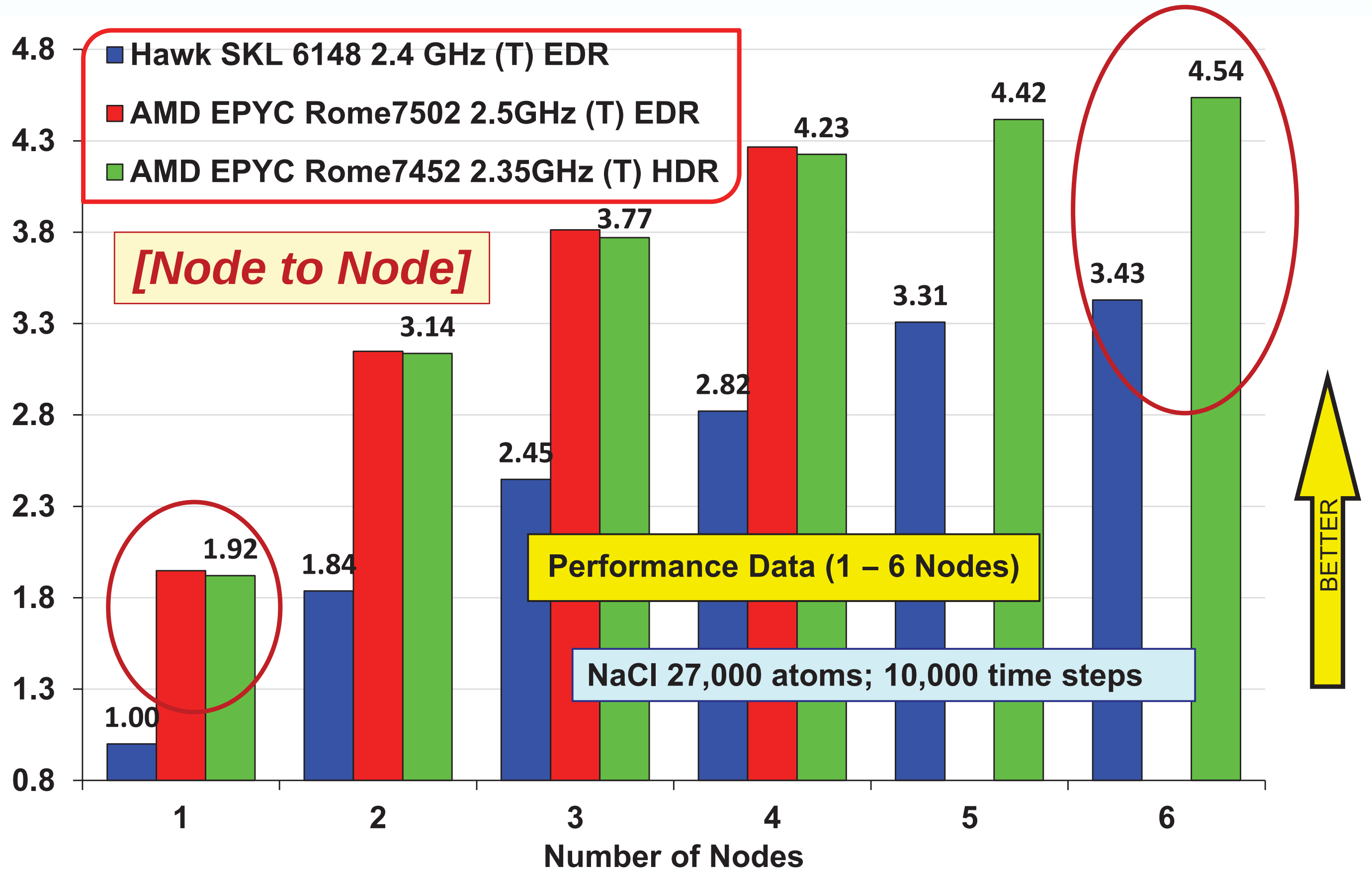
DL_POLY Classic – NaCl Simulation

Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*



DL_POLY Classic – NaCl Simulation

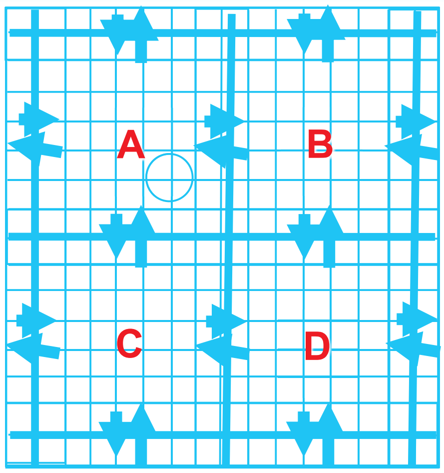
Performance *Relative to the Hawk SKL 6148 2.4 GHz (1 node)*



DL_POLY 4 – Distributed data

Domain Decomposition - Distributed data:

- Distribute atoms, forces across the nodes
  - More memory efficient, can address much larger cases (10^5-10^7)
- Shake and short-ranges forces require only neighbour communication
  - communications scale linearly with number of nodes
- Coulombic energy remains global
  - Adopt **Smooth Particle Mesh Ewald** scheme
    - includes Fourier transform smoothed charge density (reciprocal space grid typically 64x64x64 - 128x128x128)



W. Smith and I. Todorov

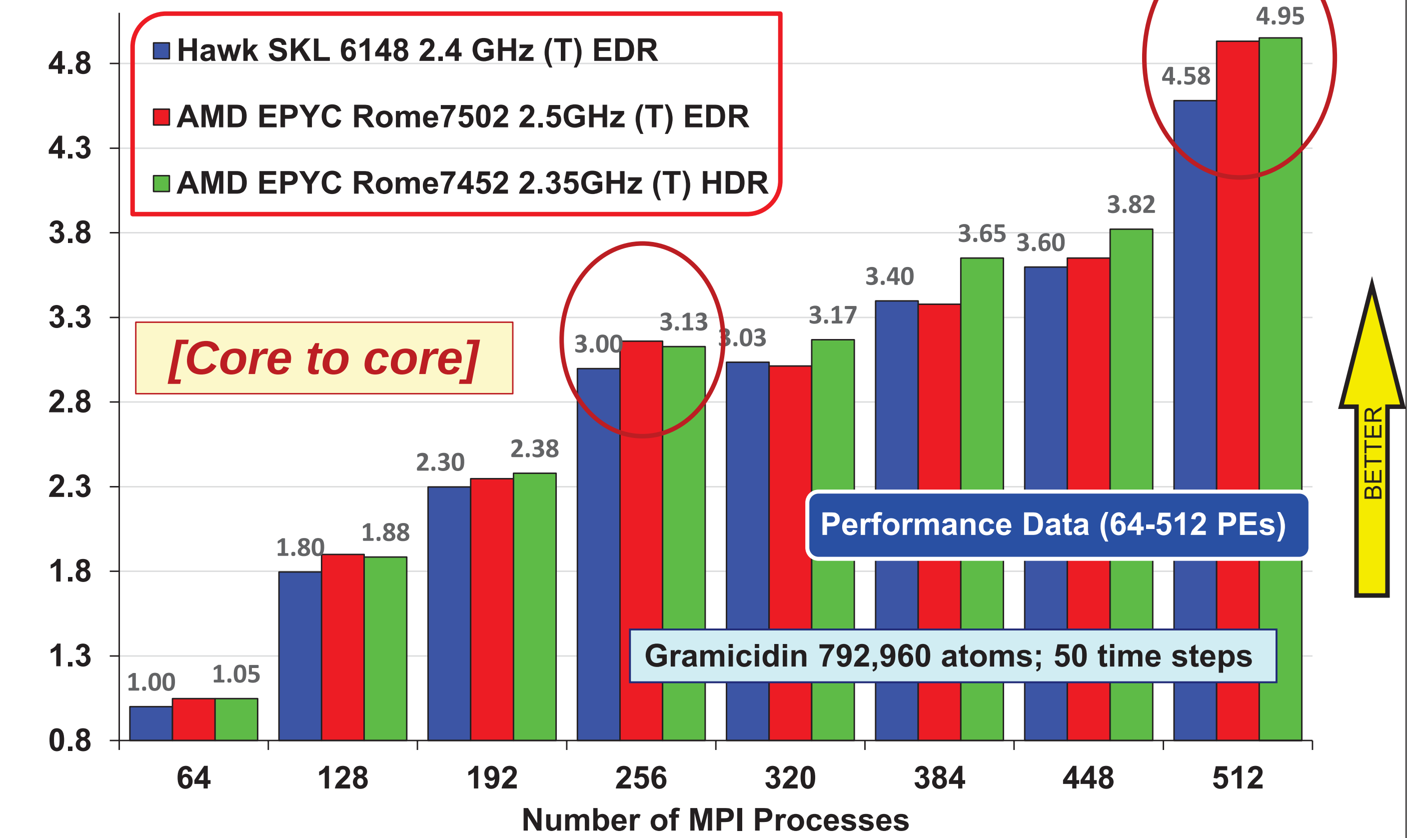
Benchmarks

1. **NaCl Simulation; 216,000 ions, 200 time steps, Cutoff=12Å**
2. **Gramicidin in water; rigid bonds + SHAKE: 792,960 ions, 50 time steps**

http://www.scd.stfc.ac.uk//research/app/ccg/software/DL_POLY/44516.aspx

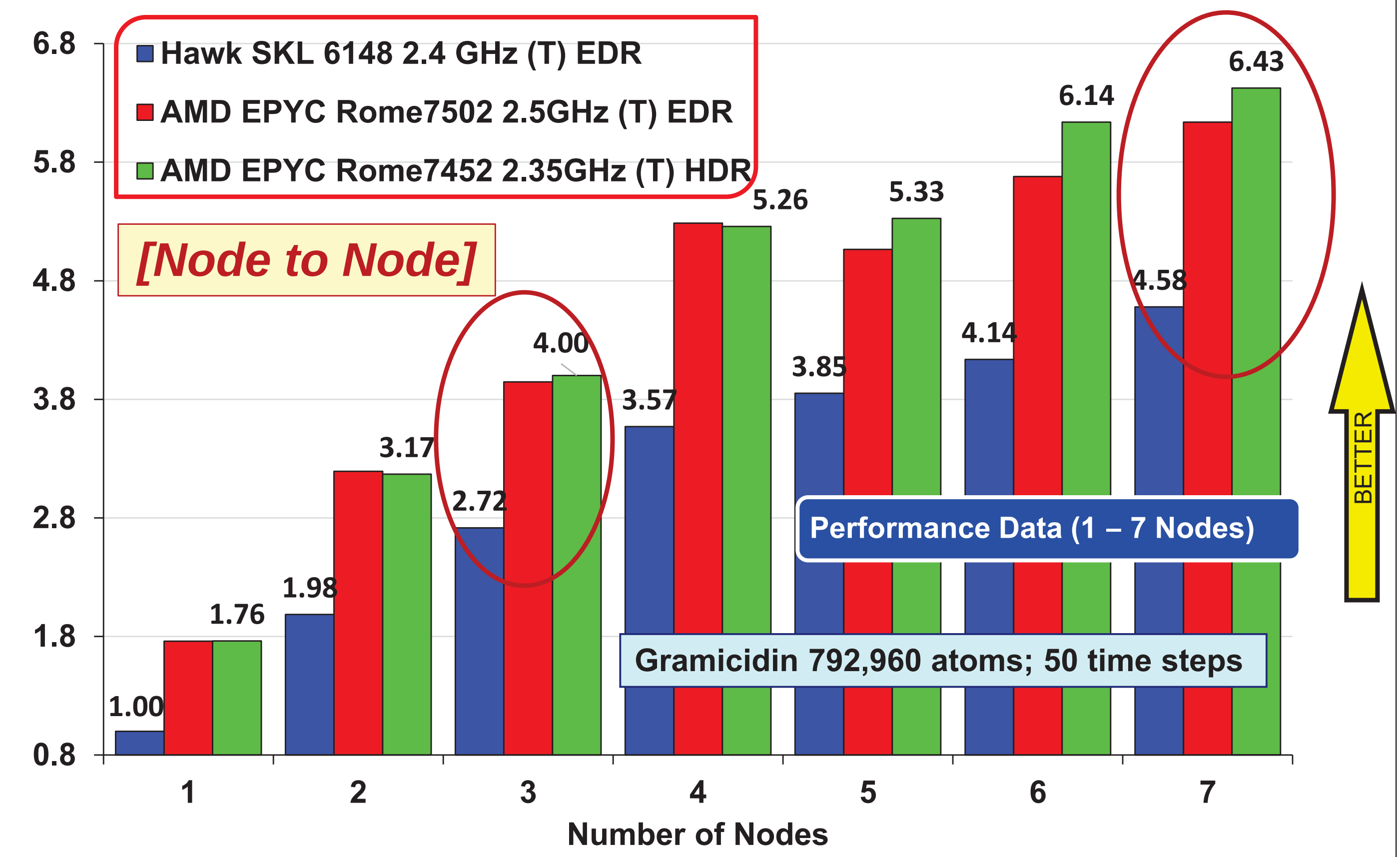
DL_POLY 4 – Gramicidin Simulation

Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*

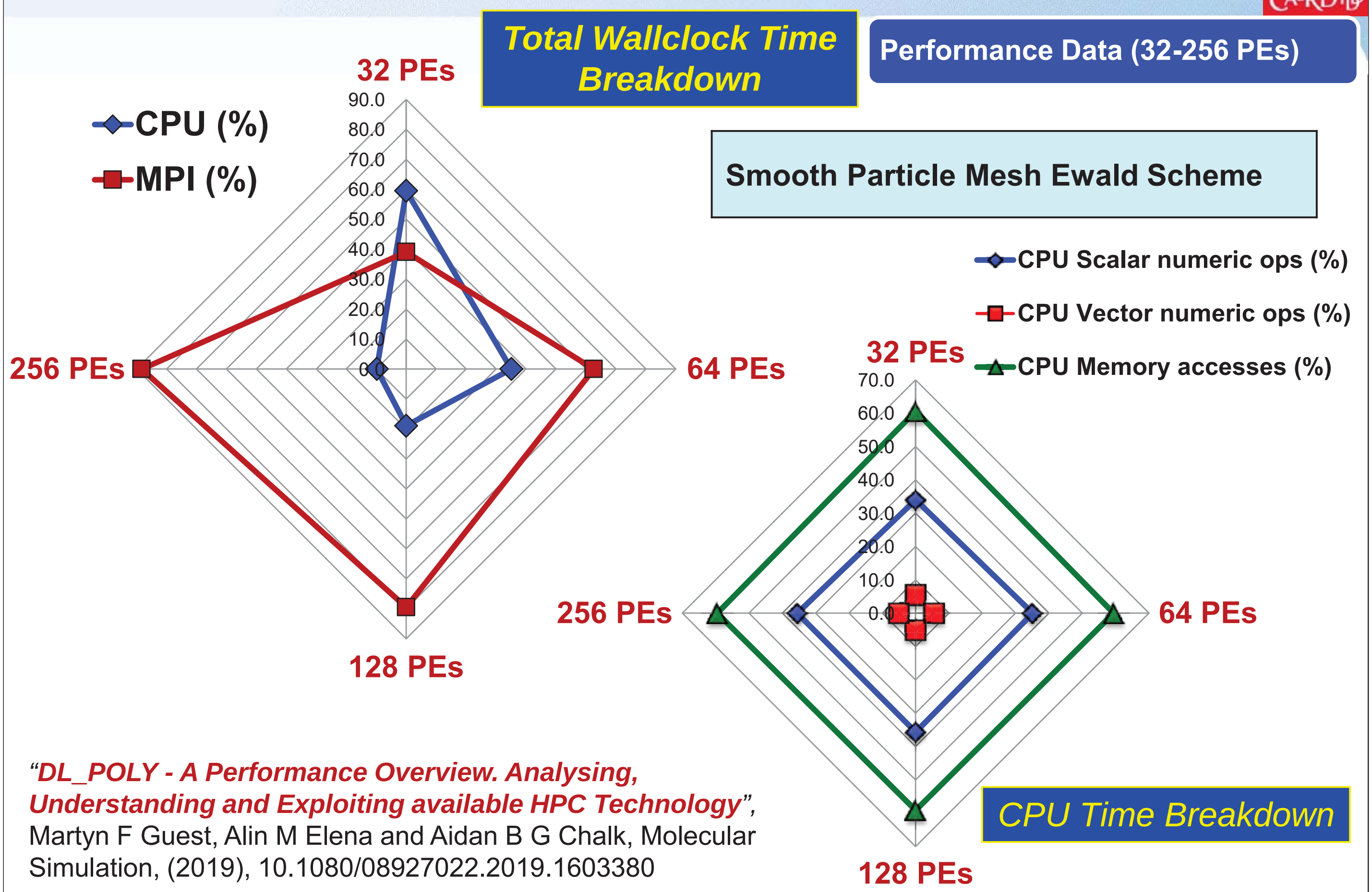


DL_POLY 4 – Gramicidin Simulation

Performance *Relative to the Hawk SKL 6148 2.4 GHz (1 Node)*



DLPOLY4 – Gramicidin Simulation Performance Report



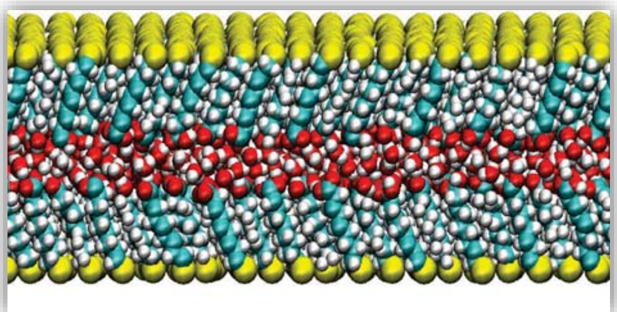
"DL_POLY - A Performance Overview. Analysing, Understanding and Exploiting available HPC Technology", Martyn F Guest, Alin M Elena and Aidan B G Chalk, Molecular Simulation, (2019), 10.1080/08927022.2019.1603380

Molecular Simulation - II. *LAMMPS*

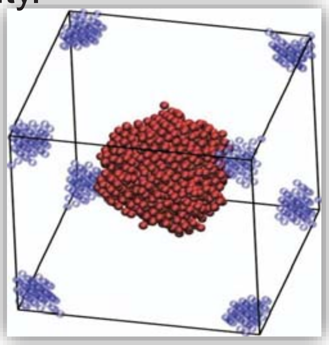
Archer Rank: 9

http://lammps.sandia.gov/index.html

- LAMMPS is a **classical molecular dynamics code**, and an acronym for Large-scale Atomic/Molecular Massively Parallel Simulator. (**LAMMPS (12 Dec 2018)** used in this study)
- LAMMPS has potentials for soft materials (biomolecules, polymers) and solid-state materials (metals, semiconductors) and coarse-grained or mesoscopic systems. It can be used to model atoms or, more generically, as a parallel particle simulator at the atomic, meso, or continuum scale.
- LAMMPS runs on single processors or in parallel using message-passing techniques and a spatial-decomposition of the simulation domain. The code is designed to be easy to modify or extend with new functionality.



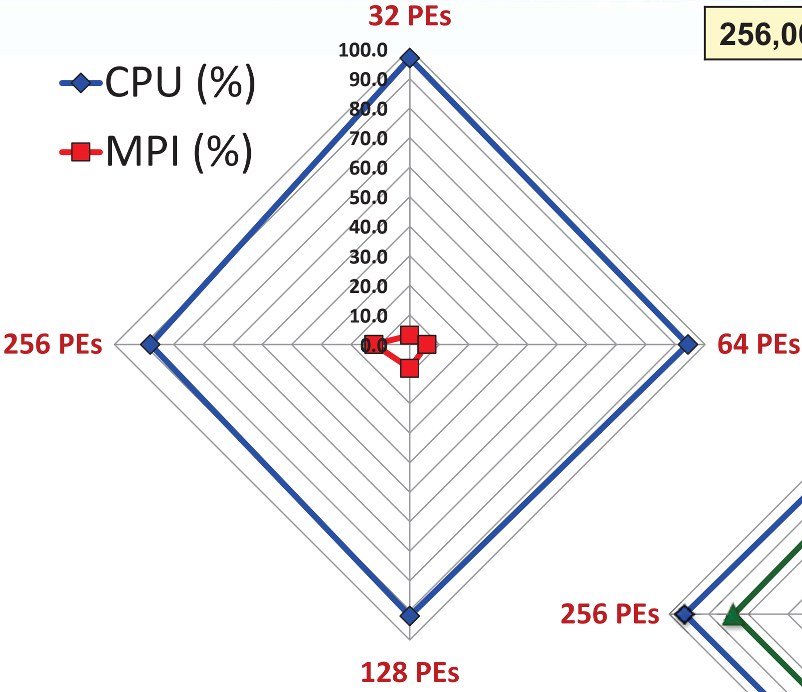
S. Plimpton, ***Fast Parallel Algorithms for Short-Range Molecular Dynamics***, J Comp Phys, 117, 1-19 (1995).



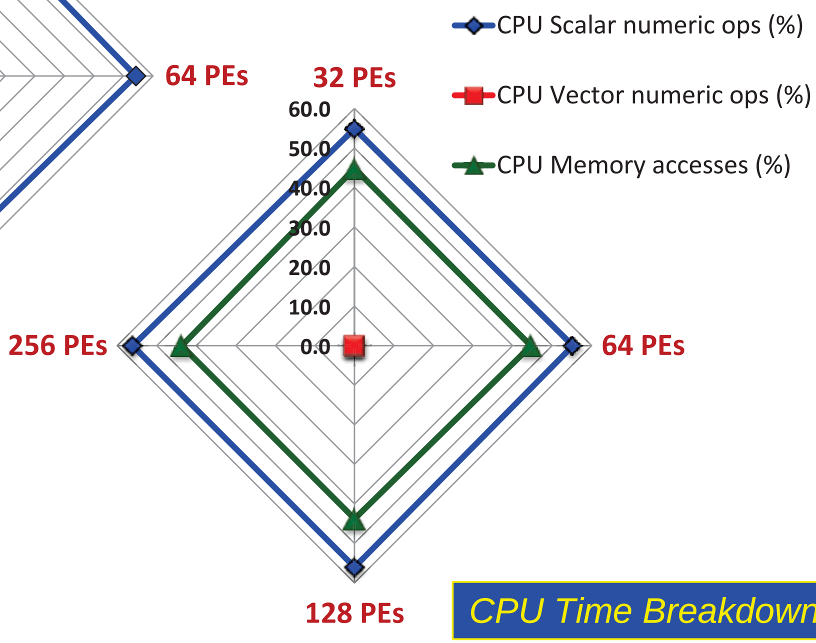
LAMMPS –Lennard-Jones Fluid - Performance Report

256,000 atoms; 5,000 time steps

Performance Data (32-256 PEs)

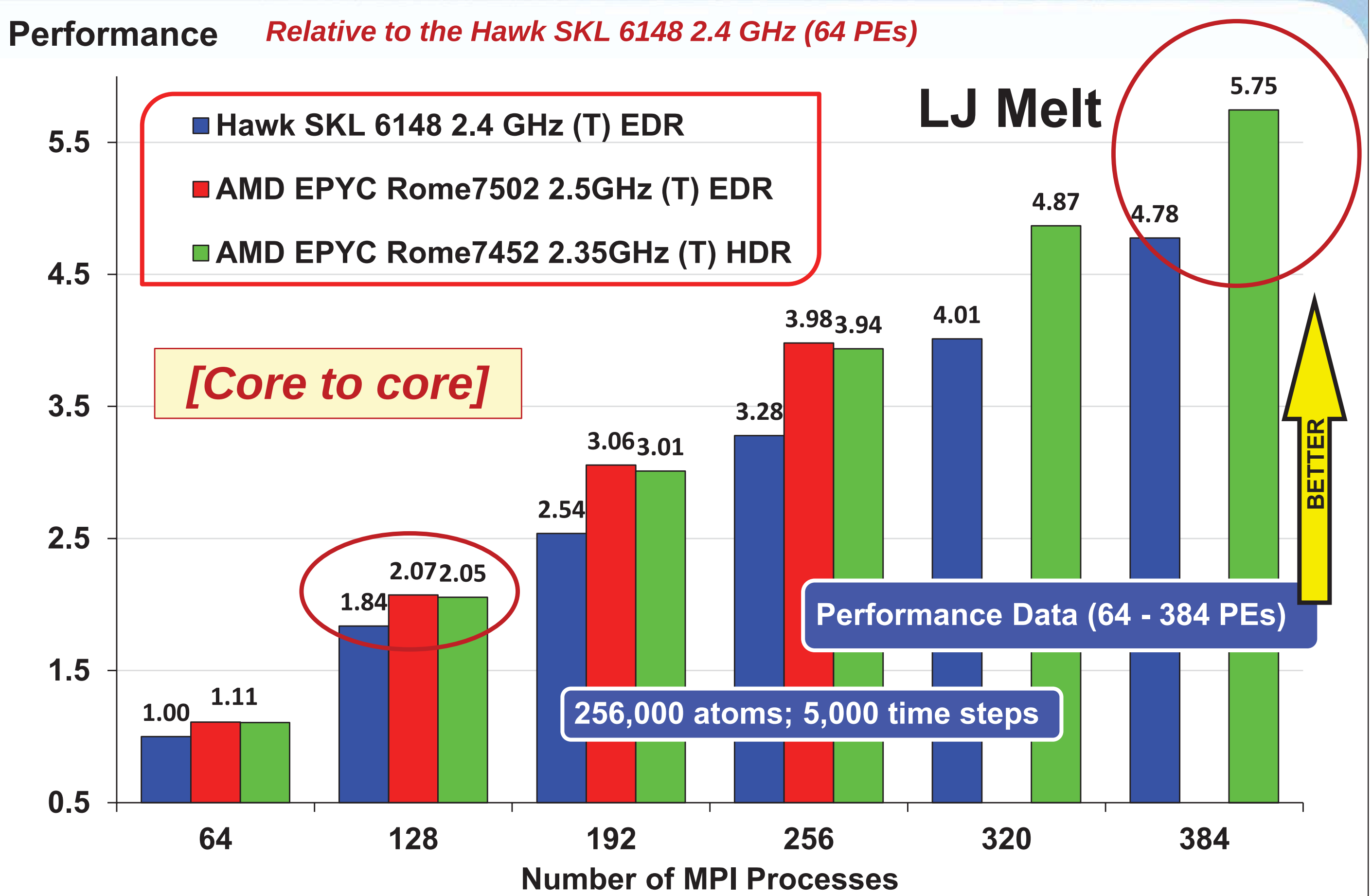


Total Wallclock Time Breakdown

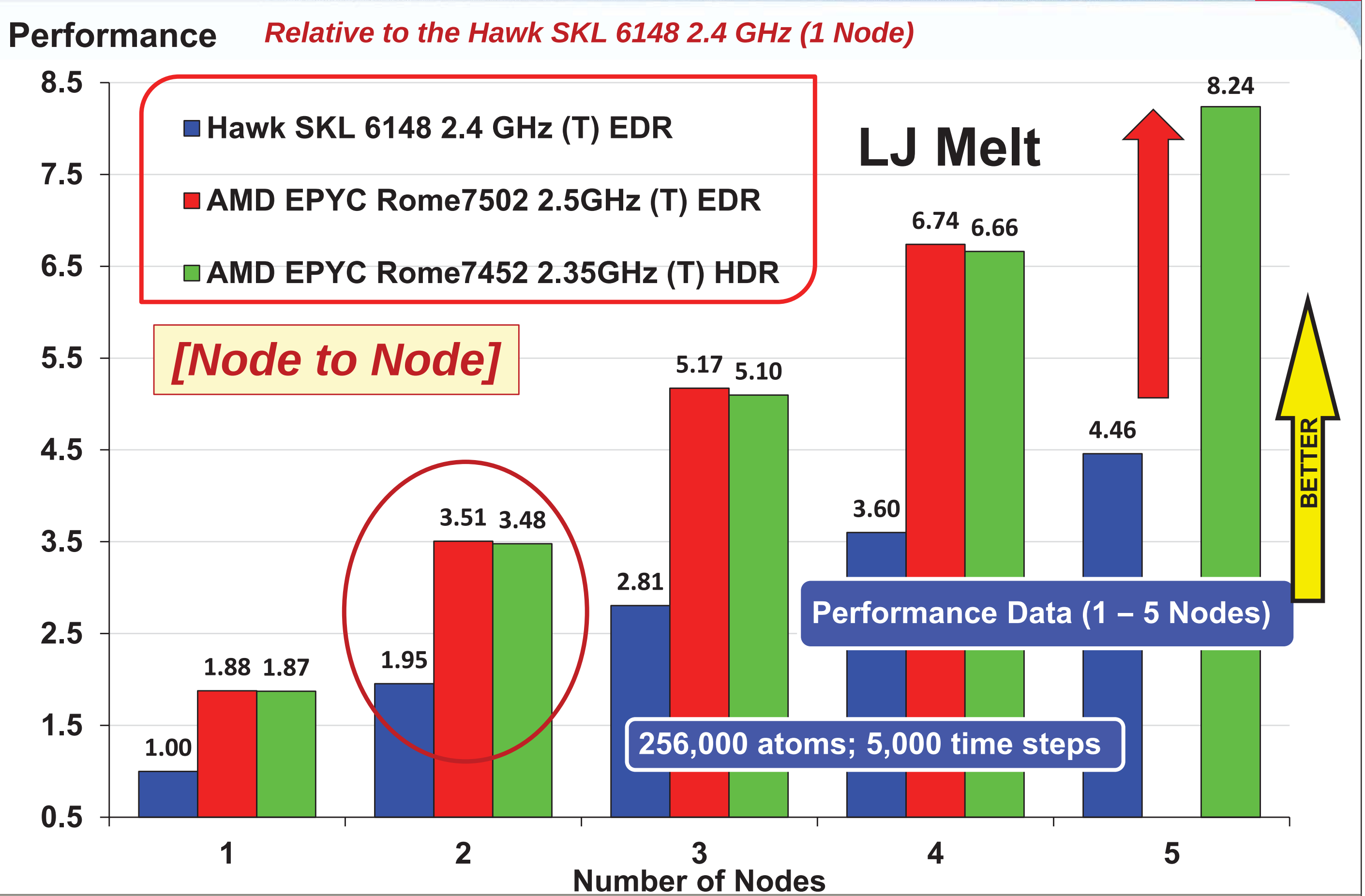


CPU Time Breakdown

LAMMPS – Atomic fluid with Lennard-Jones Potential



LAMMPS – Atomic fluid with Lennard-Jones Potential



Molecular Simulation - III. NAMD

NAMD Scalable Molecular Dynamics

Archer Rank: 21 **http://www.ks.uiuc.edu/Research/namd/**

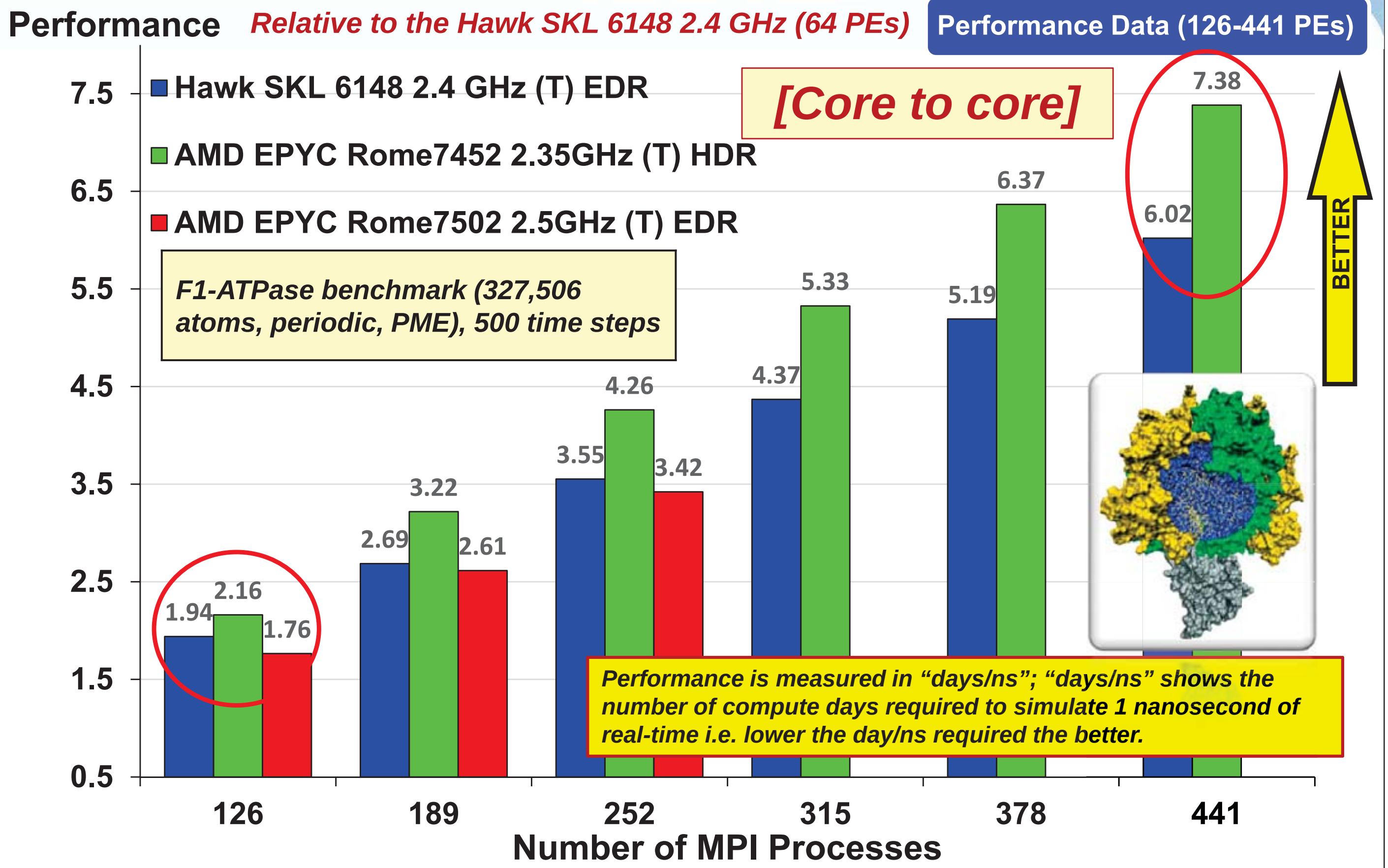
- NAMD, is a parallel molecular dynamics code designed for high-performance simulation of large bio-molecular systems. Based on Charm++ parallel objects, NAMD scales to hundreds of cores for typical simulations and beyond 500,000 cores for the largest simulations.
- NAMD uses the popular molecular graphics program VMD for simulation setup and trajectory analysis, but is also file-compatible with AMBER, CHARMM, and X-PLOR. NAMD distributed free of charge with source code.
- Using **NAMD 2.13** in this work.
- Benchmark cases – apoA1 (apolipoprotein A-I), **F1-ATPase and stmv**

VMD Visual Molecular Dynamics

VMD is a molecular visualization program for displaying, animating, and analyzing large biomolecular systems. VMD supports computers running MacOS X, Unix, or Windows.

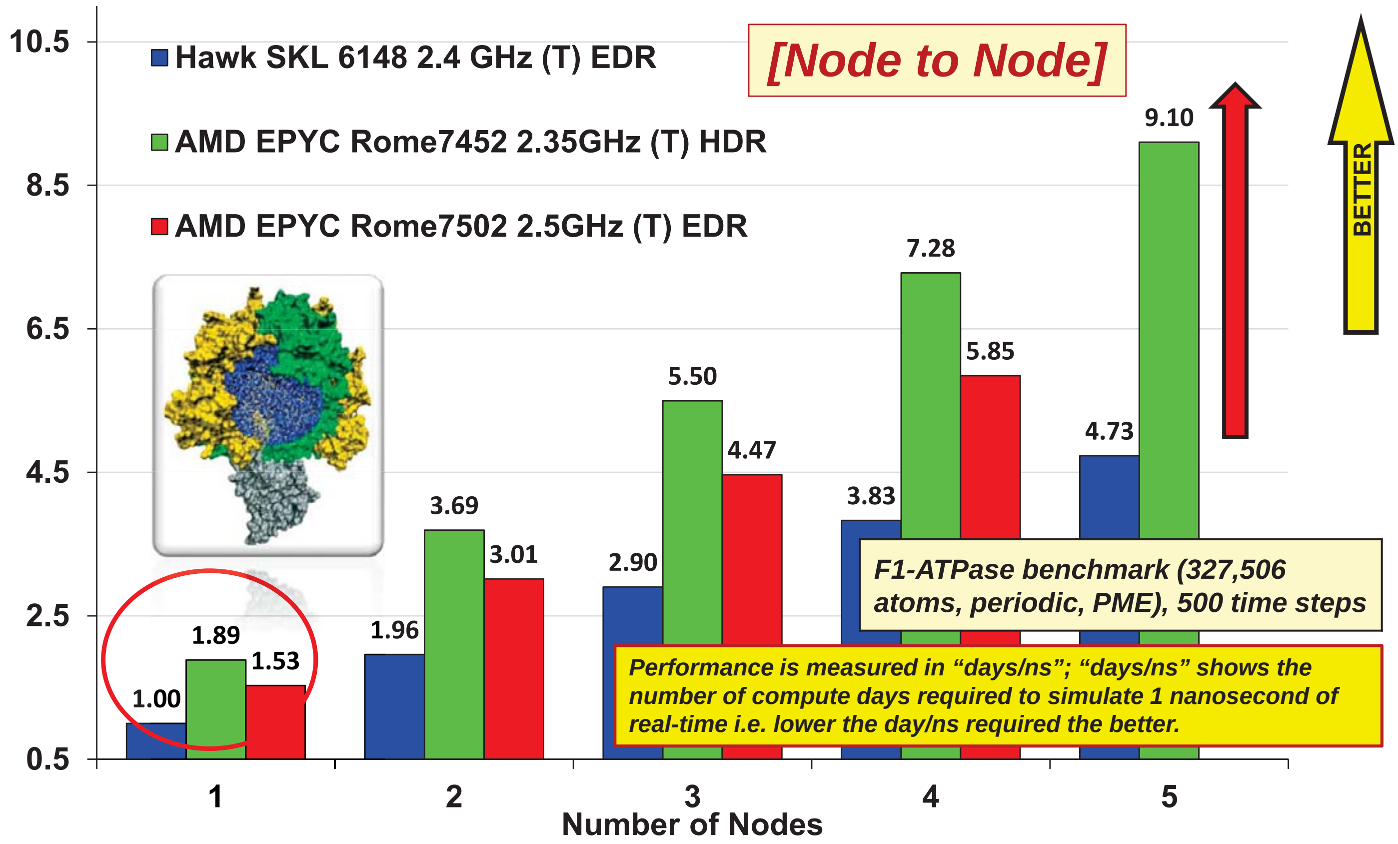
1. James C. Phillips et al., ***Scalable molecular dynamics with NAMD*** , J Comp Chem, 26, 1781-1792 (2005).
2. B. Acun, D. J. Hardy, L. V. Kale, K. Li, J. C. Phillips, & J. E. Stone. **Scalable Molecular Dynamics with NAMD on the Summit System**. *IBM Journal of Research and Development*, 2018.

NAMD – F1-ATPase Benchmark – days/ns



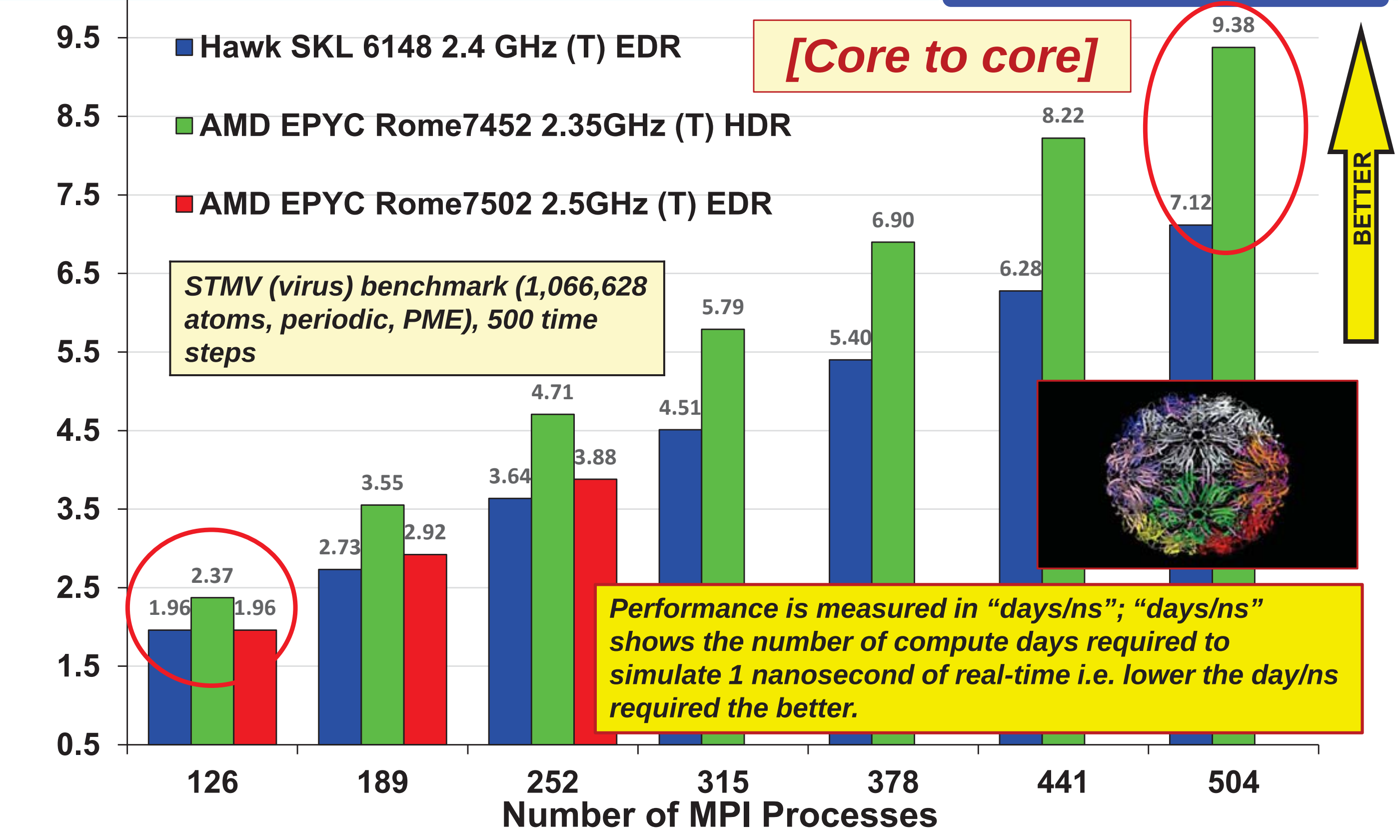
NAMD – F1-ATPase Benchmark – days/ns

Performance *Relative to the Hawk SKL 6148 2.4 GHz (1 Node)* **Performance Data (1 – 5 Nodes)**



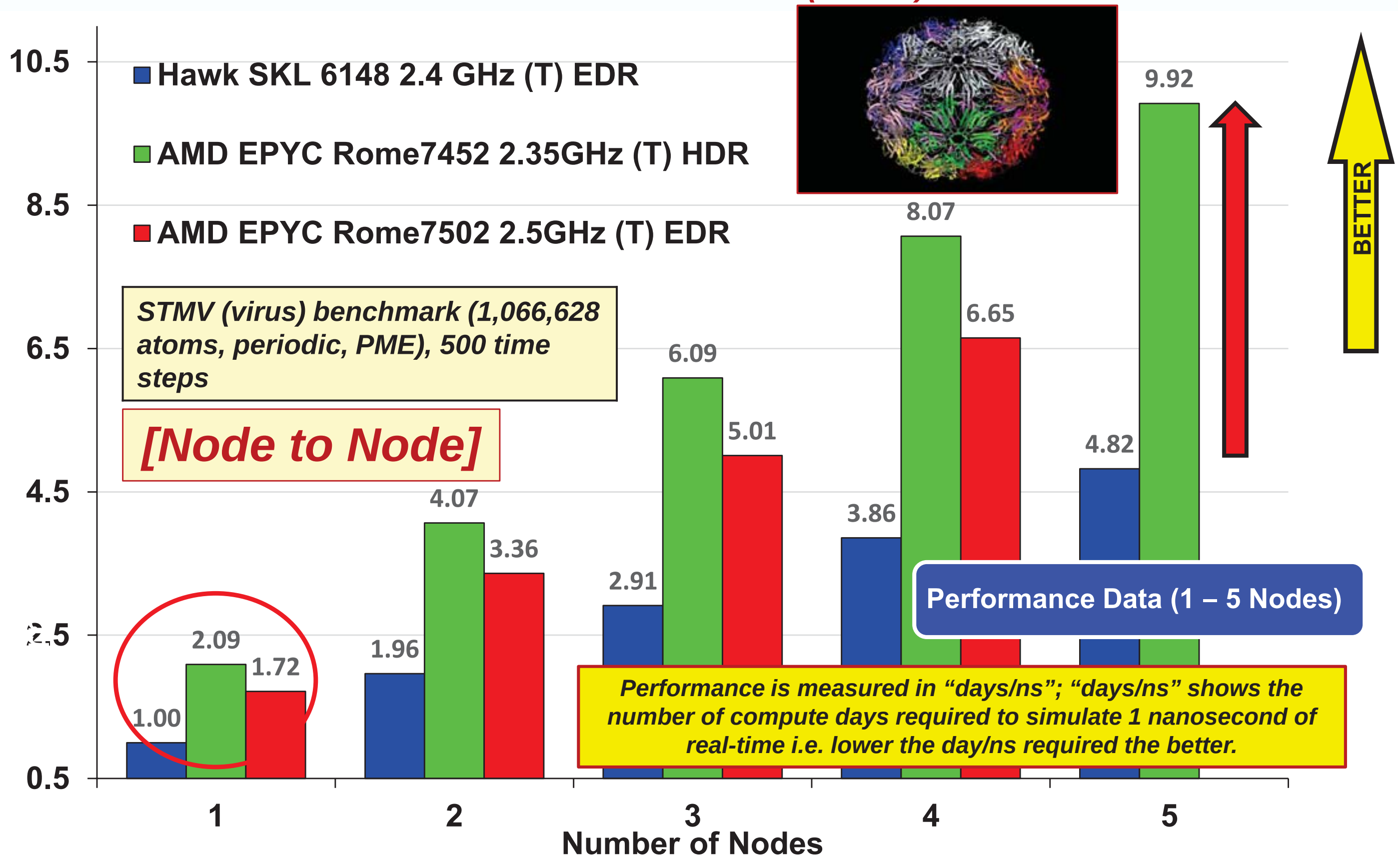
NAMD – STMV (virus) Benchmark – days/ns

Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)* **Performance Data (128-512 PEs)**



NAMD – STMV (virus) Benchmark – days/ns

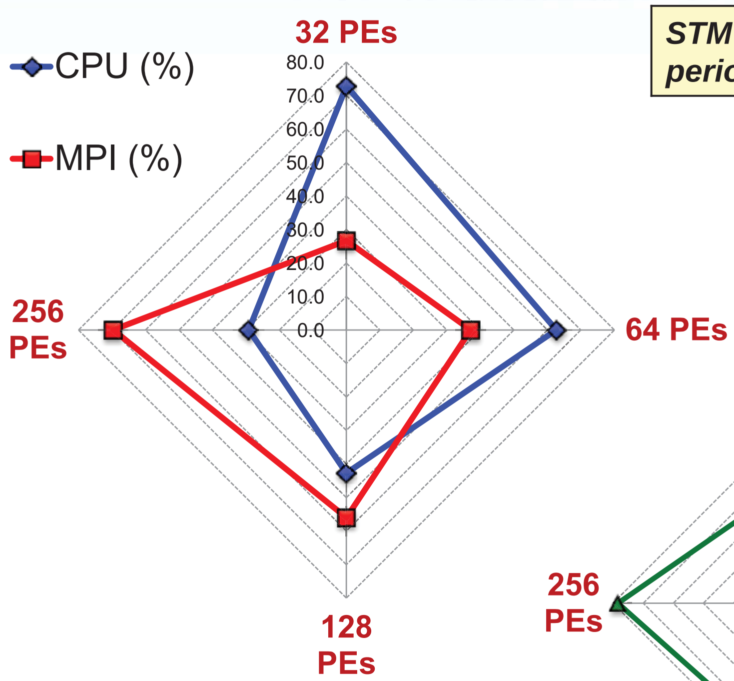
Performance *Relative to the Hawk SKL 6148 2.4 GHz (1 Node)*



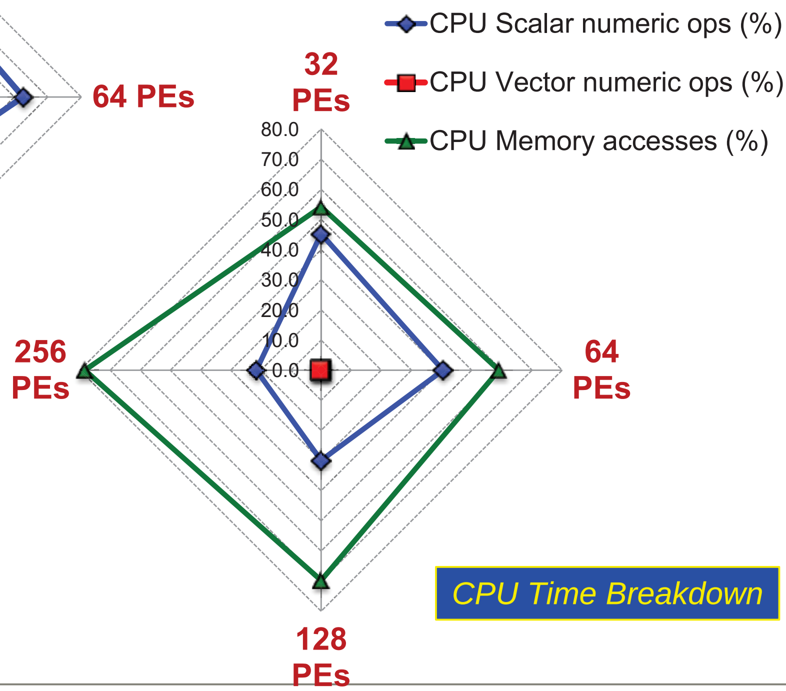
NAMD – STMV (virus) Performance Report

STMV (virus) benchmark (1,066,628 atoms, periodic, PME), 500 time steps

Performance Data (32-256 PEs)



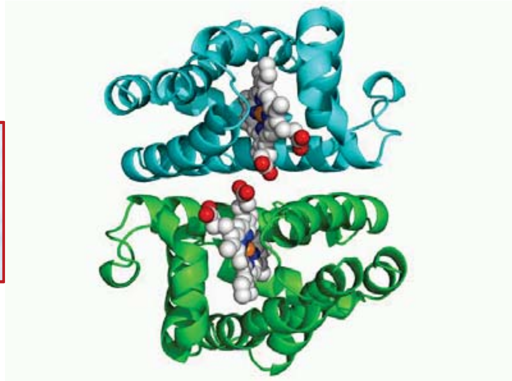
Total Wallclock Time Breakdown



CPU Time Breakdown

Molecular Simulation - IV. GROMACS *Archer Rank: 7*

GROMACS (GROningen MAchine for Chemical Simulations) is a molecular dynamics package designed for simulations of proteins, lipids and nucleic acids [University of Groningen]



Versions under Test:

Version 4.6.1 – 5 March 2013

Version 5.0.7 – 14 October 2015

Version 2016.3 – 14 March 2017

Version 2018.2 – 14 June 2018 (optimised for Hawk by Ade Fewings)

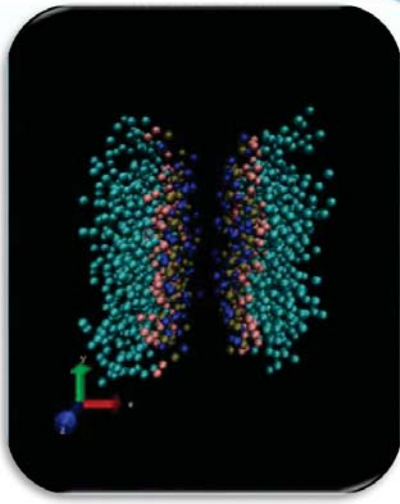
- **Berk Hess et al. "*GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*". *Journal of Chemical Theory and Computation* 4 (3): 435–447.**

http://manual.gromacs.org/documentation/

GROMACS Benchmark Cases

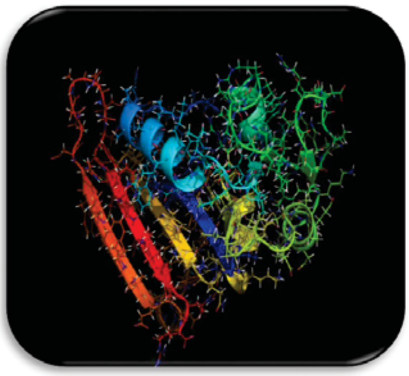
Ion channel system

- The 142k particle ion channel system is the membrane protein GluCl - a pentameric chloride channel embedded in a DOPC membrane and solvated in TIP3P water, using the Amber ff99SB-ILDN force field. This system is a **challenging** parallelization case due to the small size, but is one of the **most wanted target sizes** for biomolecular simulations

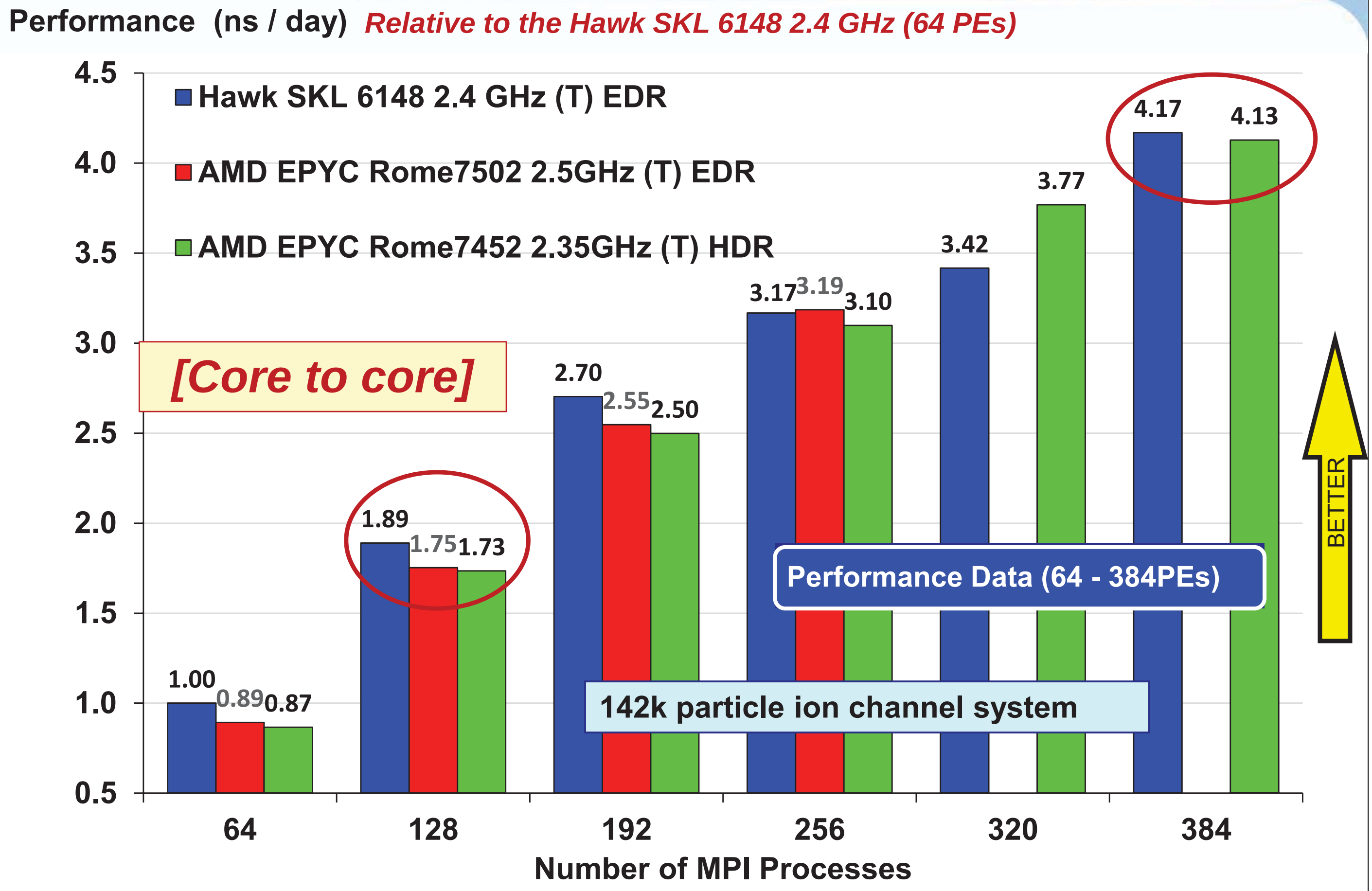


Lignocellulose

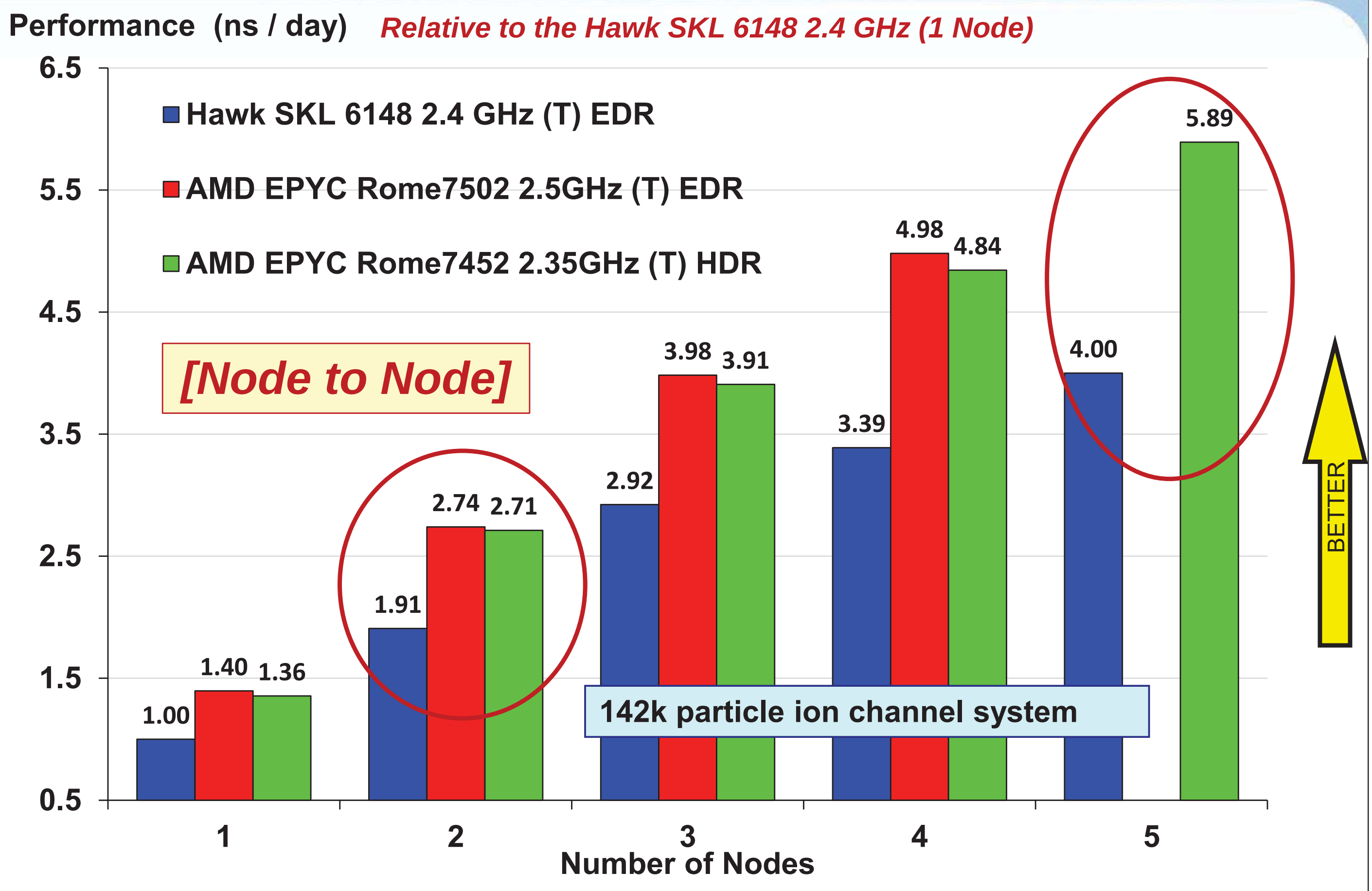
- Gromacs Test Case B from the UEA Benchmark Suite. A model of cellulose and lignocellulosic biomass in an aqueous solution. This system of 3.3M atoms is inhomogeneous, and uses **reaction-field electrostatics** instead of PME and therefore should scale well.



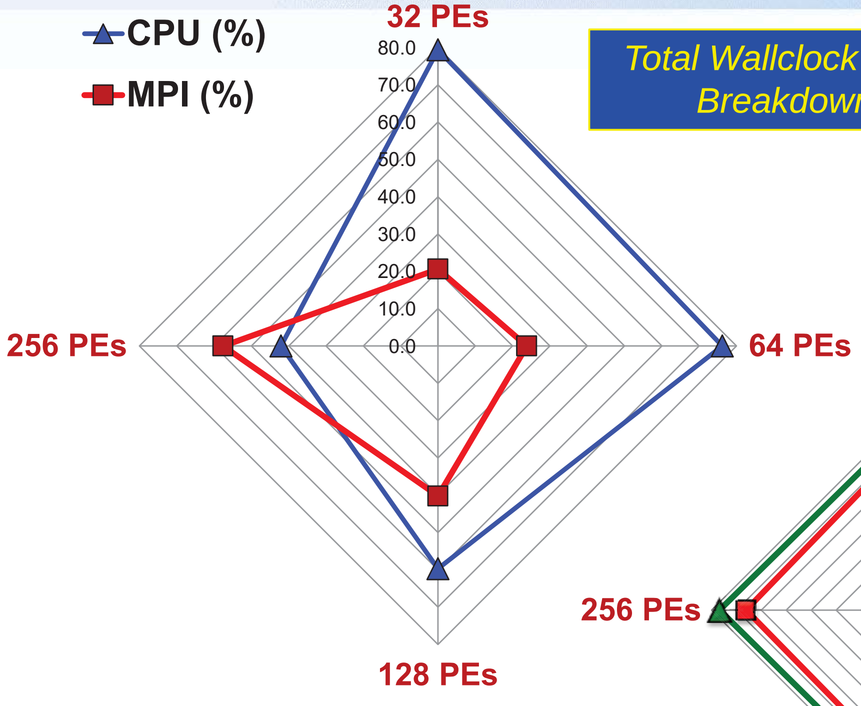
GROMACS – Ion Channel Simulation



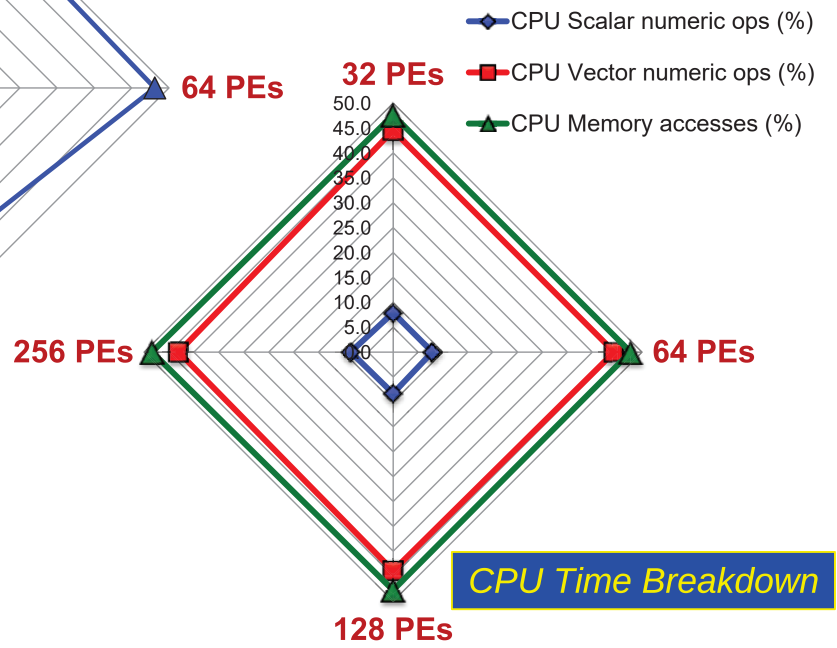
GROMACS – Ion Channel Simulation



GROMACS – Ion-channel Performance Report



Total Wallclock Time Breakdown



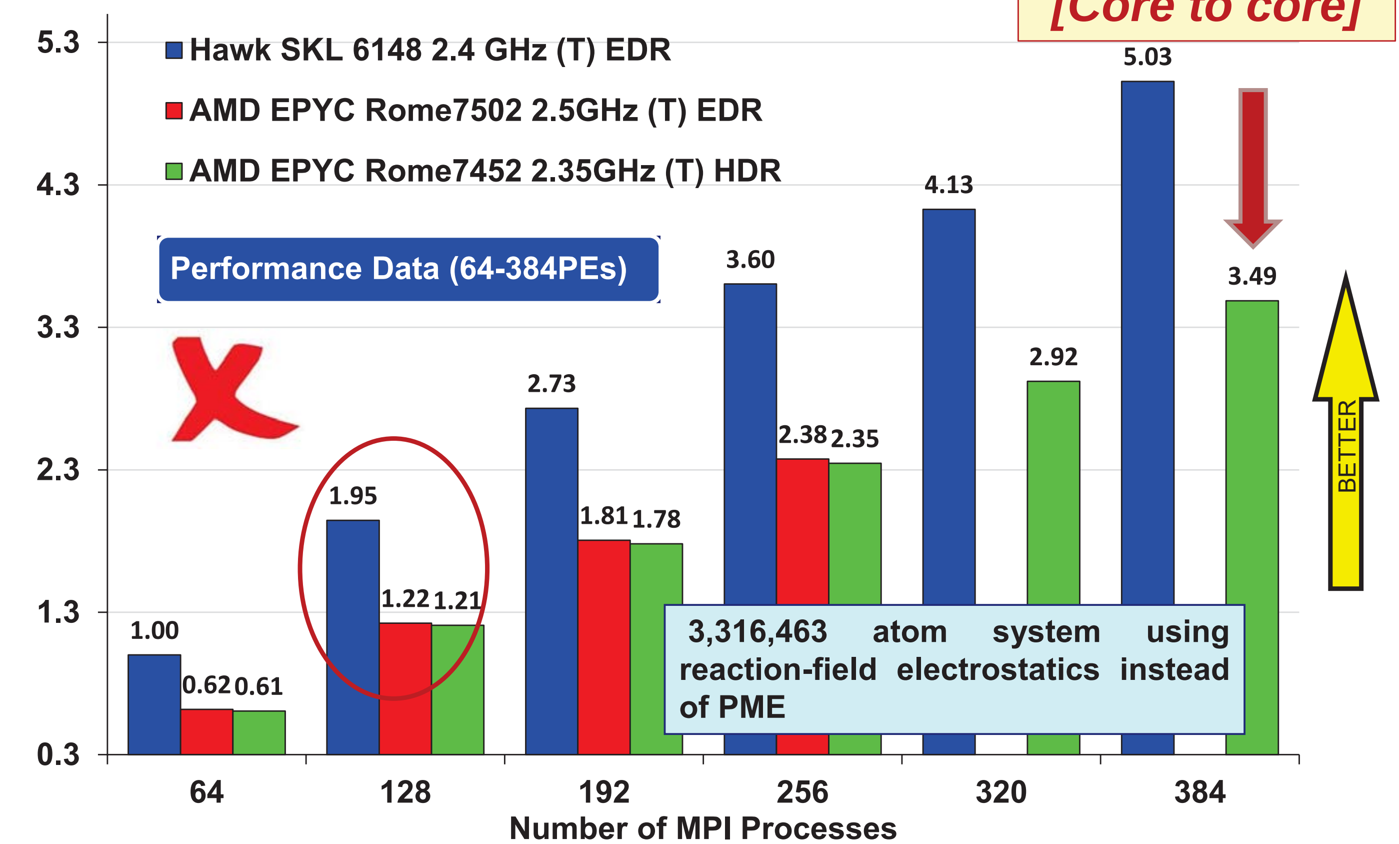
Performance Data (32-256 PEs)

CPU Time Breakdown

GROMACS – Lignocellulose Simulation

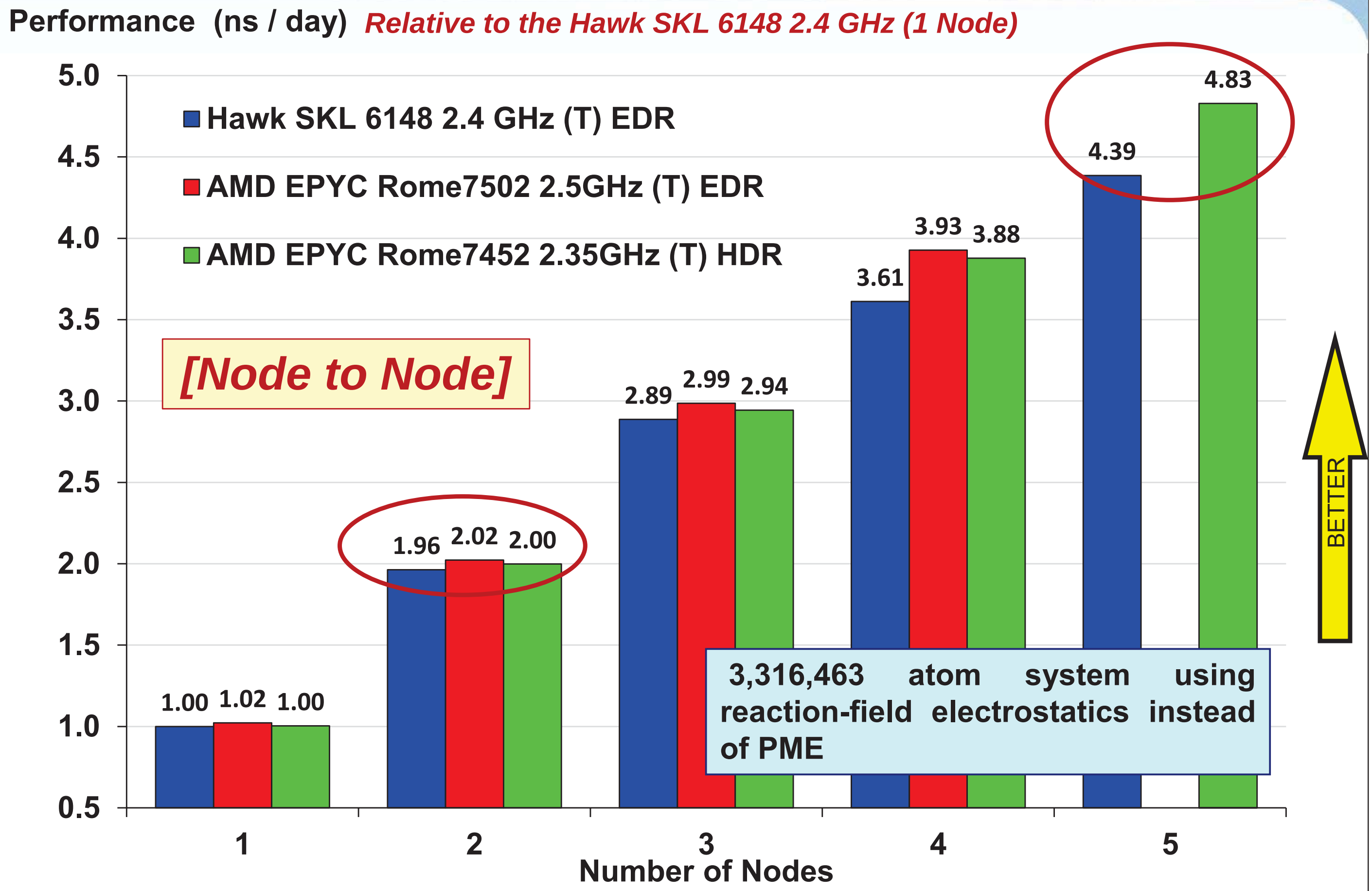
Performance (ns / day) *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*

[Core to core]

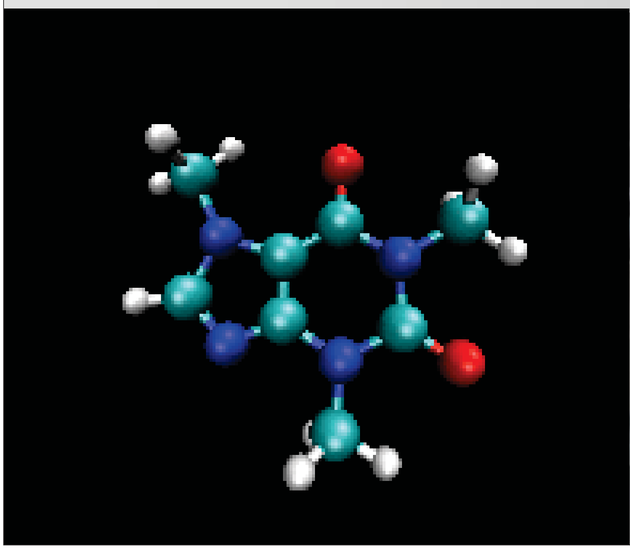


3,316,463 atom system using reaction-field electrostatics instead of PME

GROMACS – Lignocellulose Simulation



Performance Analysis of the AMD EPYC Rome Processors

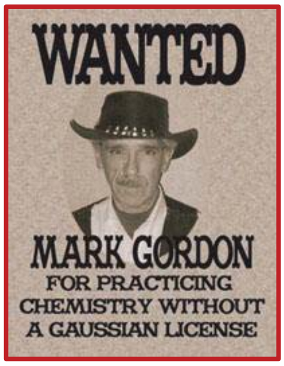


2. Electronic Structure – GAMESS-US and GAMESS-UK (MPI)

Molecular Quantum Chemistry – GAMESS (US)

https://www.msg.chem.iastate.edu/gamess/capabilities.html

- GAMESS can compute **SCF wavefunctions** ranging from RHF, ROHF, UHF, GVB, and MCSCF.
- **Correlation corrections** to these SCF wavefunctions include CI, second order PT and CC approaches, as well as the DFT approximation.
- **Excited states** by CI, EOM, or TD-DFT procedures.
- Nuclear gradients available for **automatic geometry optimisation**, TS searches, or reaction path following.
- Computation of the energy Hessian permits prediction of **vibrational frequencies**, with IR or Raman intensities.
- **Solvent effects** may be modelled by the discrete EF potentials, or continuum models e.g., PCM.
- Numerous **relativistic computations** are available.
- The **Fragment Molecular Orbital** method permits use on very large systems, by dividing the computation into small fragments.



"*Advances in electronic structure theory: GAMESS a decade later*" M.S.Gordon, M.W.Schmidt pp. 1167-1189, in "Theory and Applications of Computational Chemistry: the first forty years" C.E.Dykstra, G.Frenking, K.S.Kim, G.E.Scuseria (editors), Elsevier, Amsterdam, 2005.

Quantum Chemistry Codes:* *Gaussian, GAMESS, NWChem, Dalton, Molpro, Abinit, ACES, Columbus, Turbomole, Spartan, ORCA etc

GAMESS (US) – The DDI Interface

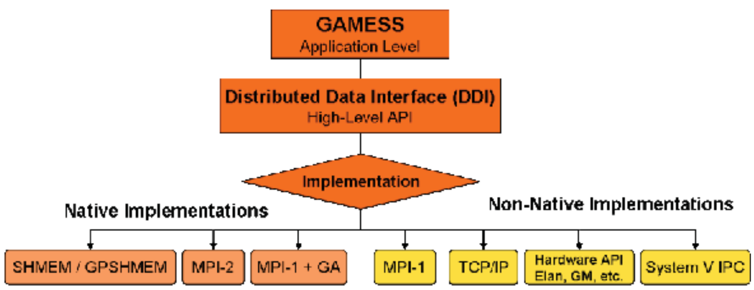
https://www.msg.chem.iastate.edu/gamess/capabilities.html

- The **Distributed Data Interface designed** to permit storage of large data arrays in the aggregate memory of distributed memory, message passing systems.
- Design of this relatively small library discussed, in regard to its implementation over SHMEM, MPI-1, or socket based message libraries.
- Good performance of a MP2 program using DDI demonstrated on both PC and workstation cluster computers
- DDI Developed to avoid using the **Global Arrays** (NWChem) (GDF)!

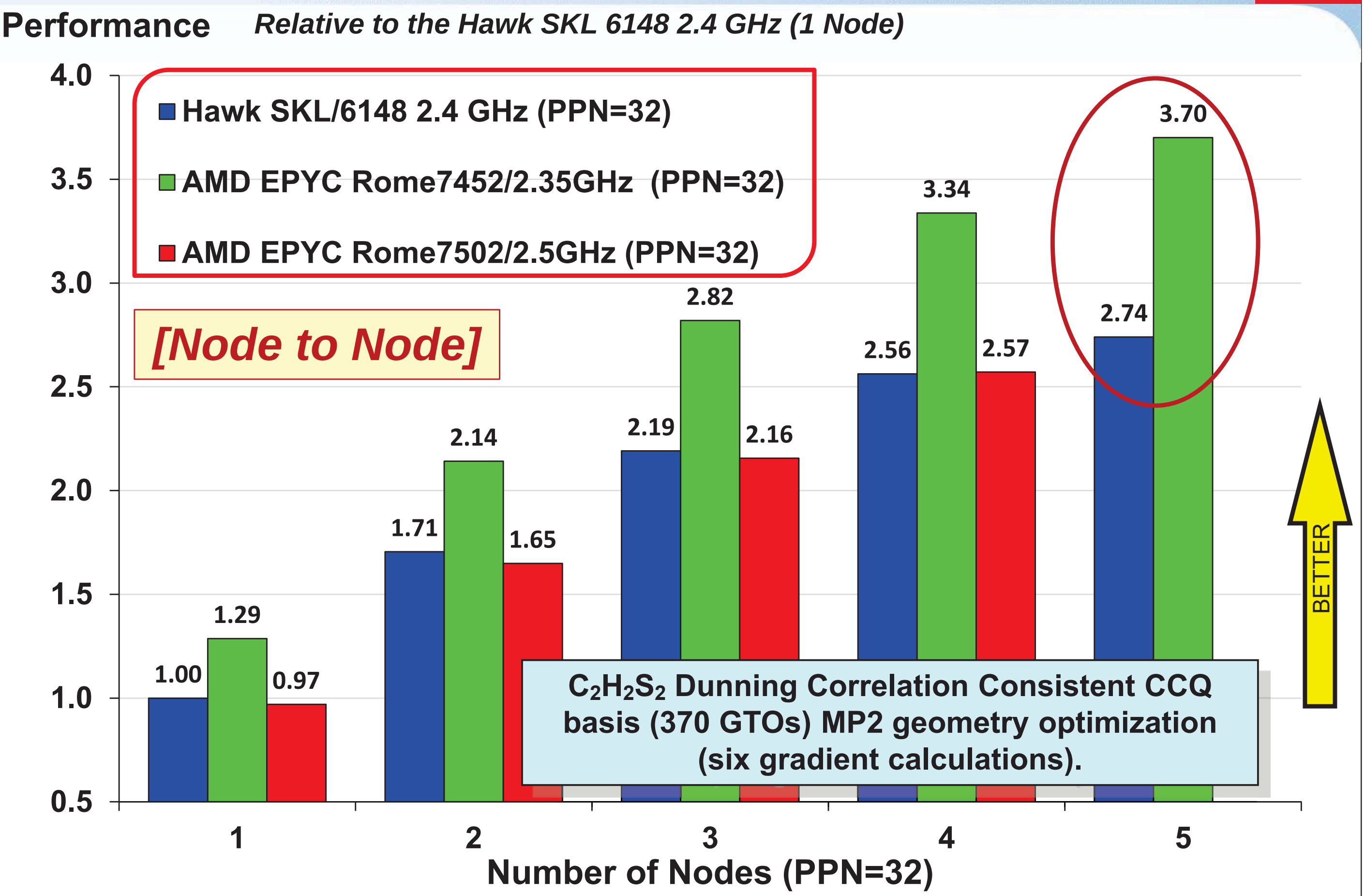
Distributed data interface in GAMESS, June 2000, Computer Physics Communications 128(s 1–2):190–200, DOI: 10.1016/S0010-4655(00)00073-4

Examples

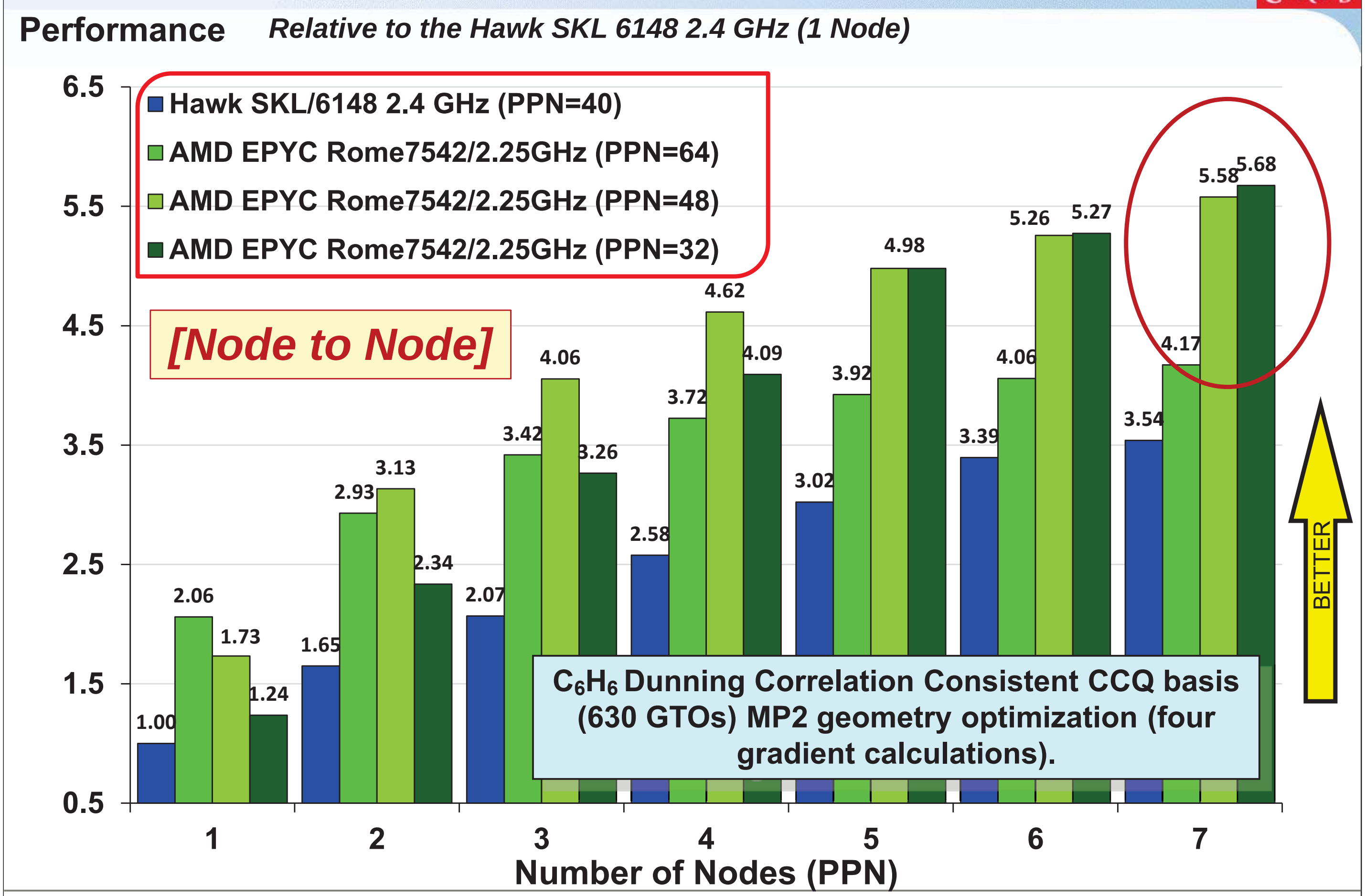
1. **$C_2H_2S_2$: Dunning Correlation Consistent CCQ basis (370 GTOs) MP2 geometry optimization (six gradient calculations).**
2. **C_6H_6 : Dunning Correlation Consistent CCQ basis (630 GTOs) MP2 geometry optimization (four gradient calculations).**



GAMESS (US) Performance – $C_2H_2S_2$ (MP2)



GAMESS (US) Performance – C_6H_6 (MP2)

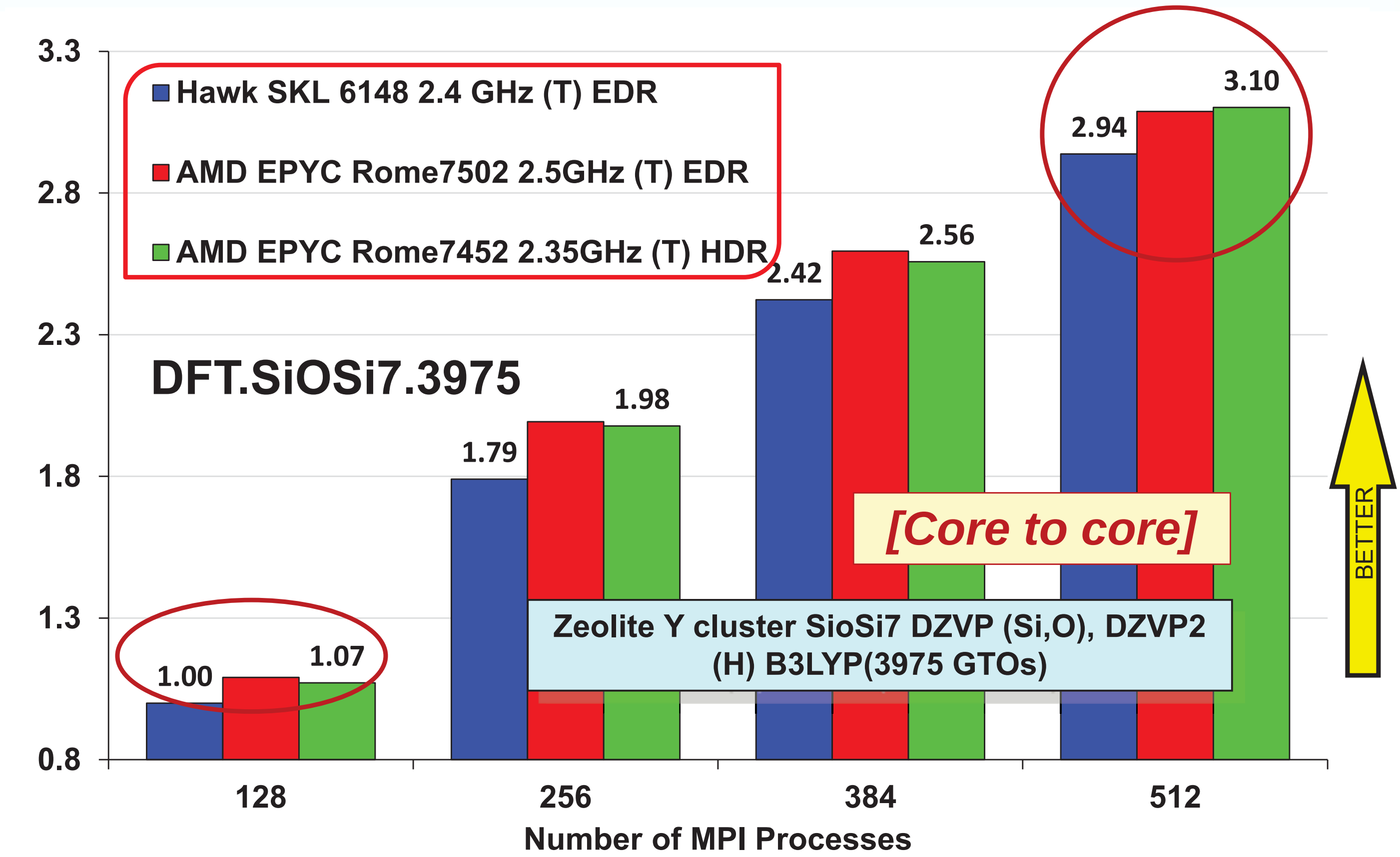


Parallel *Ab-Initio* Electronic Structure Calculations

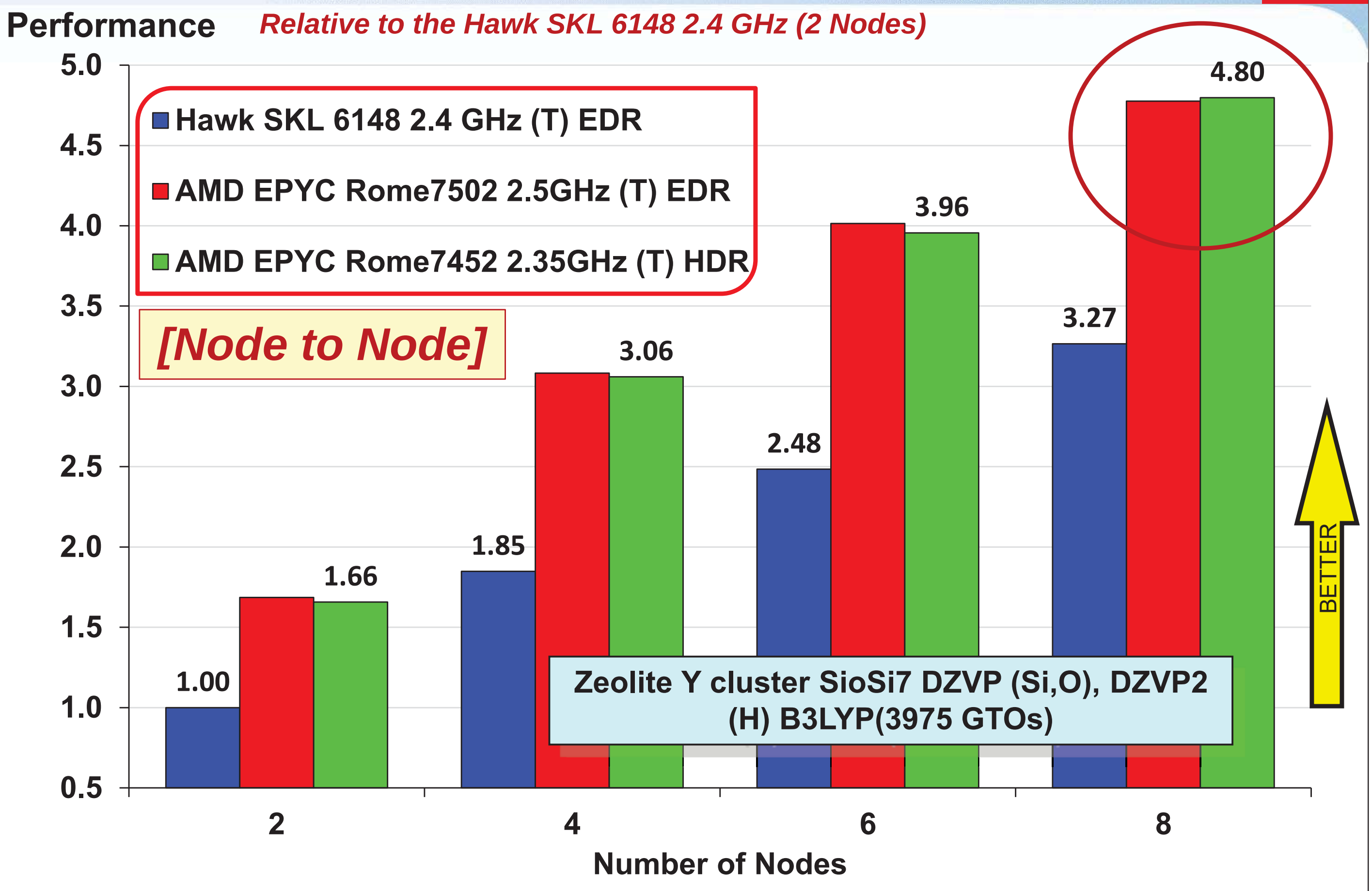
- GAMESS-UK now has **two parallelisation** schemes:
  - The traditional version based on the Global Array tools
    - **retains a lot of replicated data; limited to about 4000 atomic basis functions**
  - Developments by **Ian Bush** (now at Oxford University via NAG Ltd. and Daresbury) extended the system sizes by both GAMESS-UK (molecular systems) and CRYSTAL (periodic systems)
    - **Partial introduction of "Distributed Data" architecture…**
    - **MPI/ScaLAPACK based**
- Three representative examples of increasing complexity.
- **Cyclosporin 6-31g-dp** basis (1855 GTOs) DFT B3LYP (direct SCF)
- **Valinomycin** (dodecadepsipeptide) in water; **DZVP2 DFT** basis, HCTH functional (1620 GTOs) (direct SCF)
- **Zeolite Y cluster SioSi7** DZVP (Si,O), DZVP2 (H) B3LYP(3975 GTOs)

GAMESS-UK Performance - Zeolite Y cluster

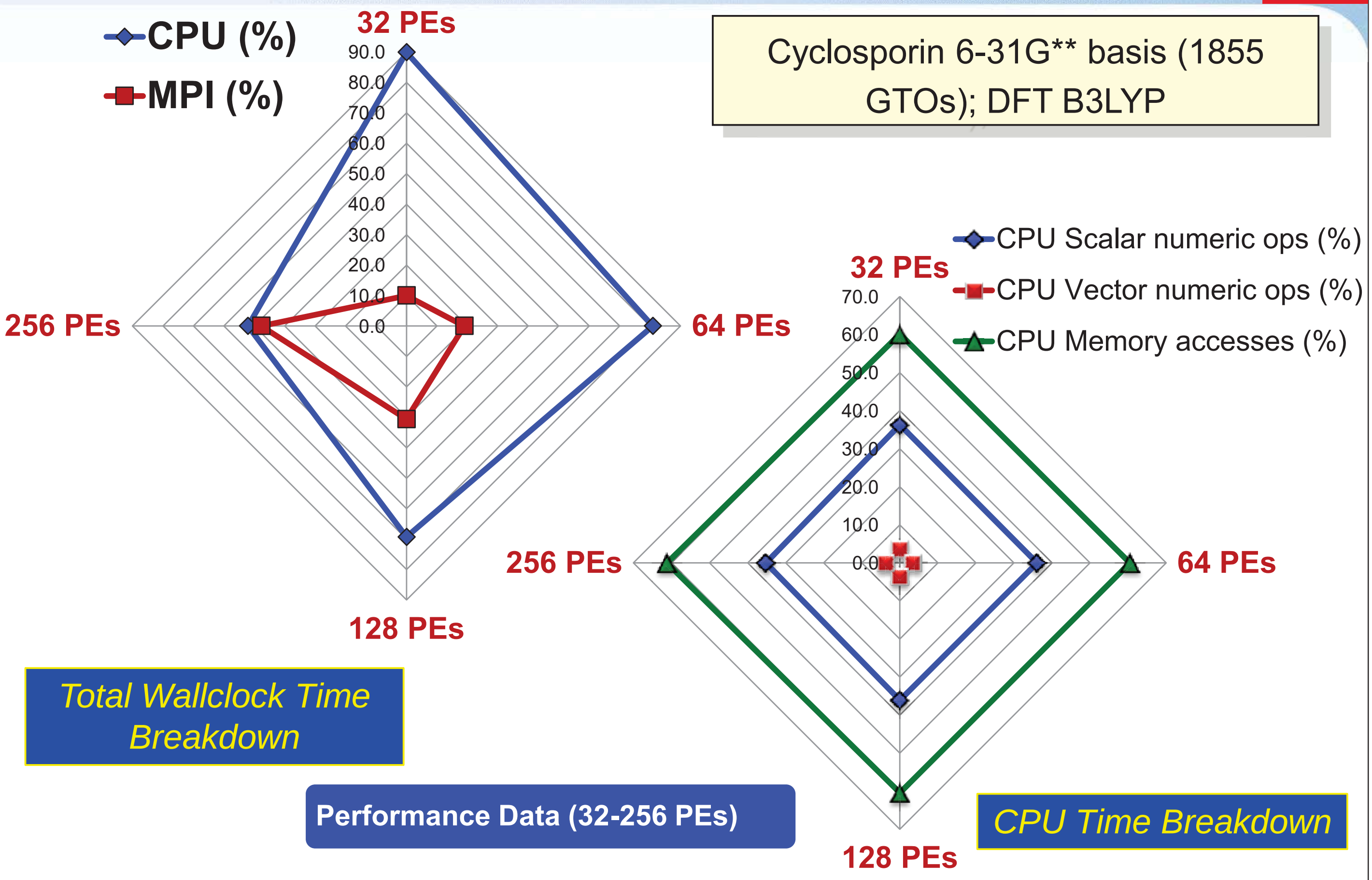
Performance *Relative to the Hawk SKL 6148 2.4 GHz (128 PEs)*



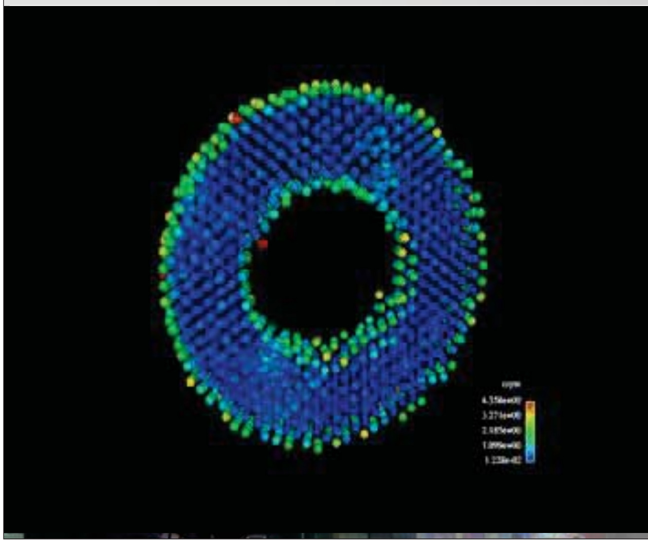
GAMESS-UK Performance - Zeolite Y cluster



GAMESS-UK.MPI DFT – DFT Performance Report



Performance Analysis of the AMD EPYC Rome Processors



3. Advanced Materials Software; Quantum Espresso and VASP

Advanced Materials Software

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

Computational Materials

- **VASP** – performs ab-initio QM molecular dynamics (MD) simulations using **pseudopotentials** or the projector-augmented wave method and a plane wave basis set.
- **Quantum Espresso** – an integrated suite of Open-Source computer codes for electronic-structure calculations and materials modelling at the nanoscale. It is based on density-functional theory (**DFT**), plane waves, and **pseudopotentials**
- **SIESTA** - an *O*(N) **DFT** code for electronic structure calculations and *ab initio* molecular dynamics simulations for molecules and solids. It uses norm-conserving **pseudopotentials** and linear combination of numerical atomic orbitals (LCAO) basis set.
- **CP2K** is a program to perform atomistic and molecular simulations of solid state, liquid, molecular, and biological systems. It provides a framework for different methods such as e.g., **DFT** using a mixed Gaussian & plane waves approach (GPW) and classical pair and many-body potentials.
- **ONETEP** (Order-N Electronic Total Energy Package) is a linear-scaling code for quantum-mechanical calculations based on **DFT**.

VASP – Vienna *Ab-initio* Simulation Package

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

Archer Rank: 1

VASP (**5.4.4**) performs ab-initio QM molecular dynamics (MD) simulations using pseudopotentials or the projector-augmented wave method and a plane wave basis set.

Benchmark	Details
MFI Zeolite	Zeolite ($Si_{96}O_{192}$), 2 k-points, FFT grid: (65, 65, 43); 181,675 points
Pd-O complex	Palladium-Oxygen complex ($Pd_{75}O_{12}$), 10 k-points, FFT grid: (31, 49, 45), 68,355 points

Pd-O Benchmark

- Pd-O complex – $Pd_{75}O_{12}$, 5X4 3-layer supercell running a single point calculation and a planewave cut off of 400eV. Uses the RMM-DIIS algorithm for the SCF and is calculated in real space.
- 10 k-points; maximum number of plane-waves: 34,470
- FFT grid; NGX=31, NGY=49, NGZ=45, giving a total of 68,355 points

Zeolite Benchmark

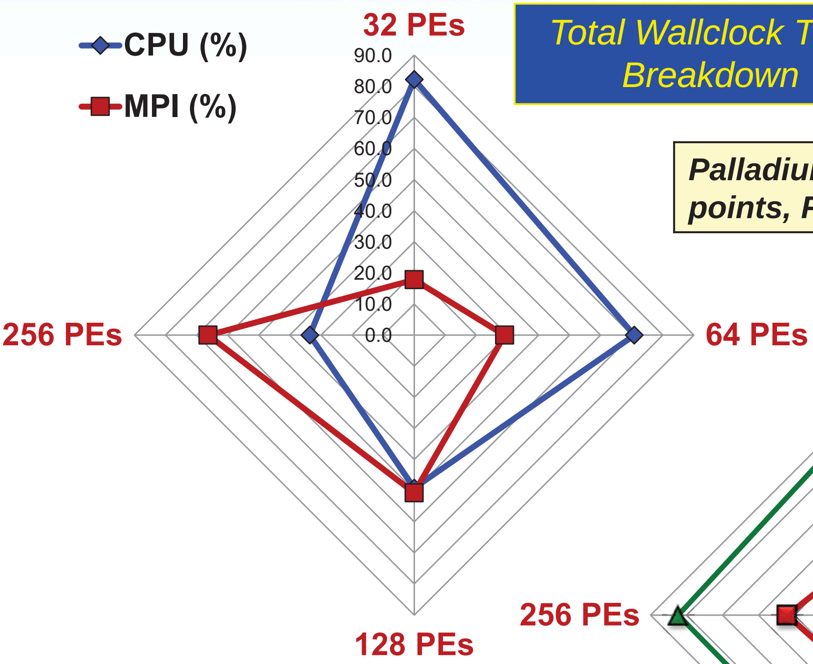
- Zeolite with the MFI structure unit cell running a single point calculation and a planewave cut off of 400eV using the PBE functional
- 2 k-points; maximum number of plane-waves: 96,834
- FFT grid; NGX=65, NGY=65, NGZ=43, giving a total of 181,675 points

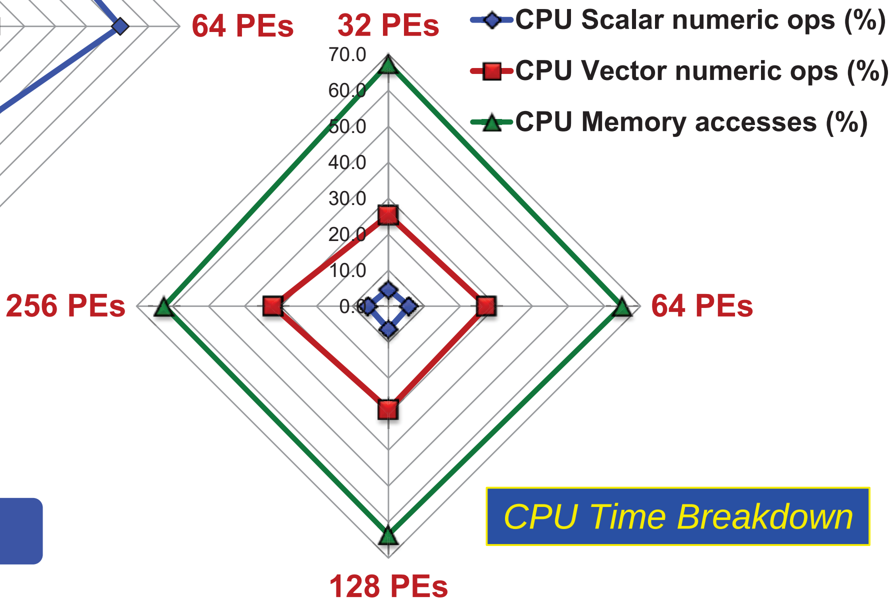
VASP – Pd-O Benchmark Performance Report

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

Total Wallclock Time Breakdown

Palladium-Oxygen complex ($Pd_{75}O_{12}$), 8 k-points, FFT grid: (31, 49, 45), 68,355 points



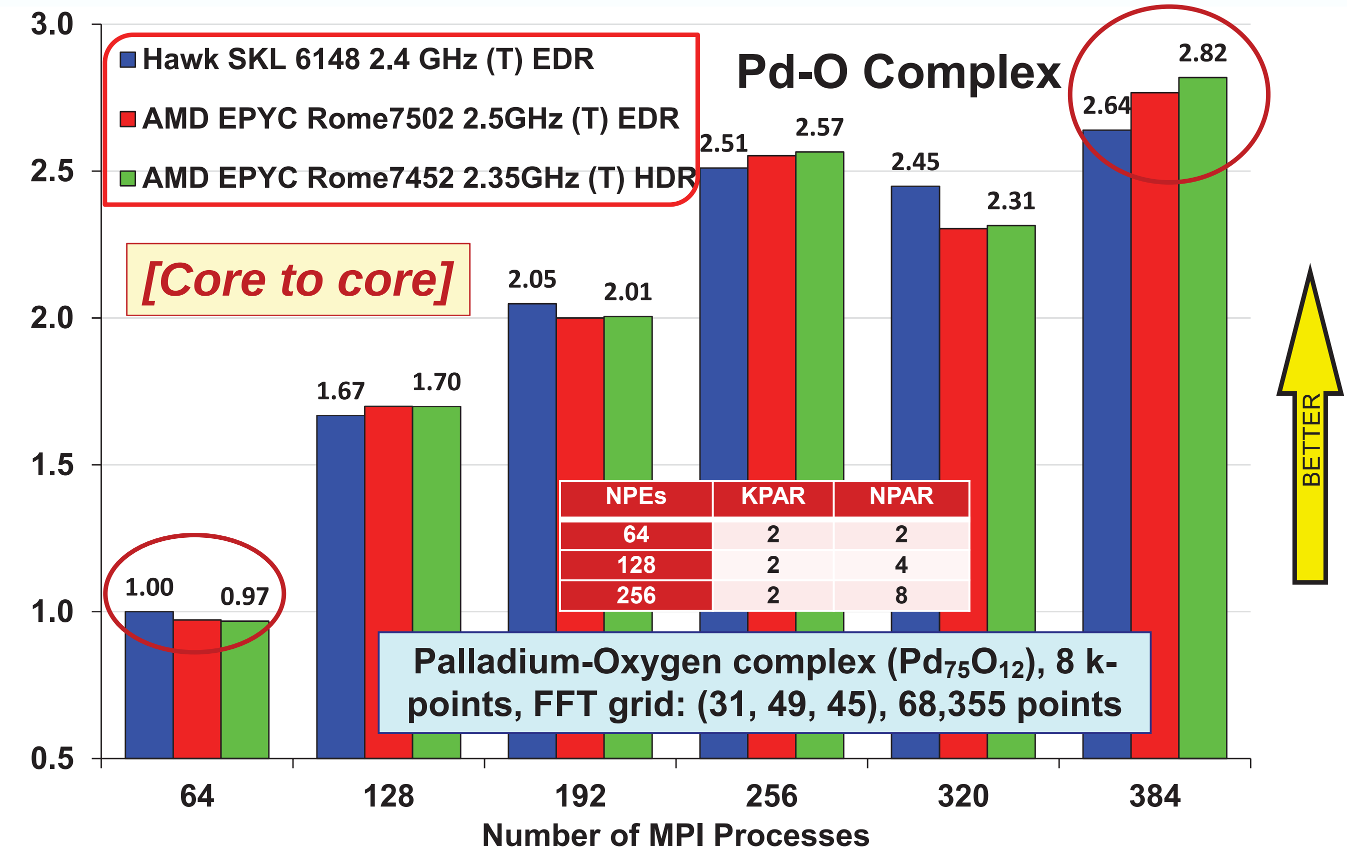


Performance Data (32-256 PEs)

CPU Time Breakdown

VASP 5.4.4 – Pd-O Benchmark - Parallelisation on k-points

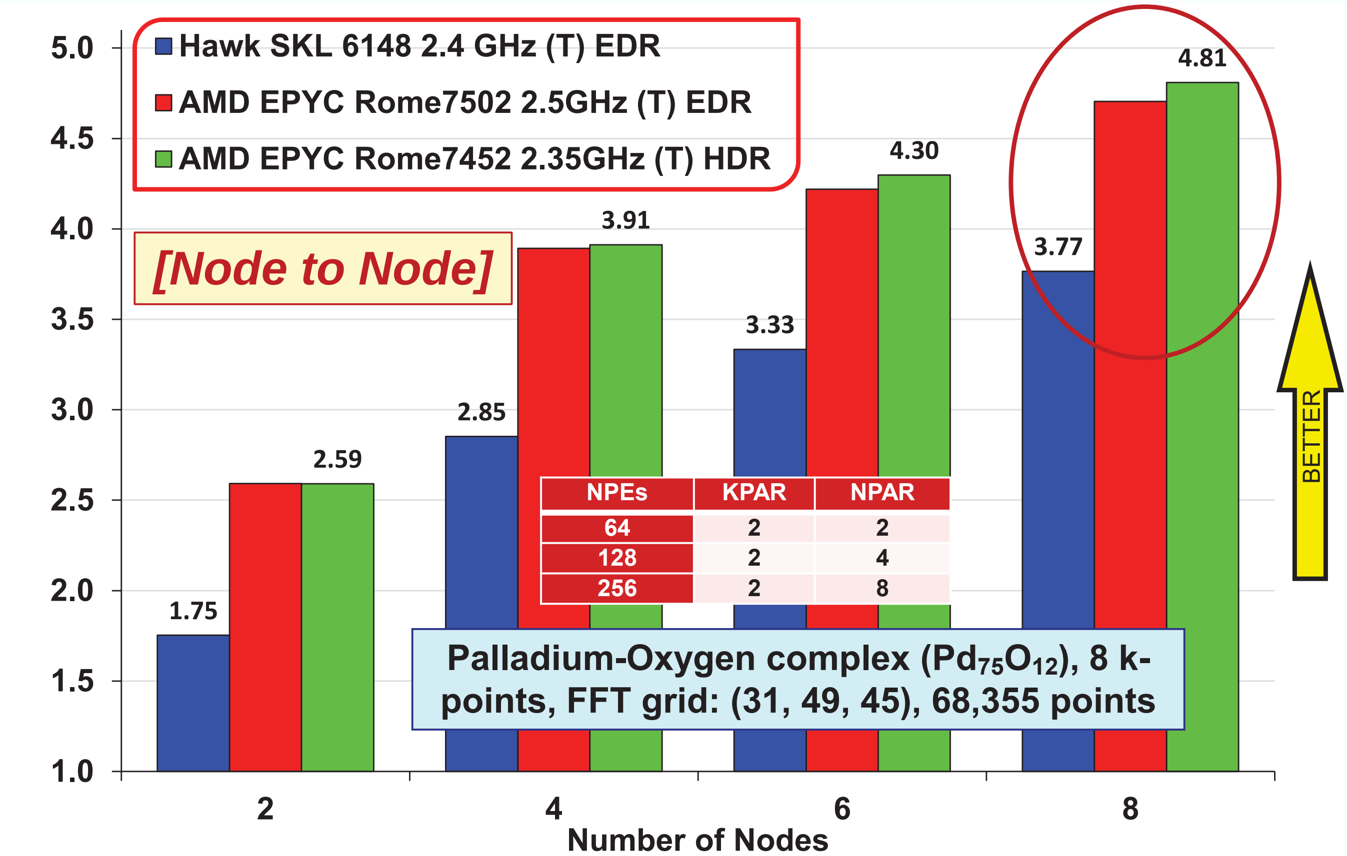
Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*



NPEs	KPAR	NPAR
64	2	2
128	2	4
256	2	8

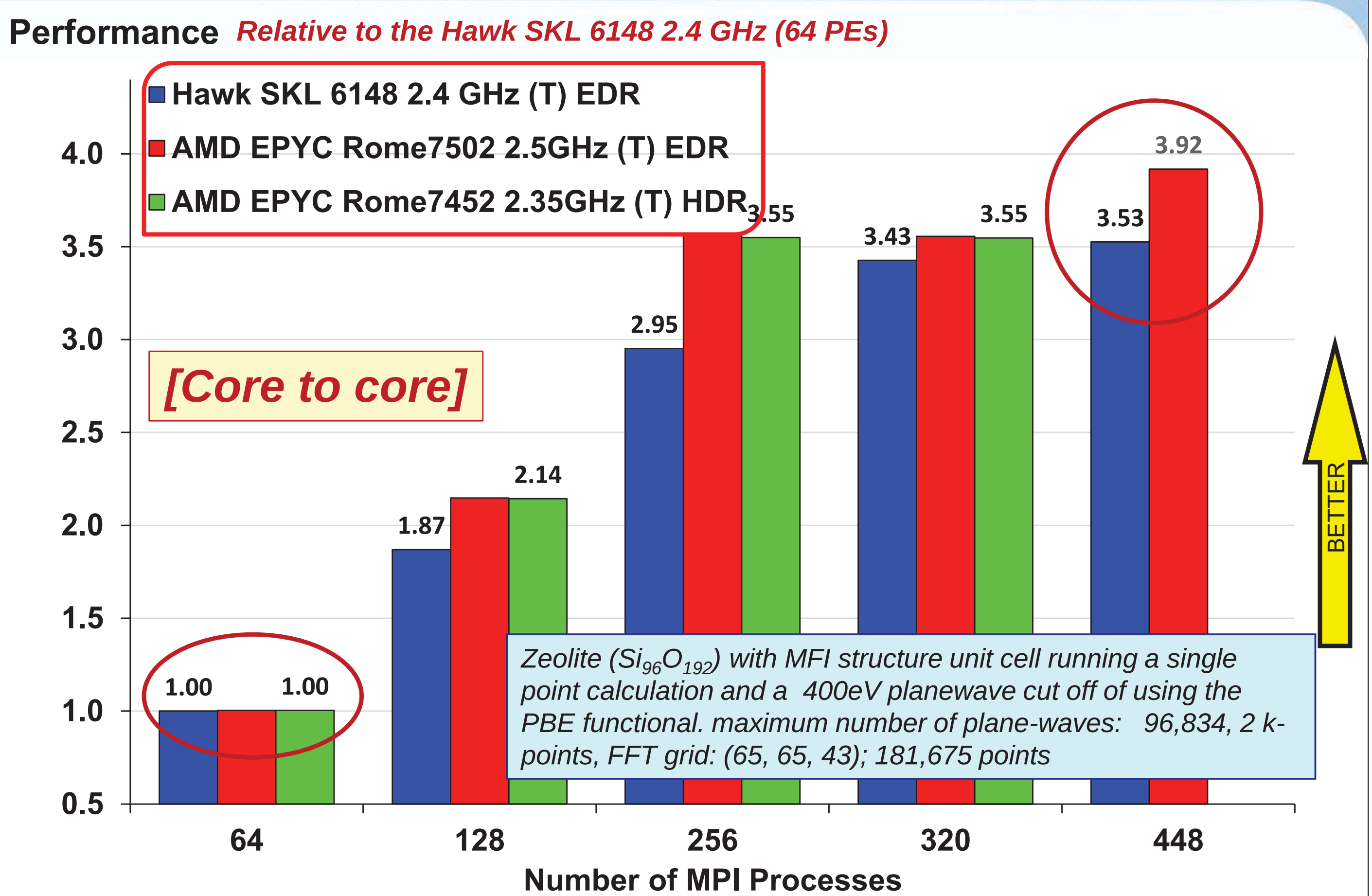
VASP 5.4.4 – Pd-O Benchmark - Parallelisation on k-points

Performance *Relative to the Hawk SKL 6148 2.4 GHz (1 Node)*

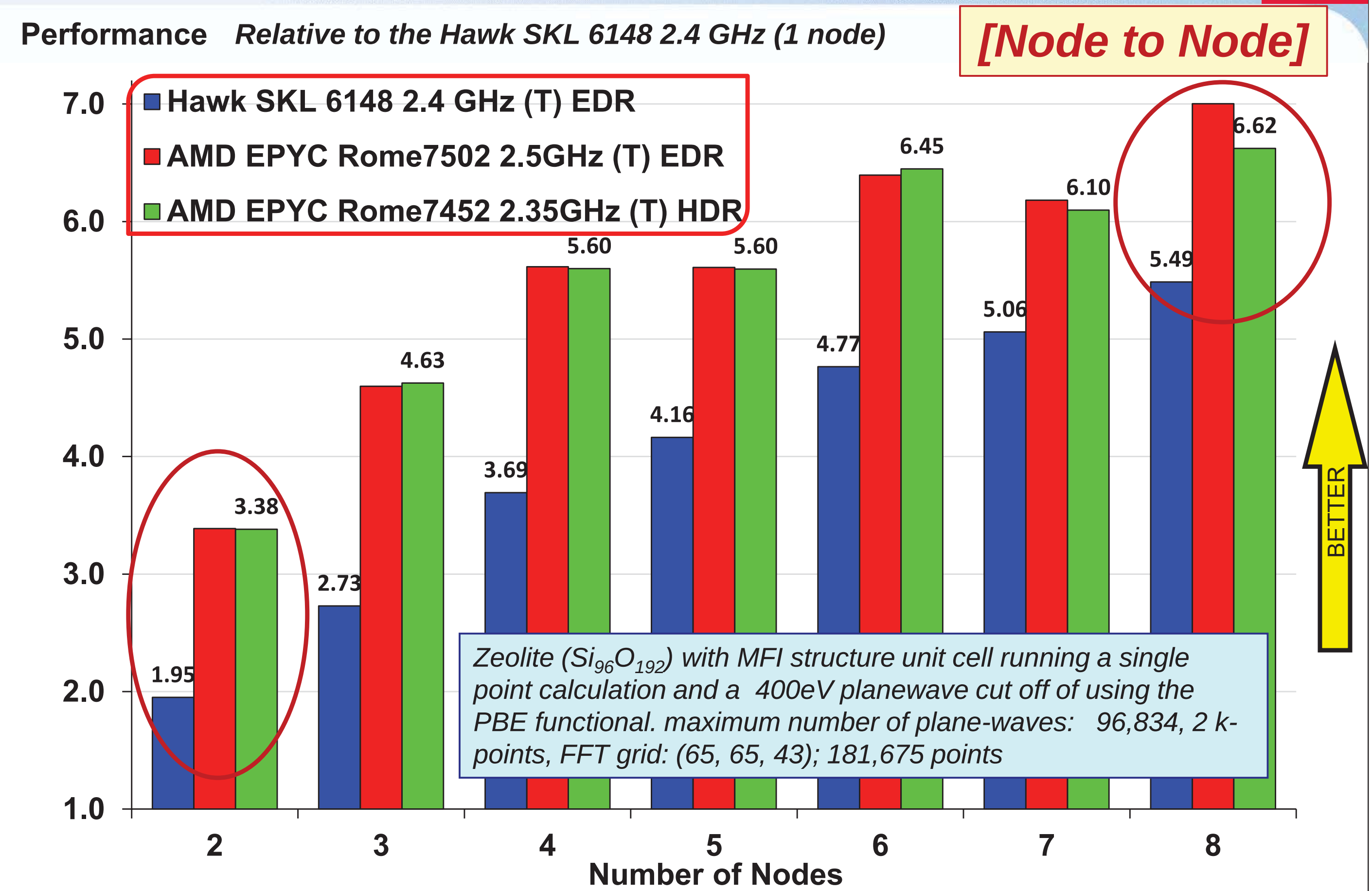


NPEs	KPAR	NPAR
64	2	2
128	2	4
256	2	8

VASP 5.4.4 – Zeolite Benchmark - Parallelisation on k-points



VASP 5.4.4 – Zeolite Benchmark - Parallelisation on k-points



Quantum Espresso v 6.1

Archer Rank: 14

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

- ***Ground-state calculations.***
- Structural Optimization.
- Transition states & minimum energy paths.
- Ab-initio molecular dynamics.
- Response properties (DFPT).
- Spectroscopic properties.
- Quantum Transport.

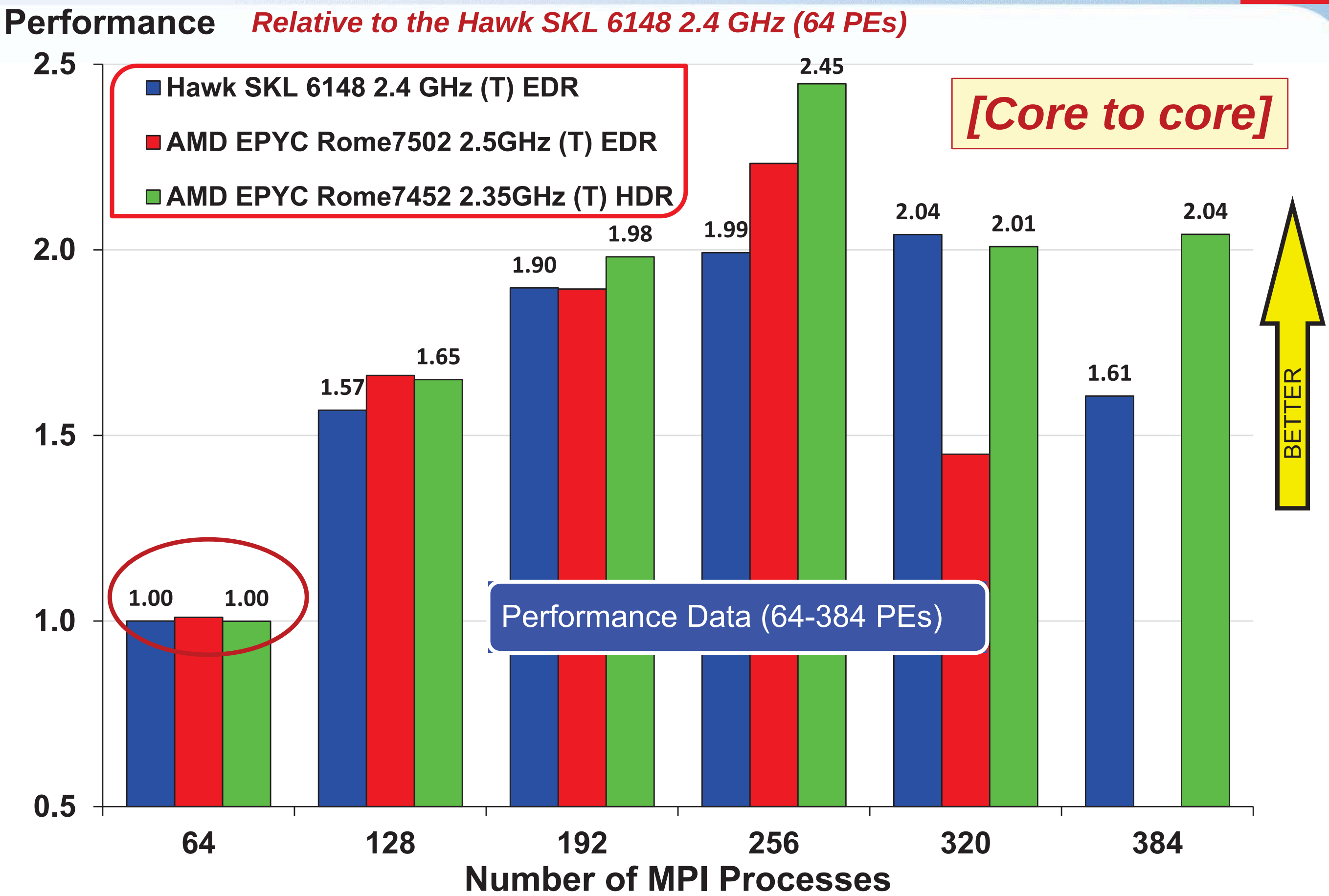
QUANTUMESPRESSO

Benchmark	Details
DEISA AU112	Au complex (Au_{112}), 2,158,381 G-vectors, 2 k-points, FFT dimensions: (180, 90, 288)
PRACE GRIR443	Carbon-Iridium complex ($C_{200}Ir_{243}$), 2,233,063 G-vectors, 8 k-points, FFT dimensions: (180, 180, 192)

Quantum Espresso is an integrated suite of Open-Source computer codes for electronic-structure calculations and materials modelling at the nanoscale. It is based on density-functional theory, plane waves, and pseudopotentials.

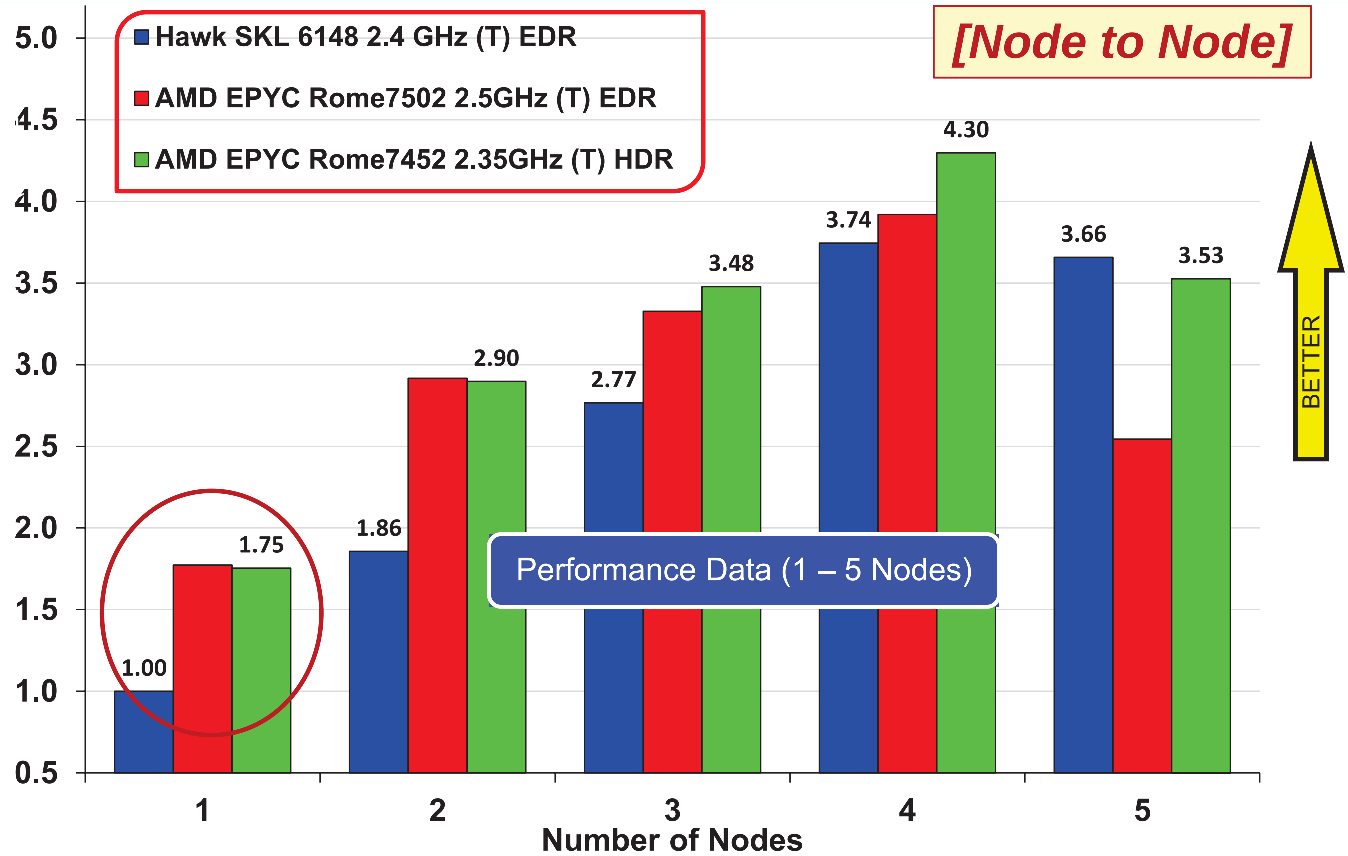
Quantum Espresso – Au_{112}

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

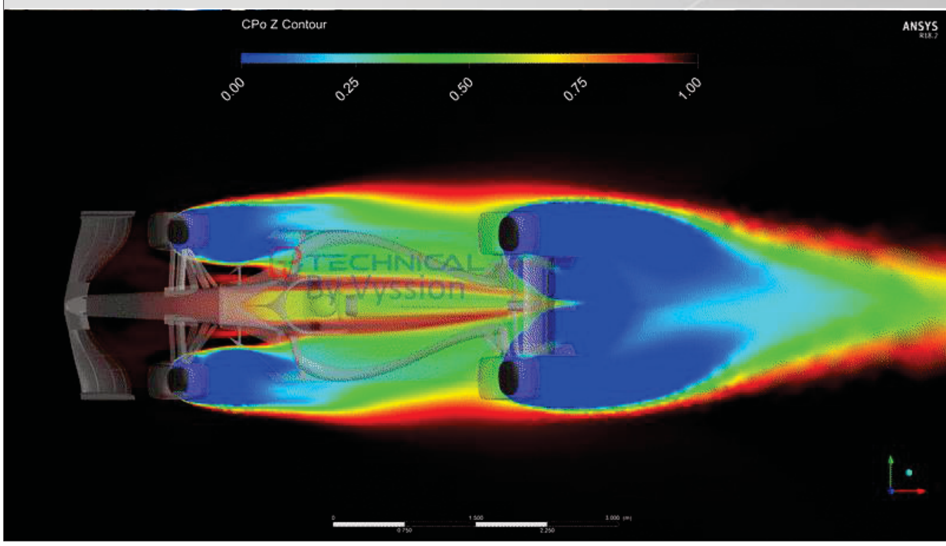


Quantum Espresso – Au_{112}

Performance *Relative to the Hawk SKL 6148 2.4 GHz (1 Node)*



Performance Analysis of the AMD EPYC Rome Processors



4. Engineering and CFD ; OpenFOAM

OpenFOAM - The open source CFD toolbox

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

The OpenFOAM® (Open Field Operation and Manipulation) CFD Toolbox is a free, open source CFD software package produced by OpenCFD Ltd.

http://www.openfoam.com/

Archer Rank: 12

Features

- OpenFOAM has an extensive range of features to solve anything from complex fluid flows involving chemical reactions, turbulence and heat transfer, to solid dynamics and electromagnetics. **v1906**

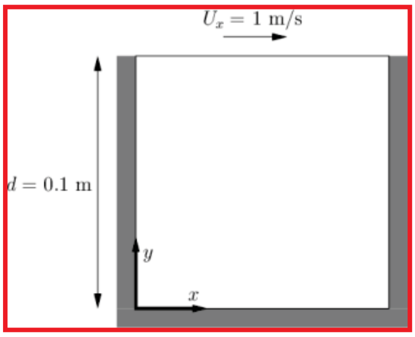
Applications

- Includes over 90 solver applications that simulate specific problems in engineering mechanics and over 180 utility applications that perform pre- and post-processing tasks, e.g. meshing, data visualisation, etc.

http://www.openfoam.org/docs/user/cavity.php#x5-170002.1.5

Lid-driven cavity flow (Cavity 3d)

- *Isothermal, incompressible flow in a 2D square domain. The geometry has all the boundaries of the square are walls. The top wall moves in the x-direction at 1 m/s while the other 3 are stationary. Initially, the flow is assumed laminar and is solved on a uniform mesh using the icoFoam solver.*

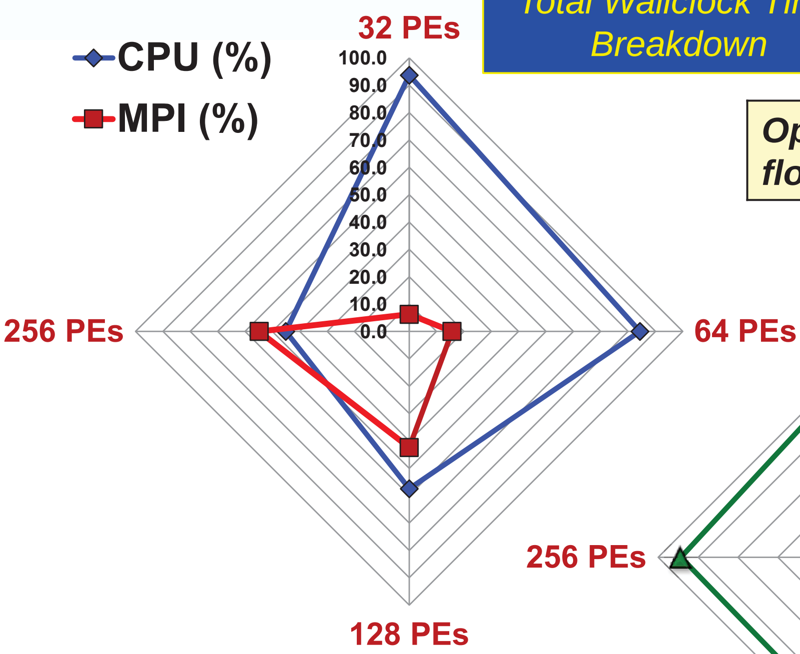


Geometry of the lid driven cavity

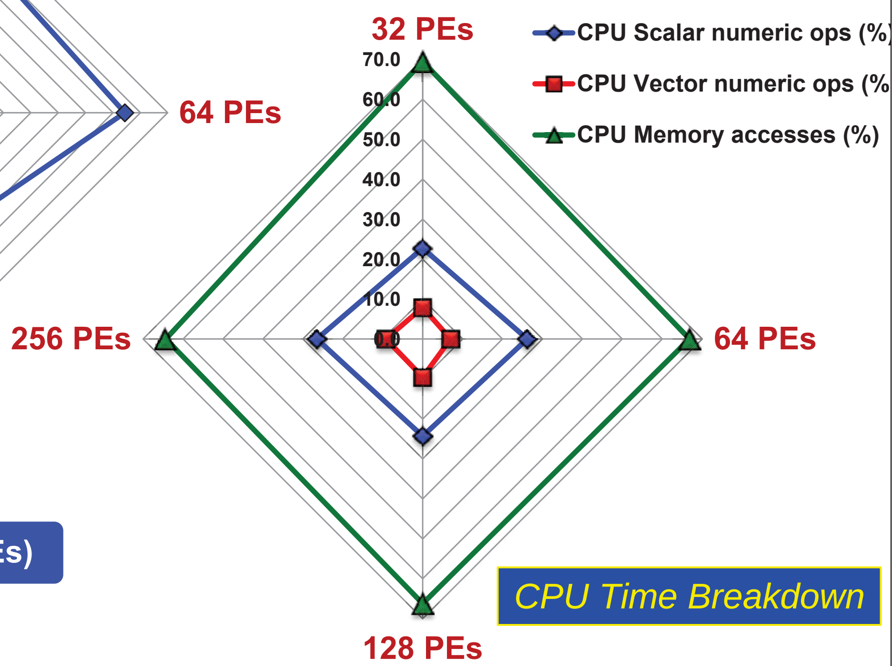
OpenFOAM – Cavity 3d-3M Performance Report

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

Total Wallclock Time Breakdown

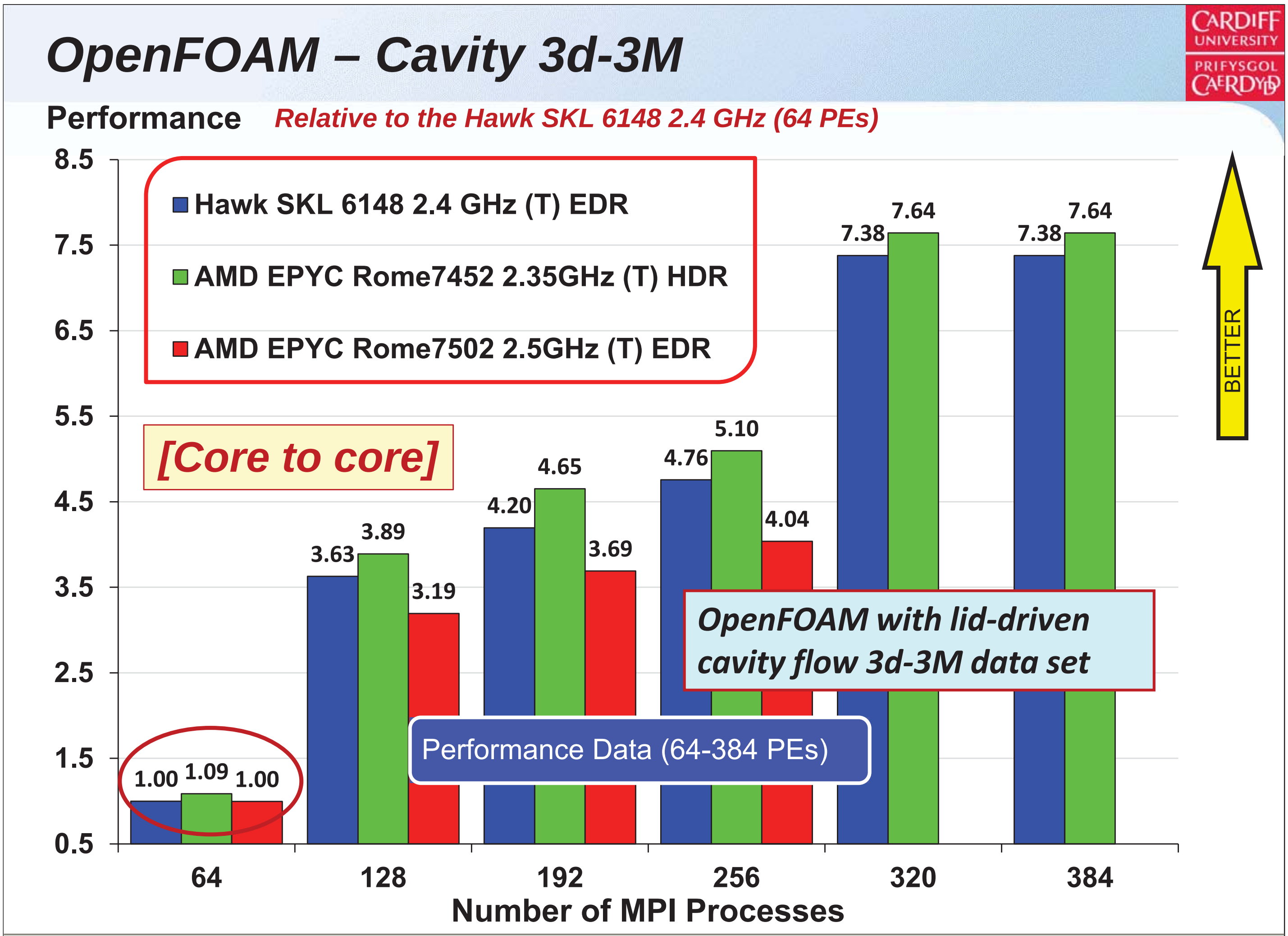


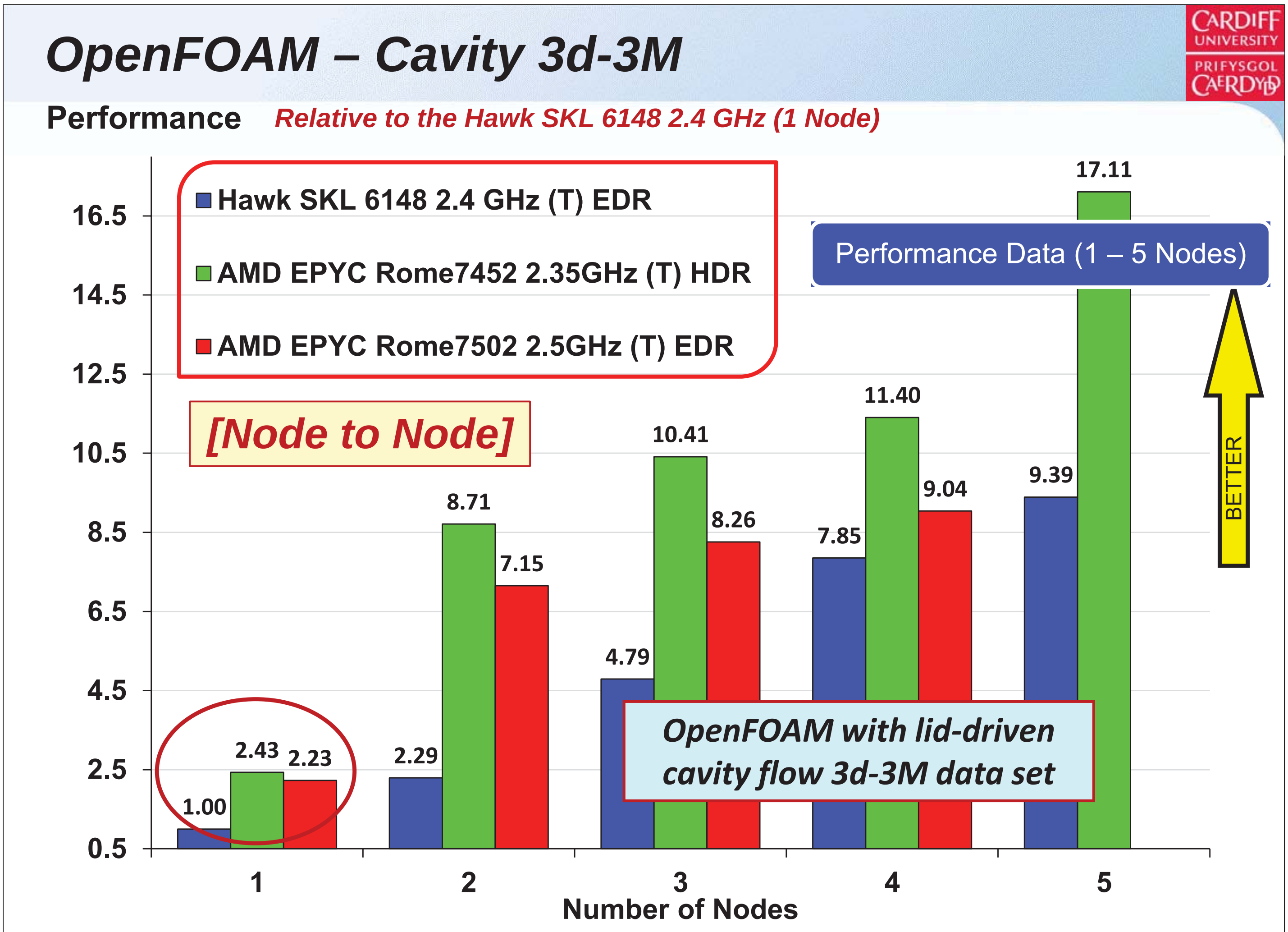
OpenFOAM with lid-driven cavity flow 3d-3M data set



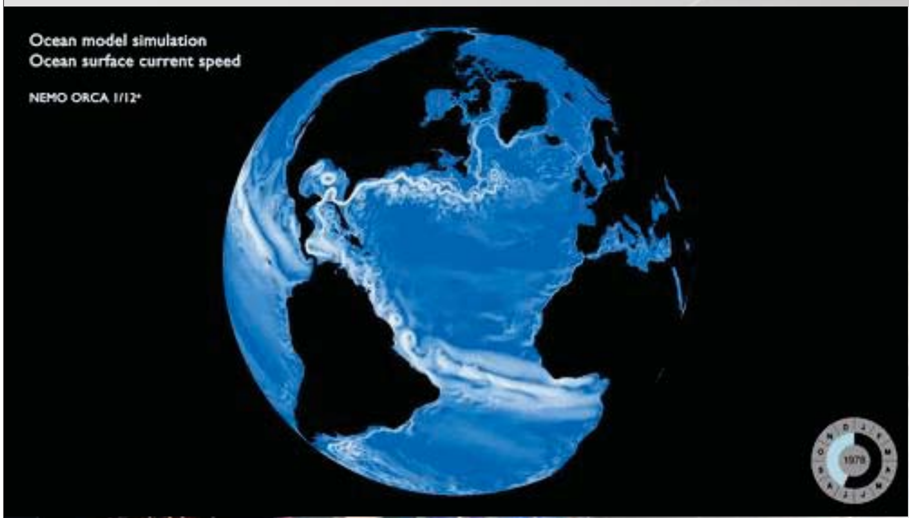
Performance Data (32-256 PEs)

CPU Time Breakdown


OpenFOAM – Cavity 3d-3M
Performance
Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)
Hawk SKL 6148 2.4 GHz (T) EDR
AMD EPYC Rome7452 2.35GHz (T) HDR
AMD EPYC Rome7502 2.5GHz (T) EDR
[Core to core]
OpenFOAM with lid-driven cavity flow 3d-3M data set
Performance Data (64-384 PEs)
BETTER
8.5
7.5
6.5
5.5
4.5
3.5
2.5
1.5
0.5
1.00 1.09 1.00
3.63 3.89 3.19
4.20 4.65 3.69
4.76 5.10 4.04
7.38 7.64
7.38 7.64
64
128
192
256
320
384
Number of MPI Processes


OpenFOAM – Cavity 3d-3M
Performance
Relative to the Hawk SKL 6148 2.4 GHz (1 Node)
Hawk SKL 6148 2.4 GHz (T) EDR
AMD EPYC Rome7452 2.35GHz (T) HDR
AMD EPYC Rome7502 2.5GHz (T) EDR
[Node to Node]
Performance Data (1 – 5 Nodes)
OpenFOAM with lid-driven cavity flow 3d-3M data set
BETTER
16.5
14.5
12.5
10.5
8.5
6.5
4.5
2.5
0.5
1.00 2.43 2.23
2.29 8.71 7.15
4.79 10.41 8.26
7.85 11.40 9.04
9.39 17.11
1
2
3
4
5
Number of Nodes

Performance Analysis of the AMD EPYC Rome Processors



5 NEMO - Nucleus for European Modelling of the Ocean

The NEMO Code

Archer Rank: 3

- NEMO (**Nucleus for European Modelling of the Ocean**) is a state-of-the-art modelling framework of ocean related engines for oceanographic research, operational oceanography, seasonal forecast and [paleo] climate studies.
- **ORCA family**: global ocean with tripolar grid; The ORCA family is a series of global ocean configurations that are run together with the LIM sea-ice model (ORCA-LIM) and possibly with **PISCES biogeochemical model** (ORCA-LIM-PISCES), **using various resolutions.**
- Analysis based on the BENCH benchmarking configurations of NEMO release **version 4.0** that are rather straightforward to set up.
- Code obtained from https://forge.ipsl.jussieu.fr using:

```
$ svn co https://forge.ipsl.jussieu.fr/nemo/svn/NEMO/releases/release-4.0
$ cd release-4.0
```

  The code relies on efficient installations of both **NetCDF and HDF5** installations.
- Executables for two BENCH variants, here named, **BENCH_ORCA_SI3_PISCES** and **BENCH_ORCA_SI3.** PISCES augments the standard model with a **bio-geochemical model**.

The NEMO Code

- To run the model in BENCH configurations either executable can be run from within a directory that contains copies of or links to the namelist_* files from the respective directory:

  ./tests/BENCH_ORCA_SI3/EXP00/ and
  ./tests/BENCH_ORCA_SI3_PISCES/EXP00

  Both variants require namelist_ref, namelist_ice_{ref,cfg}, and one of the files namelist_cfg_orca{1,025,12}_like, renamed as namelist_cfg (referred to as ORCA1, ORCA025, ORCA12 variants, respectively, where 1, 025, 12 indicate to the **nominal horizontal model resolutions of 1 degree, 1/4 of a degree, and 1/12 of a degree**); variant BENCH_ORCA_SI3_PISCES additionally requires files namelist_{top,pisces}_ref and namelist_{top,pisces}_cfg.
- In total this provides six benchmark variants.

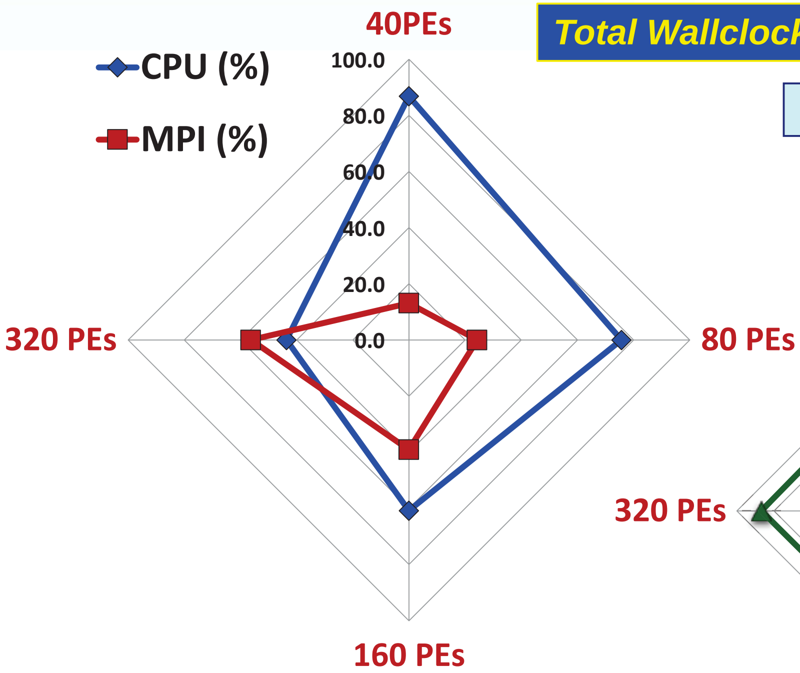
  BENCH_ORCA_SI3/ORCA1, ORCA025 and ORCA12

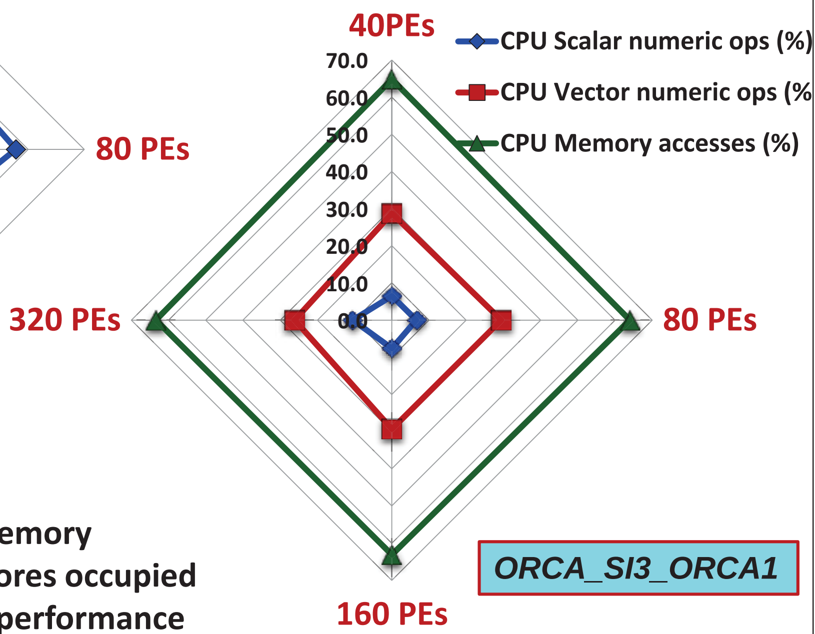
  BENCH_ORCA_SI3_PISCES/ORCA1, ORCA025 and ORCA012.
- Increasing the resolution typically increases computational resources by an **×10**. Experience limited to 5 of these configurations - **ORCA_SI3_PISCES / ORCA012** requires unrealistic memory configurations.

NEMO – ORCA_SI3 Performance Report

Total Wallclock Time Breakdown

horizontal resolutions of 1-degree





ORCA_SI3_ORCA1

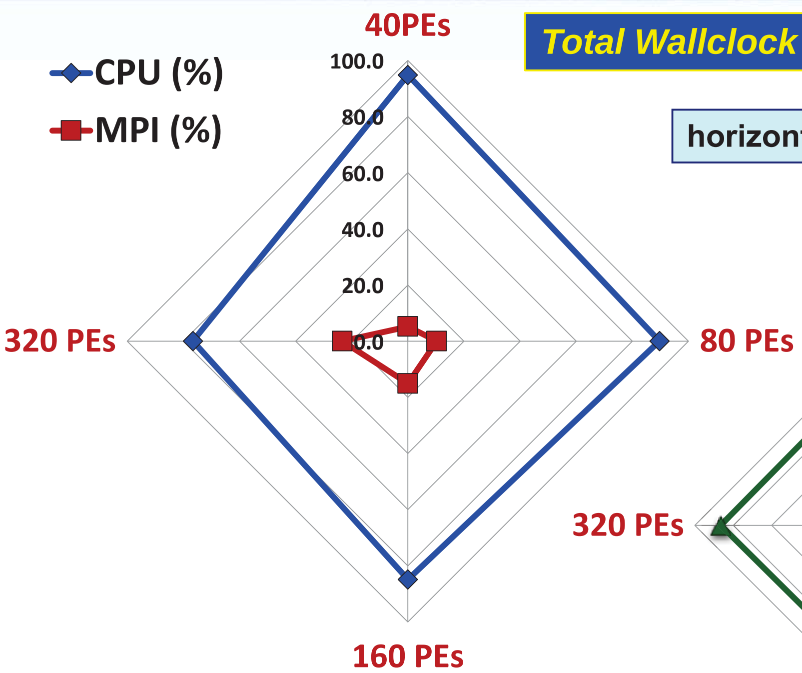
NEMO performance is dominated by memory bandwidth – running with 50% of the cores occupied on each Hawk node typically improves performance by ca. 1.6 for a fixed number of MPI processes.

CPU Time Breakdown

NEMO – ORCA_SI3_PISCES Performance Report

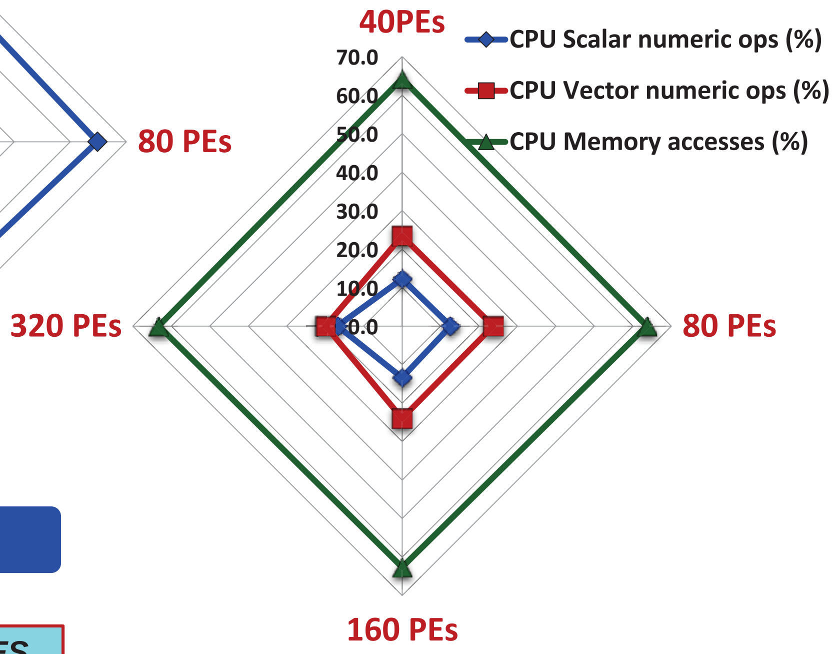
Total Wallclock Time Breakdown

horizontal resolutions of 1-degree



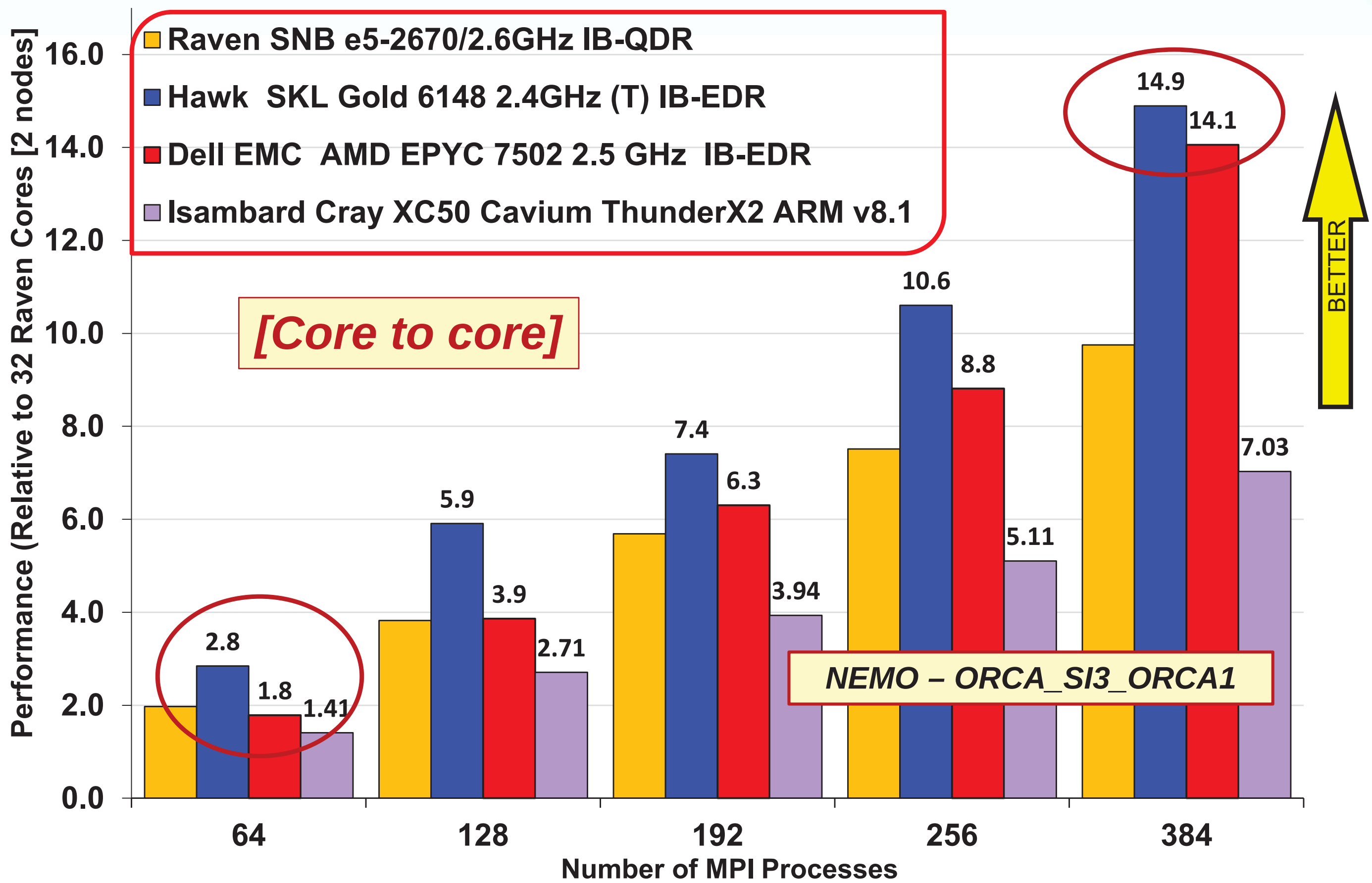
Performance Data (40-320 PEs)

NEMO – ORCA_SI3_PISCES

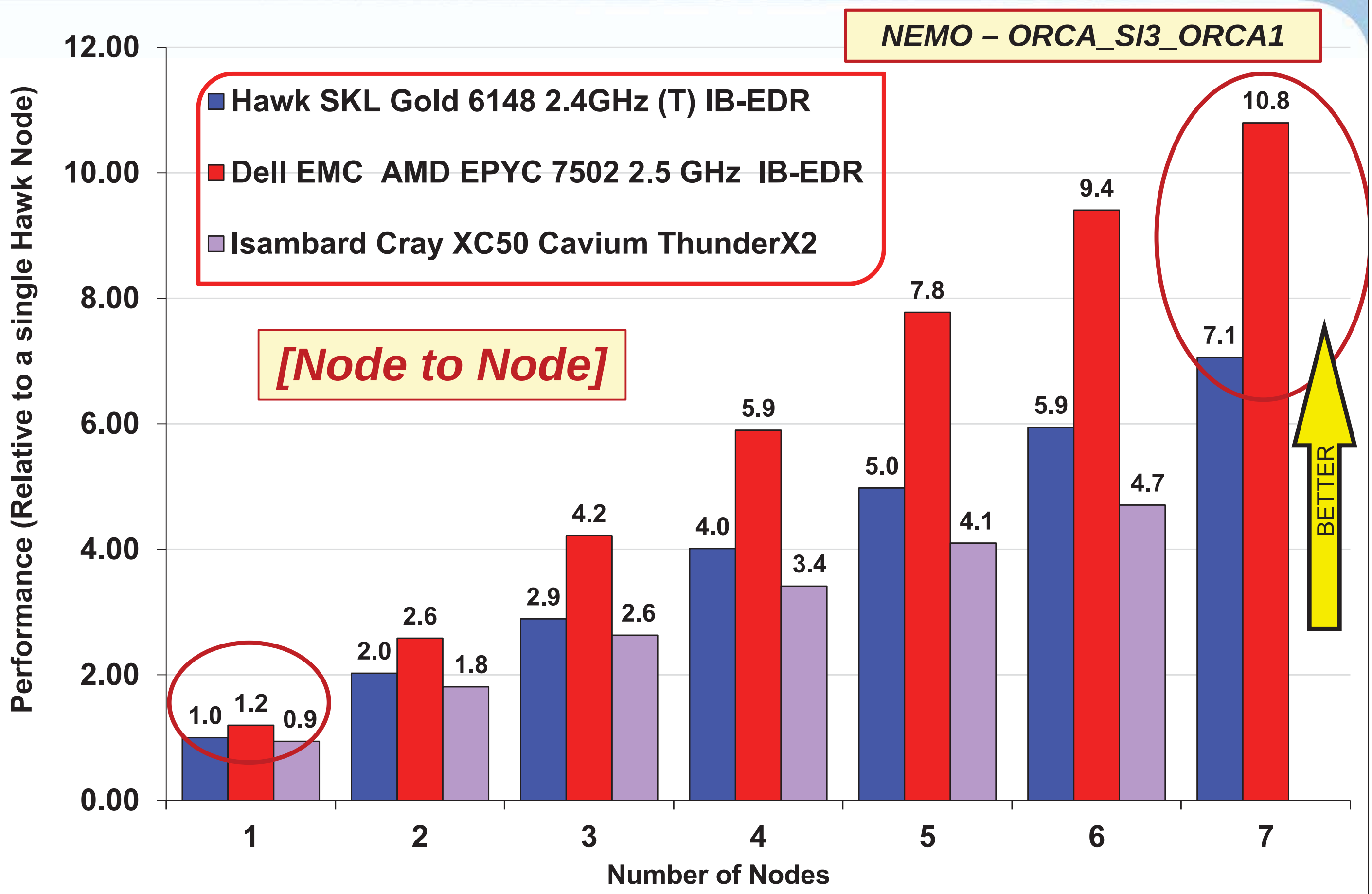


CPU Time Breakdown

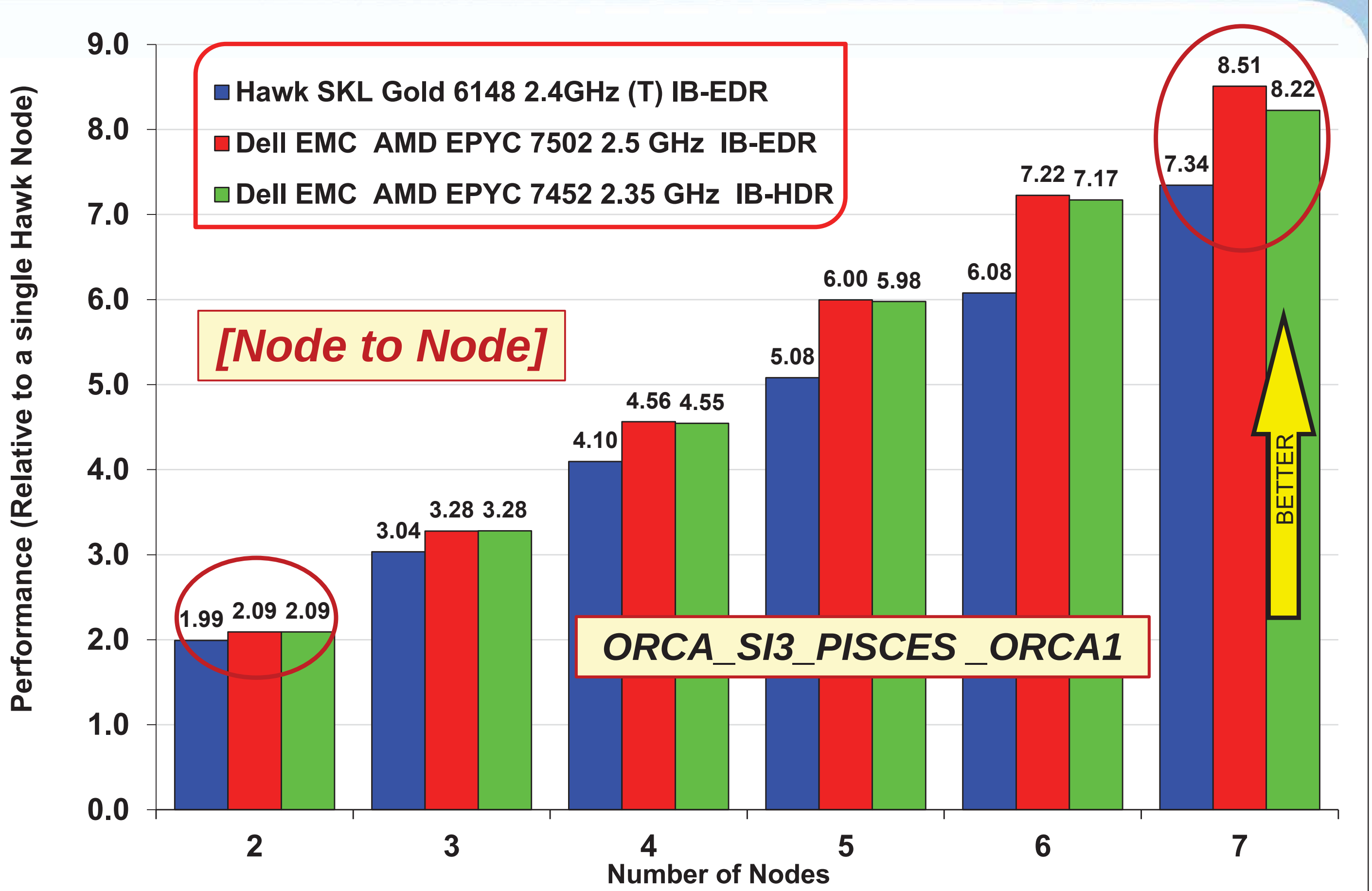
ORCA_SI3 1.0 degree : Core-to-Core Performance



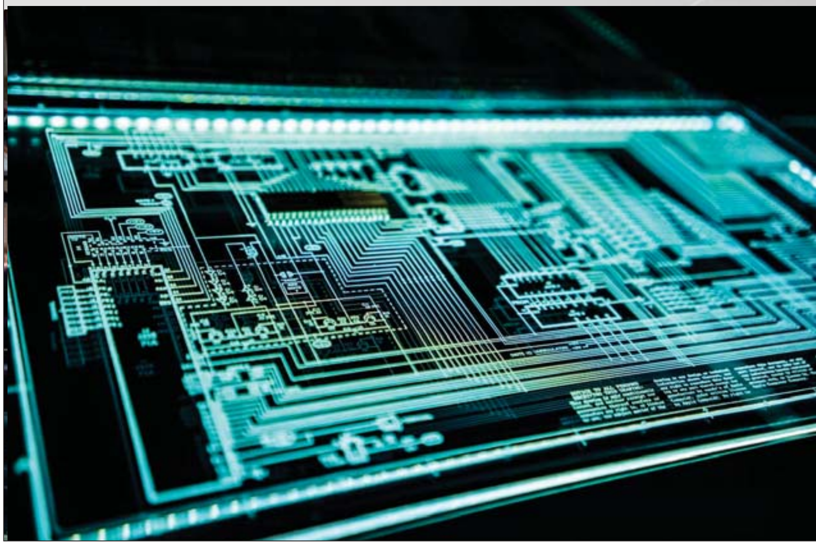
ORCA_SI3 1.0 degree : Node Performance



ORCA_SI3_PISCES 1.0 degree : Node Performance



Performance Analysis of the AMD EPYC Rome Processors

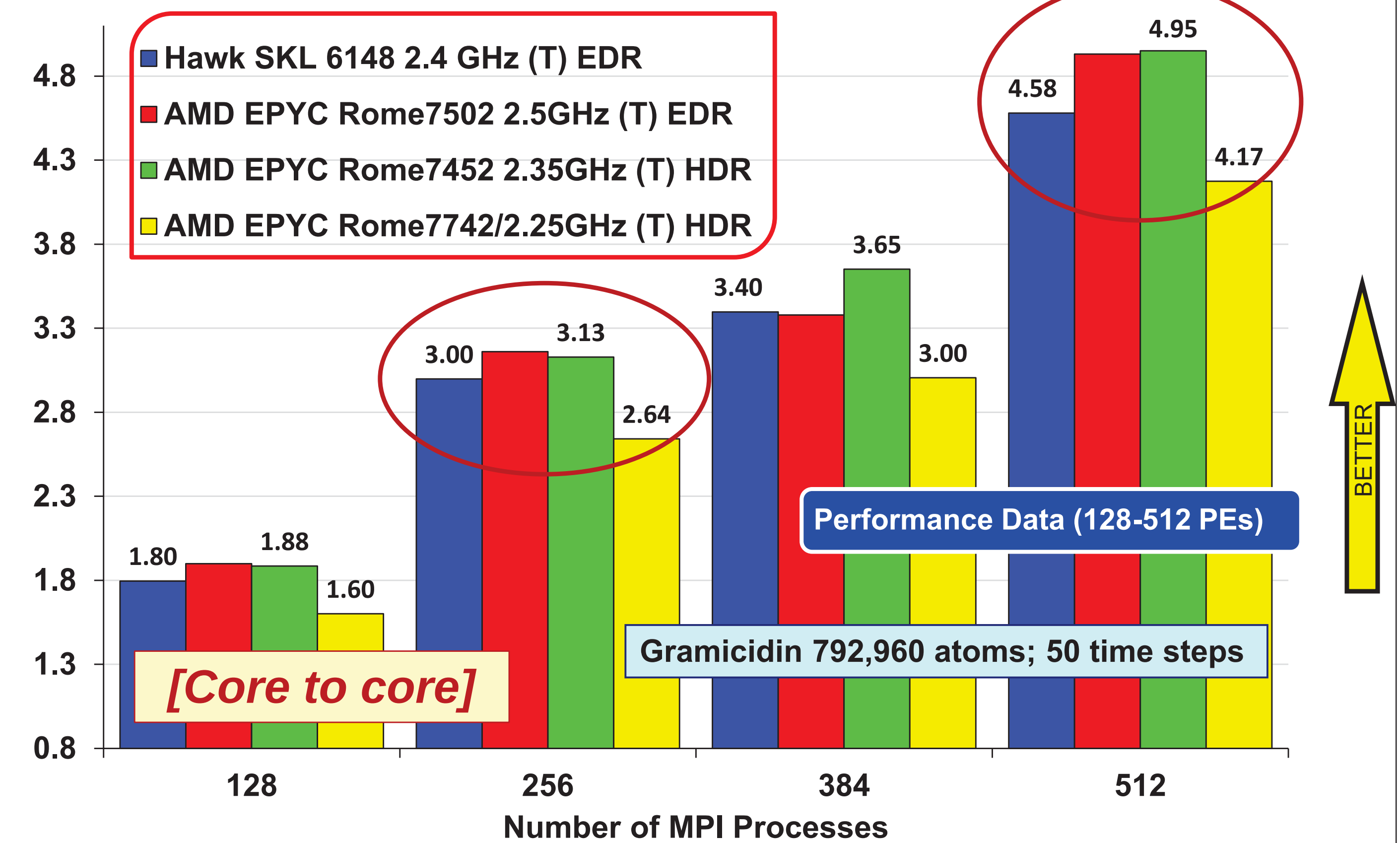


Performance Attributes of the EPYC Rome 7742 (64-core) Processor

DL_POLY 4 – Gramicidin Simulation

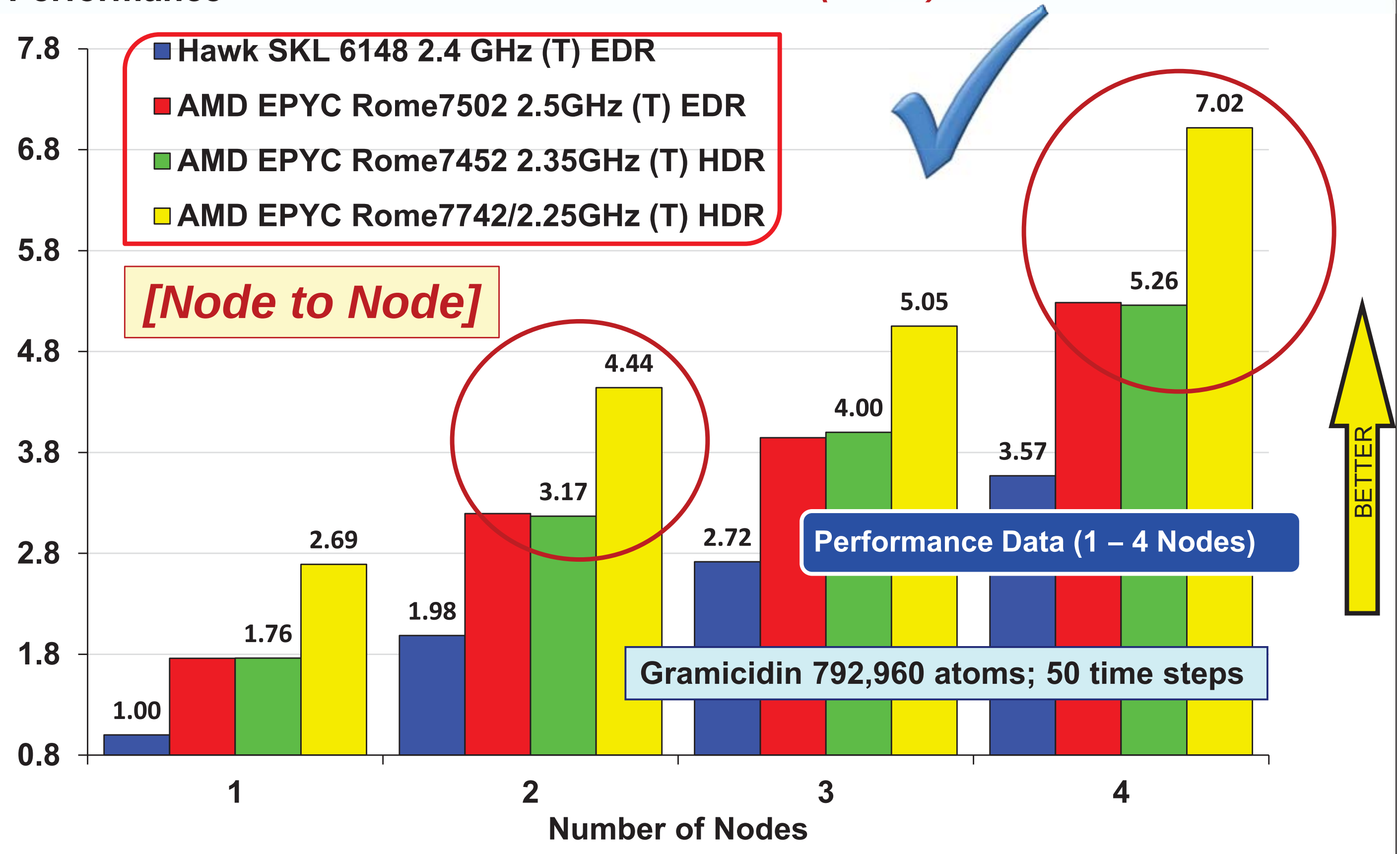
CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*



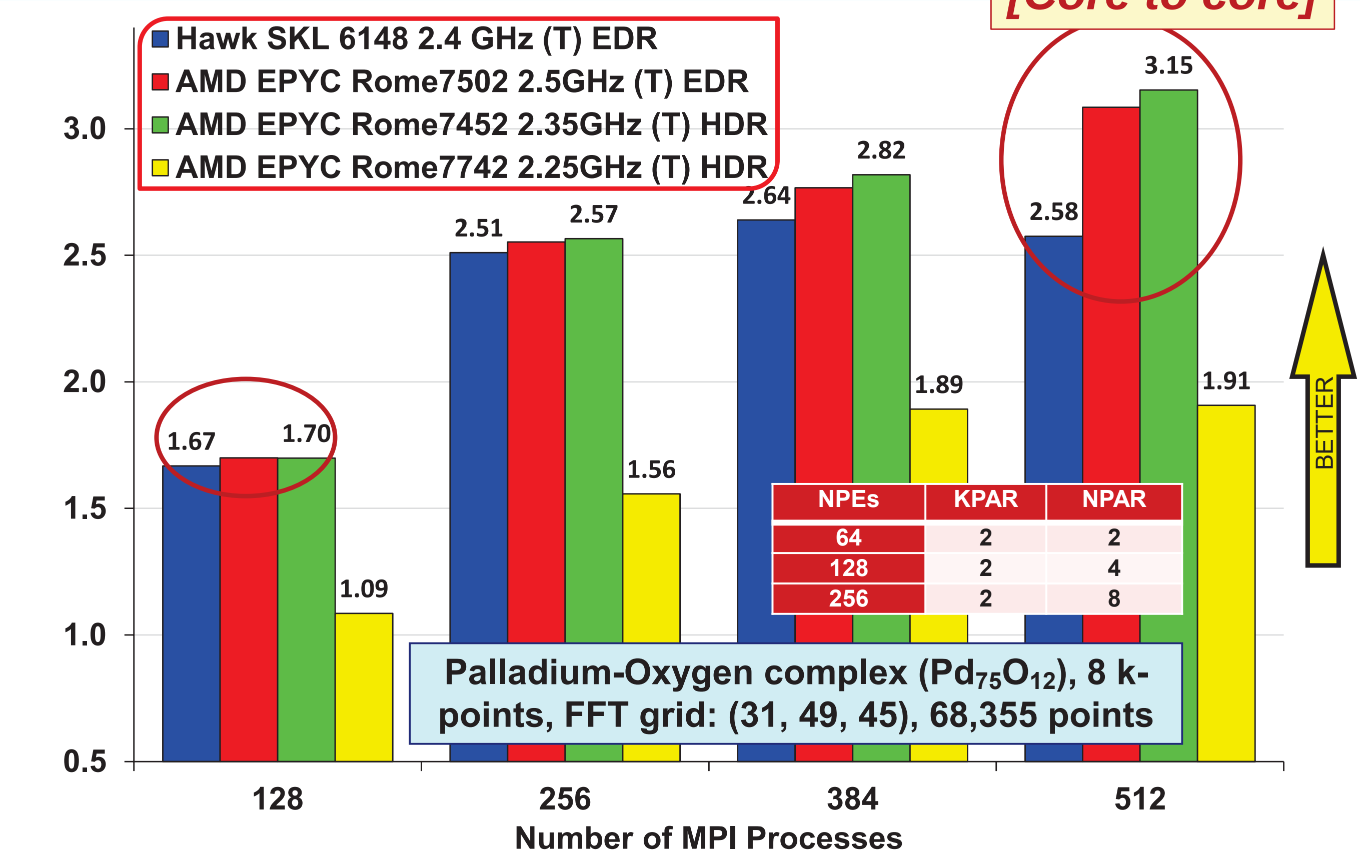
DL_POLY 4 – Gramicidin Simulation

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD



VASP 5.4.4 – Pd-O Benchmark - Parallelisation on k-points

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

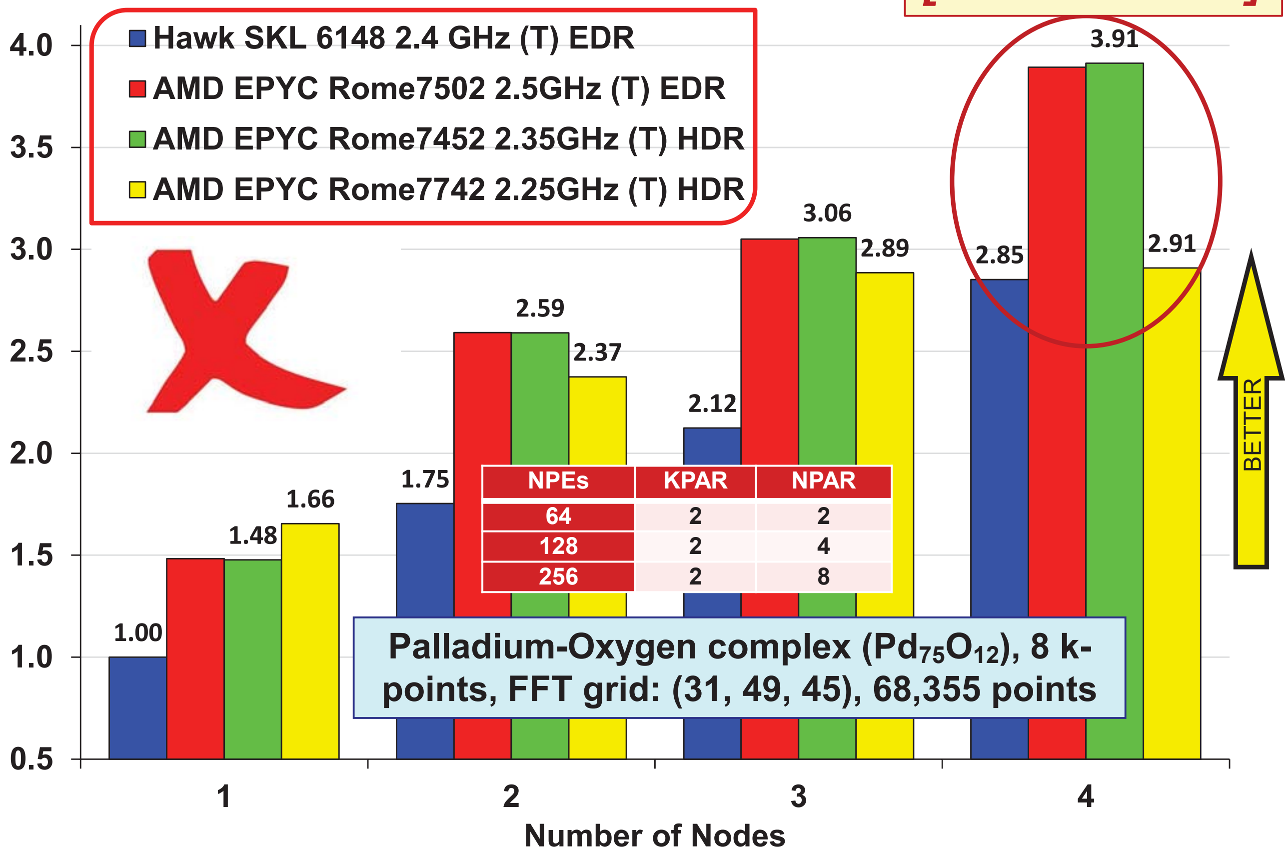


NPEs	KPAR	NPAR
64	2	2
128	2	4
256	2	8

VASP 5.4.4 – Pd-O Benchmark - Parallelisation on k-points

Performance *Relative to the Hawk SKL 6148 2.4 GHz (1 Node)*

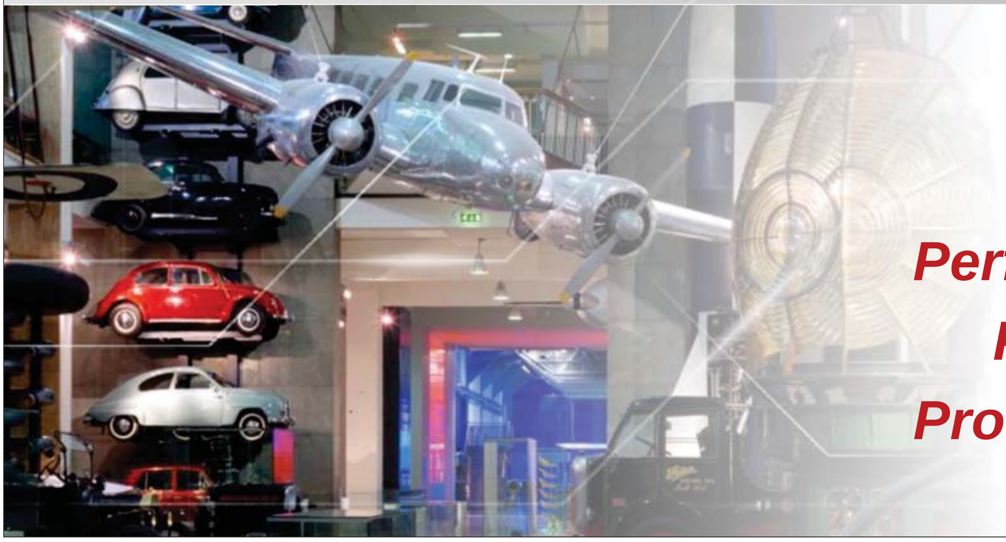
[Node to Node]



Rome Epyc 7002 family of processors from AMD

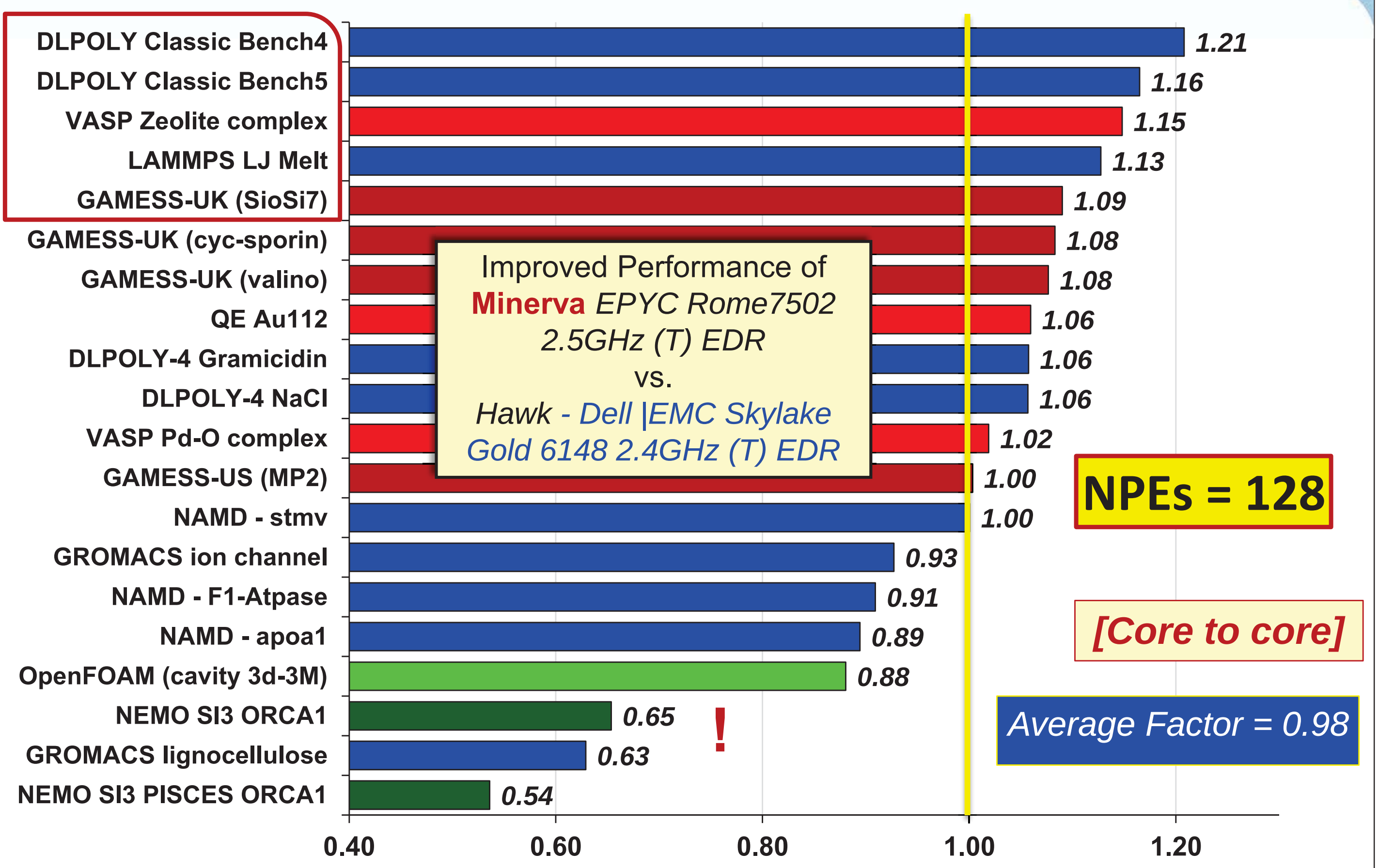
Epyc 7002 Model	Cores / Threads	Base Speed	Boost Speed	L3 Cache	TDP (Watts)	"Rome" List Price	Raw Clocks	$ / Raw Clocks	Rel Perf	$ / Rel Perf
7742	64 / 128	2.25 GHz	3.4 GHz	256 MB	225	$6,950	144.0 GHz	$48.26	24.40	$285
7702	64 / 128	2.0 GHz	3.35 GHz	256 MB	200	$6,450	128.0 GHz	$50.39	21.69	$297
7702P	64 / 128	2.0 GHz	3.35 GHz	256 MB	200	$4,425	128.0 GHz	$34.57	21.69	$204
7642	48 / 96	2.3 GHz	3.3 GHz	256 MB	225	$4,775	110.4 GHz	$43.25	18.70	$255
7552	48 / 96	2.2 GHz	3.3 GHz	192 MB	200	$4,025	105.6 GHz	$38.12	17.89	$225
7542	32 / 64	2.9 GHz	3.4 GHz	128 MB	225	$3,400	92.8 GHz	$36.64	15.72	$216
7502	32 / 64	2.5 GHz	3.35 GHz	128 MB	200	$2,600	80.0 GHz	$32.50	13.55	$192
7502P	32 / 64	2.5 GHz	3.35 GHz	128 MB	200	$2,300	80.0 GHz	$28.75	13.55	$170
7452	32 / 64	2.35 GHz	3.35 GHz	128 MB	155	$2,025	75.2 GHz	$26.93	12.74	$159
7402	24 / 48	2.8 GHz	3.35 GHz	128 MB	155	$1,783	67.2 GHz	$26.53	11.39	$157
7402P	24 / 48	2.8 GHz	3.35 GHz	128 MB	155	$1,250	67.2 GHz	$18.60	11.39	$110
7352	24 / 48	2.3 GHz	3.2 GHz	128 MB	180	$1,350	55.2 GHz	$24.46	9.35	$144
7302	16 / 32	3.0 GHz	3.3 GHz	128 MB	155	$978	48.0 GHz	$20.38	8.13	$120
7302P	16 / 32	3.0 GHz	3.3 GHz	128 MB	155	$825	48.0 GHz	$17.19	8.13	$101
7282	16 / 32	2.8 GHz	3.2 GHz	64 MB	120	$650	44.8 GHz	$14.51	7.59	$86
7272	12 / 24	2.9 GHz	3.2 GHz	64 MB	155	$625	34.8 GHz	$17.96	5.90	$106
7262	8 / 16	3.2 GHz	3.4 GHz	128 MB	120	$575	25.6 GHz	$22.46	4.34	$133
7252	8 / 16	3.1 GHz	3.2 GHz	64 MB	120	$475	24.8 GHz	$19.15	4.20	$113
7232P	8 / 16	3.1 GHz	3.2 GHz	32 MB	120	$450	24.8 GHz	$18.15	4.20	$107

Performance Analysis of the AMD EPYC Rome Processors

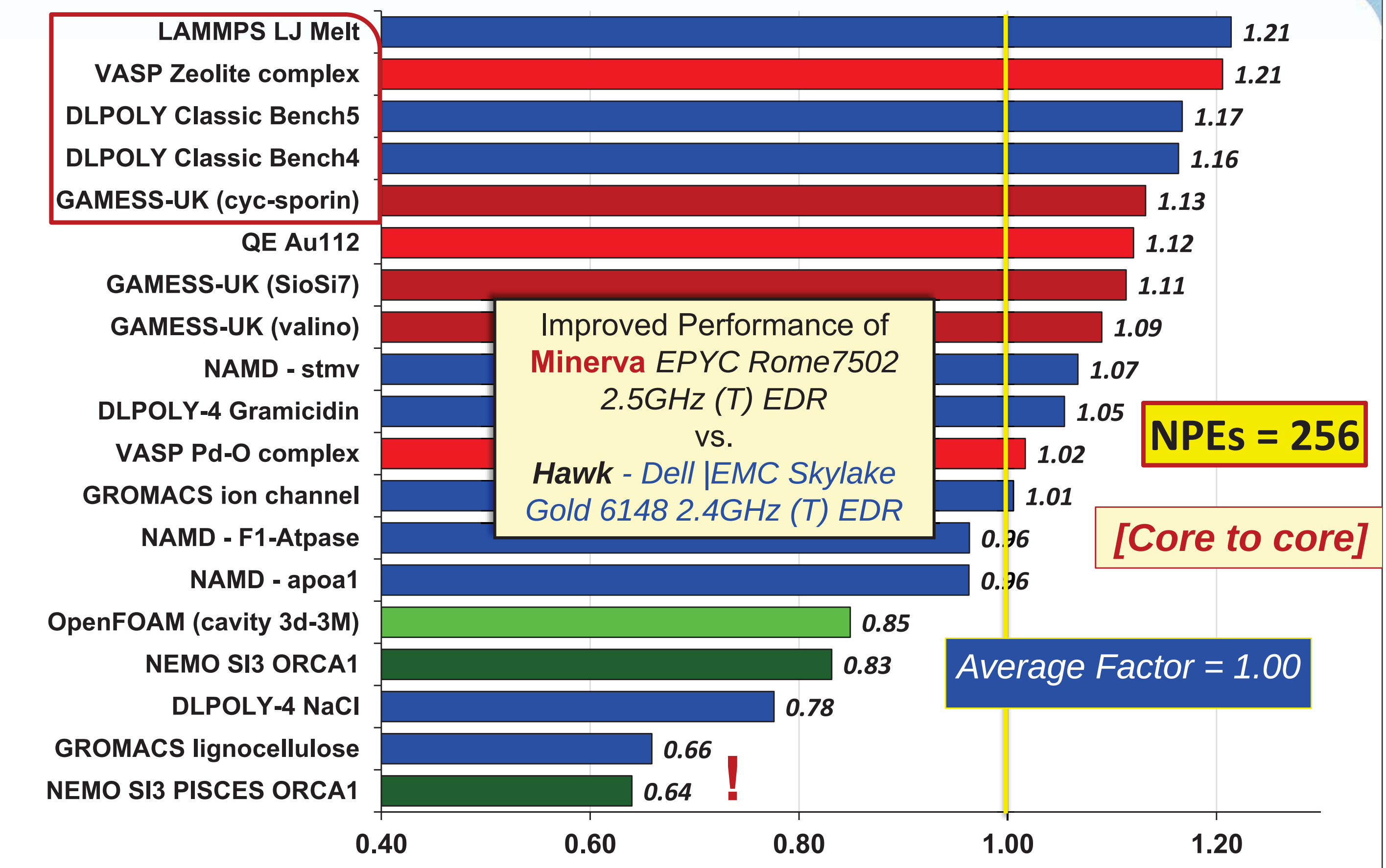


Relative Performance as a Function of Processor Family

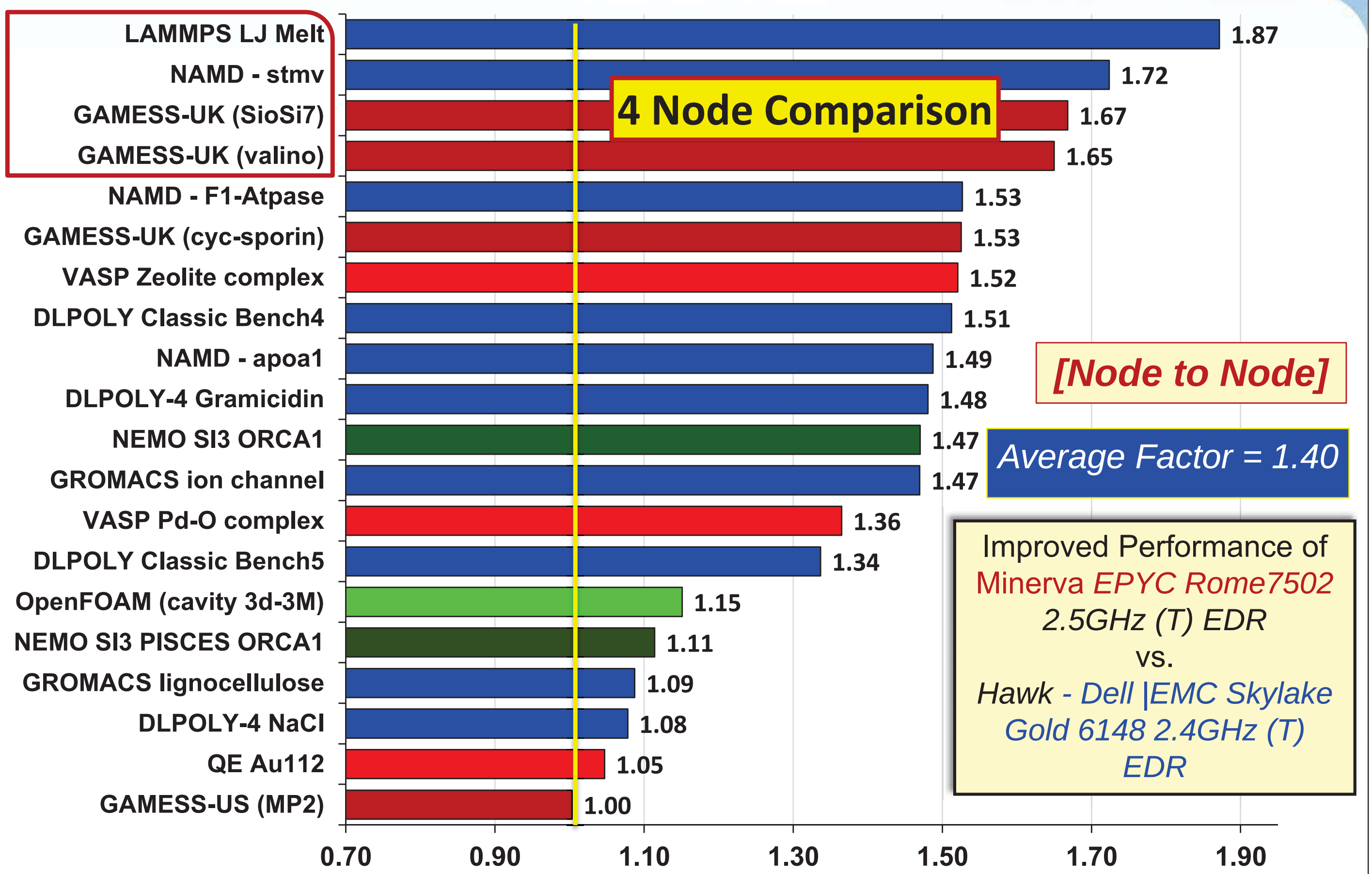
EPYC Rome 7502 2.5GHz (T) EDR vs. SKL "Gold" 6148 2.4 GHz EDR



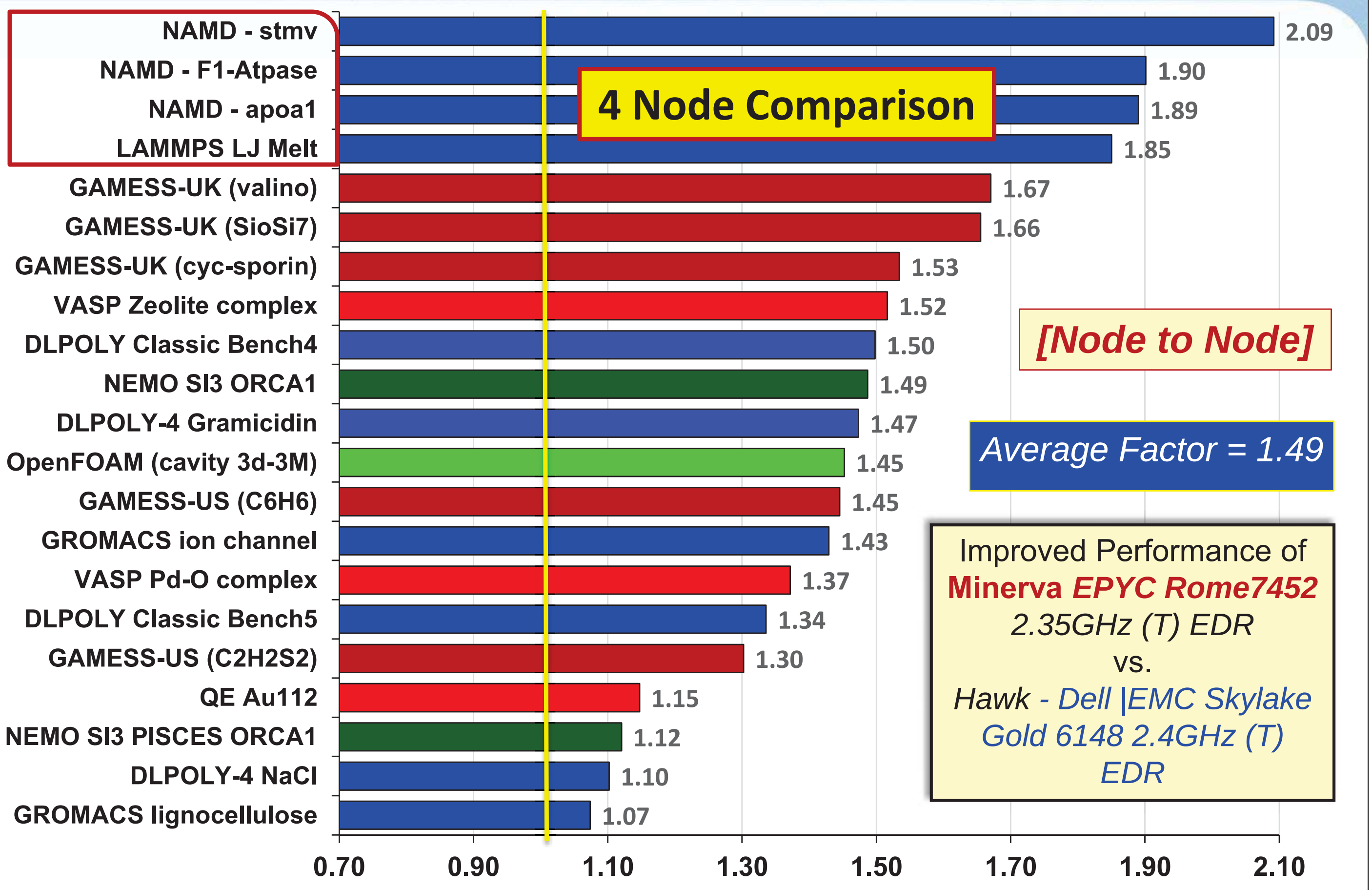
EPYC Rome 7502 2.5GHz (T) EDR vs. SKL “Gold” 6148 2.4 GHz EDR



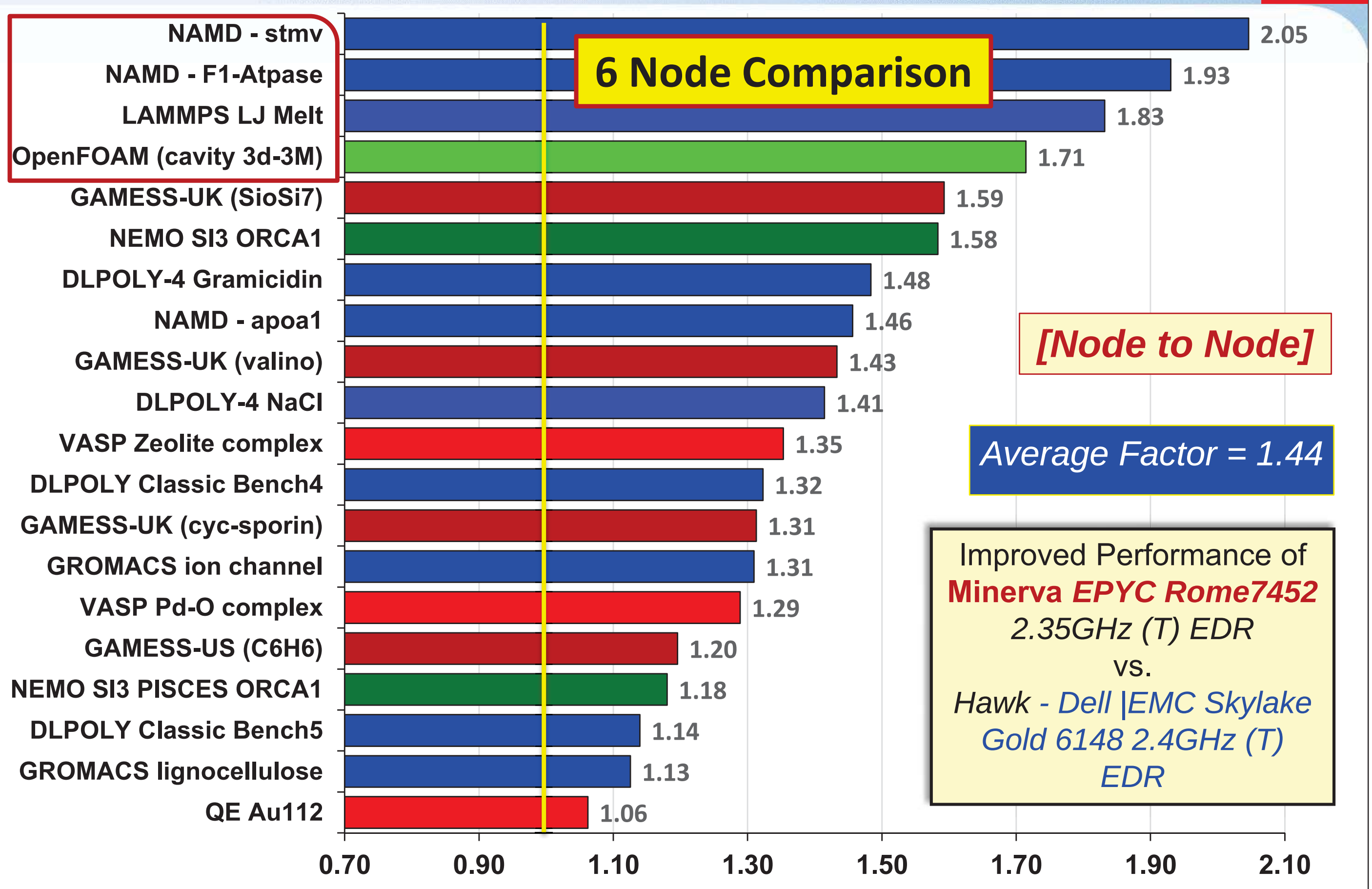
EPYC Rome 7502 2.5GHz (T) EDR vs. SKL “Gold” 6148 2.4 GHz EDR



EPYC Rome 7452 2.35GHz (T) EDR vs. SKL "Gold" 6148 2.4 GHz EDR



EPYC Rome 7452 2.35GHz (T) EDR vs. SKL "Gold" 6148 2.4 GHz EDR



Acknowledgements

- **Joshua Weage**, Dave Coughlin, Derek Rattansey, Steve Smith, Gilles Civario and Christopher Huggins for access to, and assistance with, the variety of EPYC SKUs at the Dell Benchmarking Centre.
- **Martin Hilgeman** for informative discussions and access to, and assistance with, the variety of EPYC SKUs comprising the Daytona cluster at the AMD Benchmarking Centre.
- *Ludovic Sauge*, **Enguerrand Petit** and Martyn Foster (Bull/ATOS) for informative discussions and access in 2018 to the Skylake & AMD EPYC Naples clusters at the Bull HPC Competency Centre.
- *David Cho, Colin Bridger, Ophir Maor & Steve Davey for access to the "Daytona_X" AMD 7742 cluster at the HPC Advisory Council.*

Summary

- Focus on systems featuring the **current high-end processors from AMD** (EPYC Rome SKUs – the 7502, 7452, 7702, 7742 etc.).
  - Baseline clusters include the Sandy Bridge e5-2670 system (Raven), and the recent Skylake (SKL) system, the **Gold 6148/2.4 GHz** cluster, at Cardiff University.
  - Major focus on two AMD EPYC Rome clusters featuring the 32-core **7502 2.5GHz** and **7452 2.35 GHz**.
- Considered performance of both synthetic and end-user applications. Latter include molecular simulation (**DL_POLY, LAMMPS, NAMD, Gromacs**), electronic structure (**GAMESS-UK & GAMESS-US**), materials modelling (**VASP**, **Quantum Espresso**) Engineering (**OpenFOAM**) plus the **NEMO** code (Ocean General Circulation Model) [Seven in Archer Top-30 Ranking list].
- Consideration given to scalability by **processing elements (cores)** and by **nodes** (guided by ARM Performance Reports).

Summary – Core-to-Core Comparisons

1. A ***Core-to-Core comparison*** across 20 data sets (11 applications) suggests on average that the Rome 7452 and 7502 perform on a par with the Skylake Gold (SKL) 6148/2.4 GHz.
   – **Comparable performance** averaged across a basket of codes and associated data sets when comparing the Skylake "Gold" 6148 cluster (EDR) to the AMD Rome 32 core SKUs. Thus on 128 cores, the 7502 exhibits **98% of the SKL** performance on 128 cores and **100%** (i.e. the same) performance on 256 cores.
2. Relative performance sensitive to the effective use of the AVX vector instructions.
3. Applications with low utilisation of AVX-512 leads to weaker performance of the Skylake CPUs and **better performance on the Rome-based clusters** e.g. DLPOLY, NAMD and LAMMPS.
4. A number of applications with heavy memory B/W demands perform poorly on the AMD systems e.g. NEMO. A few spurious examples e.g. Gromacs (Lignocellulose)

Summary – Node-to-Node Comparisons

1. Given comparable core performance, a ***Node-to-Node comparison*** typical of the performance when running a workload shows the Rome AMD 7452 and 7502 delivering **superior performance** compared to (i) the SKL Gold 6148 performance (64 cores vs. 40 cores), and (ii) the 64-core 7742 AMD processor.
2. Thus a **4-node benchmark** (*256 ×* ***AMD 7452 2.35 GHz*** *cores*) based on examples from 11 applications and 21 data sets show an average improvement factor of **1.49** compared to the corresponding 4 node runs (*160 cores*) on the Hawk SKL Gold 6148/2.4 GHz.
3. This factor is reduced somewhat, to **1.40** based on the **4-node** ***AMD 7502 2.5 GHz core*** benchmarks. Impact of the HDR interconnect on the 7452 cluster, or less than optimal 7502 nodes?
4. Slight reduction in improvement factor when running on **6-nodes** of the ***AMD 7452 2.35 GHz***, with an averaged factor of ***1.44*** comparing 240 SKL cores to 384 AMD Rome cores.
5. In **all applications** the **AMD Rome systems** outperform the corresponding **Skylake Gold 6148** system based on a **node-to-node comparison**.

Any Questions?

Martyn Guest 029-208-79319

Christine Kitchen 029-208-70455

Jose Munoz 029-208-70626

CIUK 2019 KEYNOTE PRESENTATION

Debora Sijacki

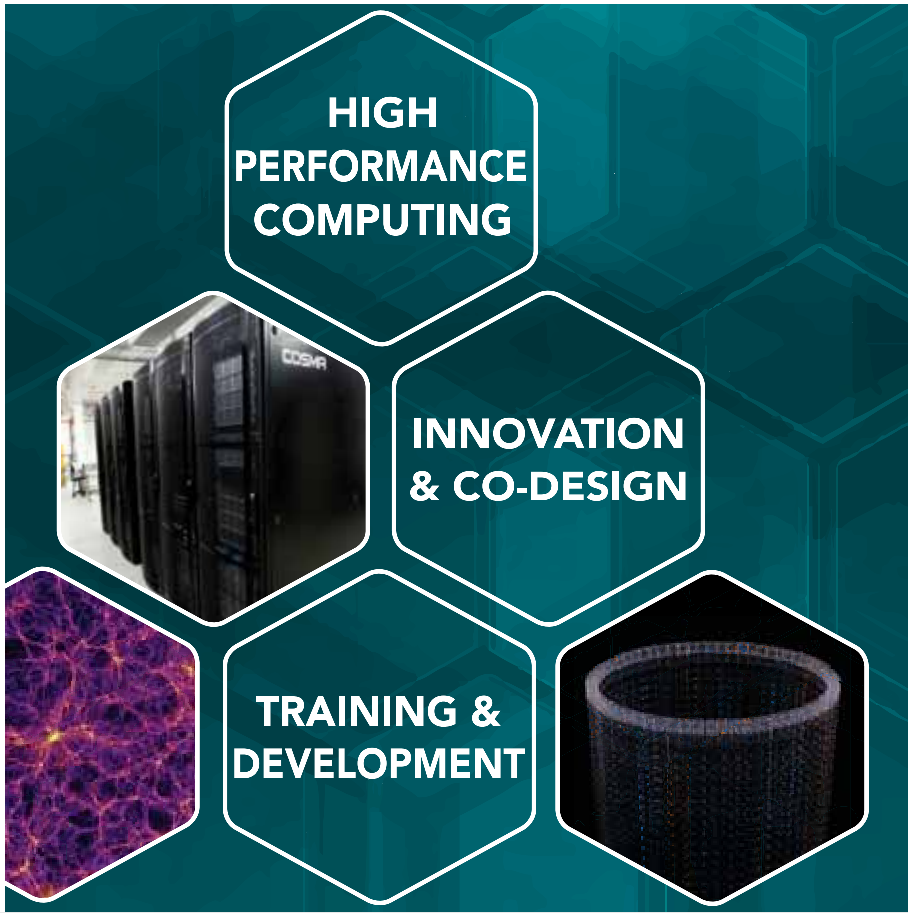
University of Cambridge, Institute of Astronomy

Winner of the 2019 PRACE Ada Lovelace Award for HPC

Towards Next Generation Computing in Cosmological Simulations: Prospects and Challenges

Please Note: due to a personal emergency Debora Sijacki was unable to travel and this presentation was given by her colleague Mark Wilkinson (DiRAC) on her behalf.

DiRAC



DiRAC

Diverse science cases require heterogenous architectures

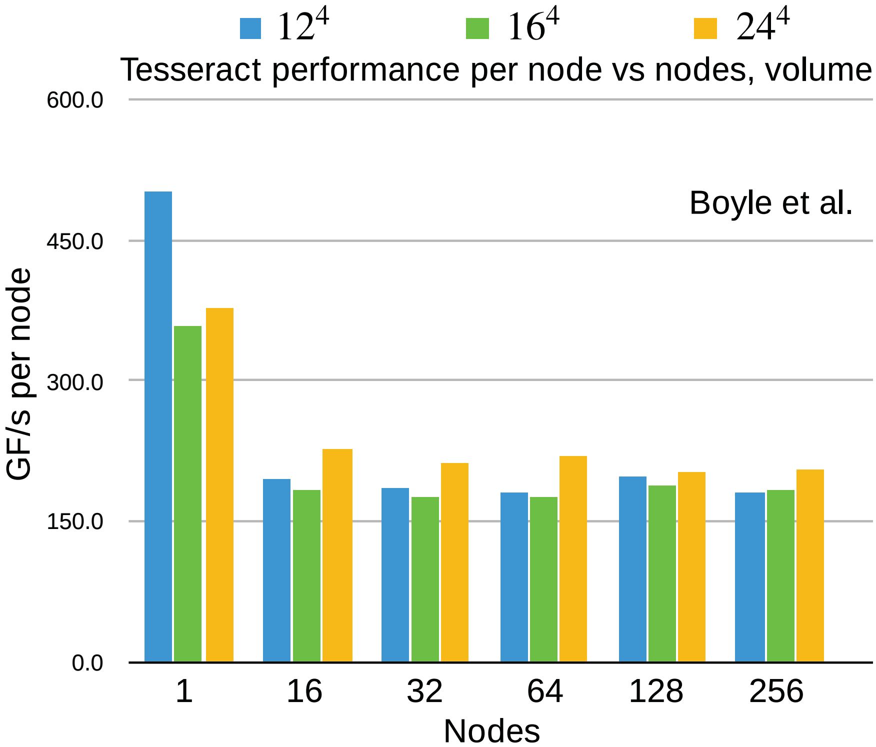
Extreme Scaling "Tesseract" (Edinburgh)	Data Intensive "DIaL" and "CSD3" (Leicester & Cambridge)	Memory Intensive "COSMA" (Durham)
		
2 Pflop/s to support largest lattice-QCD simulations	Heterogeneous architecture to support complex simulation and modelling workflows	230 TB RAM to support largest cosmological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

DiRAC

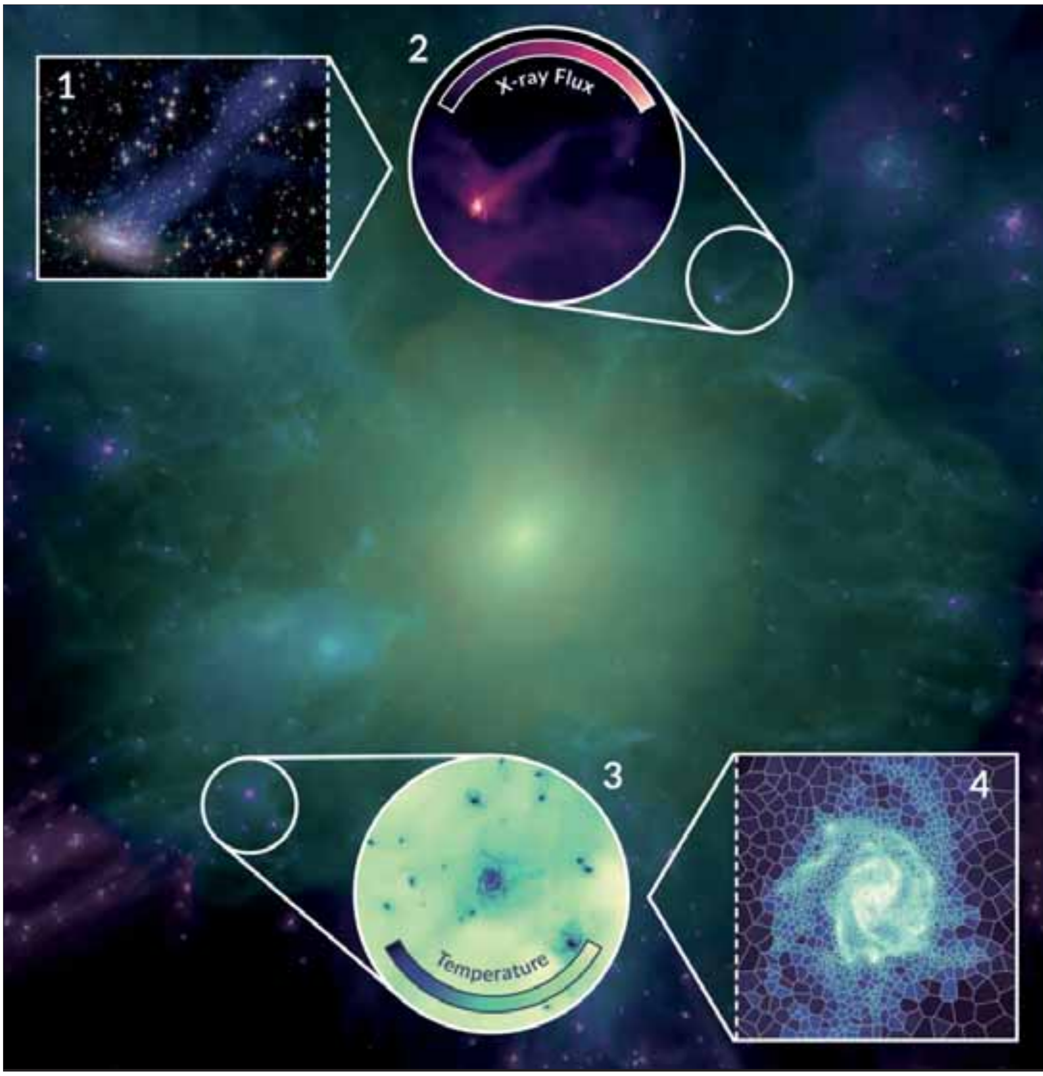
Extreme Scaling: Tesseract

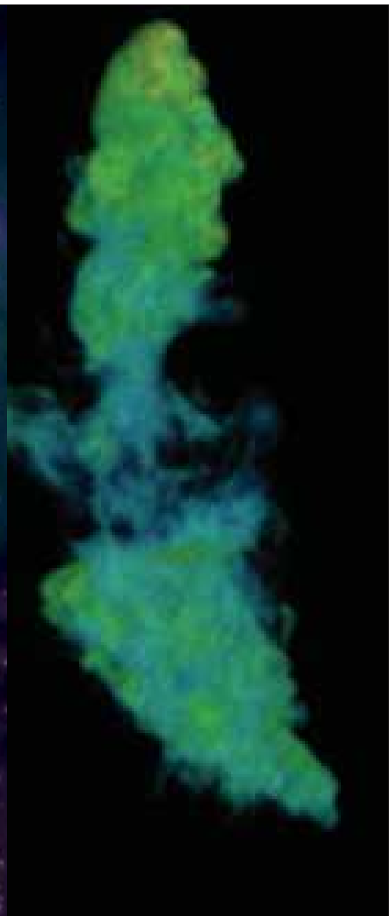
- Example of co-design in action - matching application to network topology
- Embed 2^n QCD torus inside hypercube so that nearest neighbour comms travel single hop: 4x speed up over default MPI Cartesian communicators on large systems

$\Longrightarrow$ 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





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

Jonas Markussen

Dolphin Interconnect Solutions

NVMe Over PCIe Fabrics Using Device Lending

Jonas Markussen is an R&D software engineer at Dolphin Interconnect Solutions. His work is focused on new applications for PCI Express clustering. Jonas is also currently working on his PhD, where his research interests are distributed shared-memory applications, computer networks and cluster interconnects.

Abstract

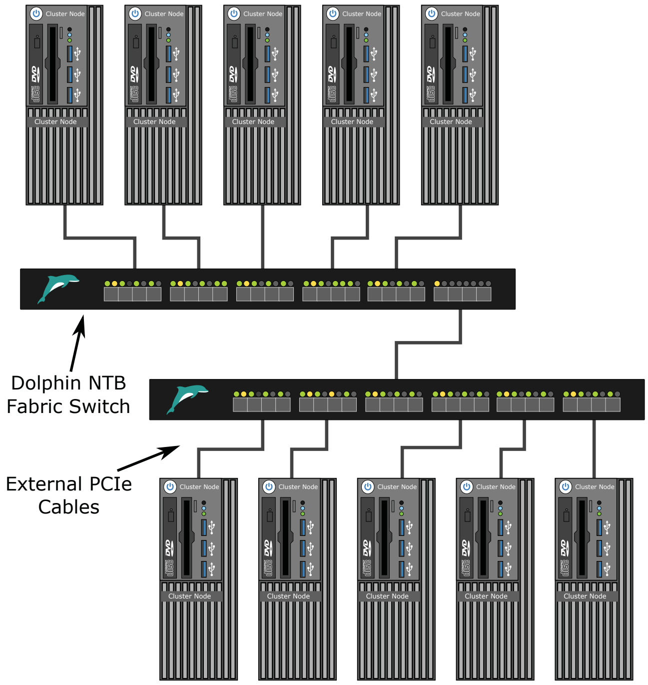
The emerging standard for accessing NVMe drives over a network today is NVMe over Fabrics (NVMe-oF). By relying on RDMA protocols, NVMe-oF is able to provide access to remote storage devices with very little overhead. However, encapsulating I/O commands and forwarding them over a network transport has an unavoidable latency cost compared to accessing a local device.

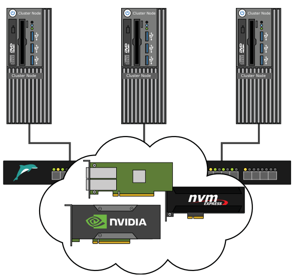
Using a mechanism called device lending, Dolphin's SmartIO technology allows a local system to access a remote NVMe drive as if it was locally installed. Device lending uses the inherent memory-mapping capabilities in PCIe to decouple NVMe drives and other PCIe devices from the hosts they physically reside in. This allows data to be moved using native DMA without relying on RDMA.

This talk will demonstrate how device lending is used to share NVMe drives in a PCIeinterconnected cluster. We compare the performance of our approach to state-of-the-art NVMe-oF RDMA solutions. We will also show examples of advanced use-cases, such as direct access to NVMe drives from GPUs, and demonstrate how a user can create a composable infrastructure optimized for data flow using our SmartIO technology.

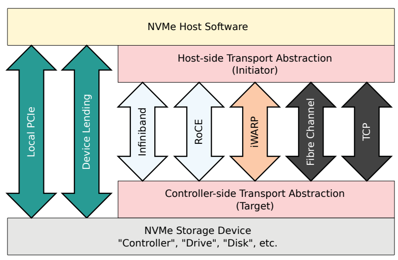
NVMe over PCI Express Fabrics using Device Lending

Jonas Markussen
Software Architect and PhD Student
jonas@dolphinics.com

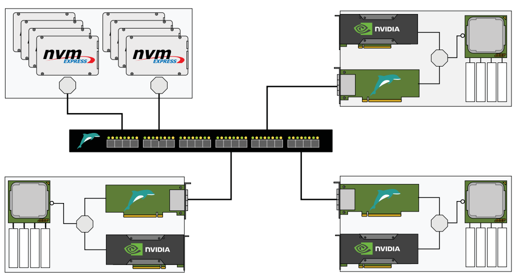




Resource pooling with
Device Lending



NVMe over Fabrics

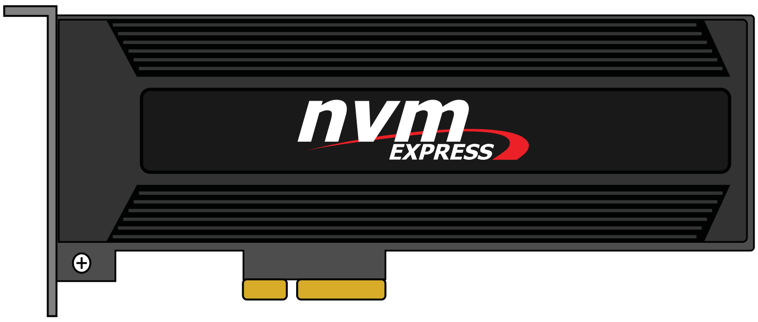


Flexible storage workflows

SSDs that are attached to the PCI Express (PCIe) bus use the Non-Volatile Memory Express (NVMe) interface standard

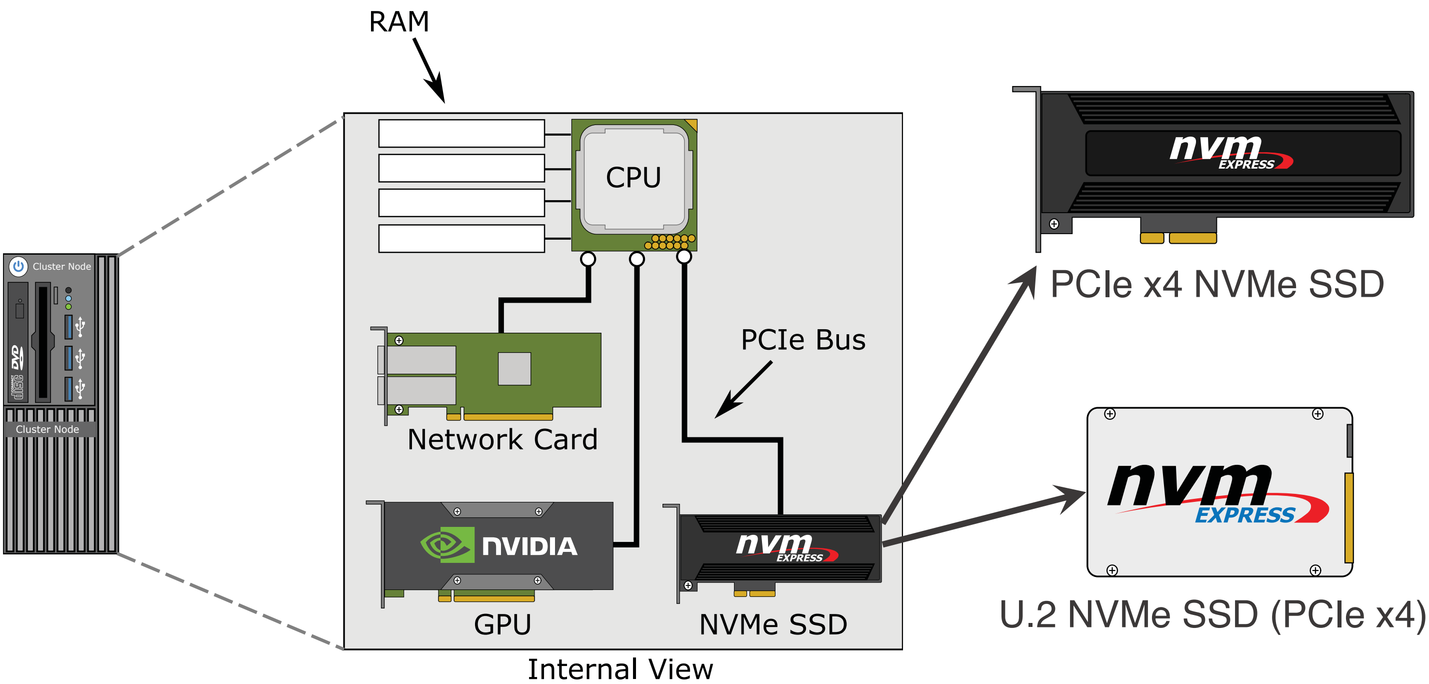
NVM **Express** = PCI **Express**

U.2 NVMe SSD (PCIe x4)

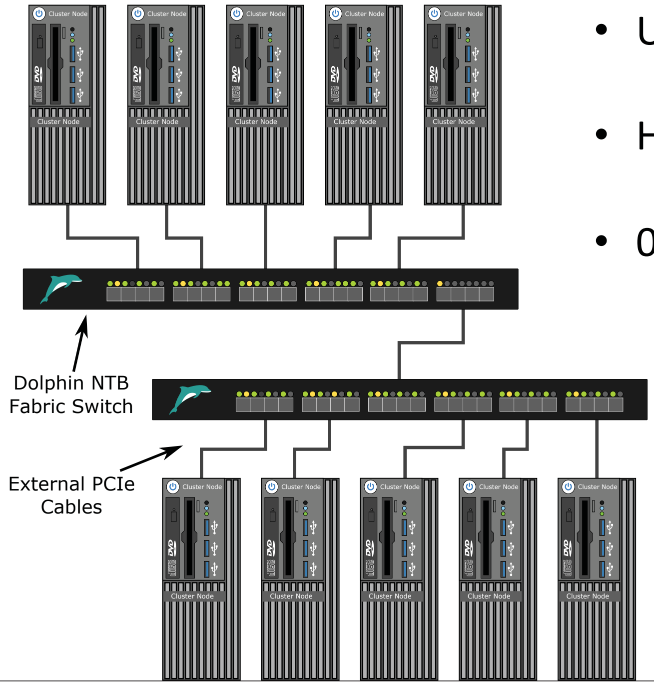


PCIe x4 NVMe SSD

PCIe is the most widely used standard for connecting I/O devices to a computer systems today

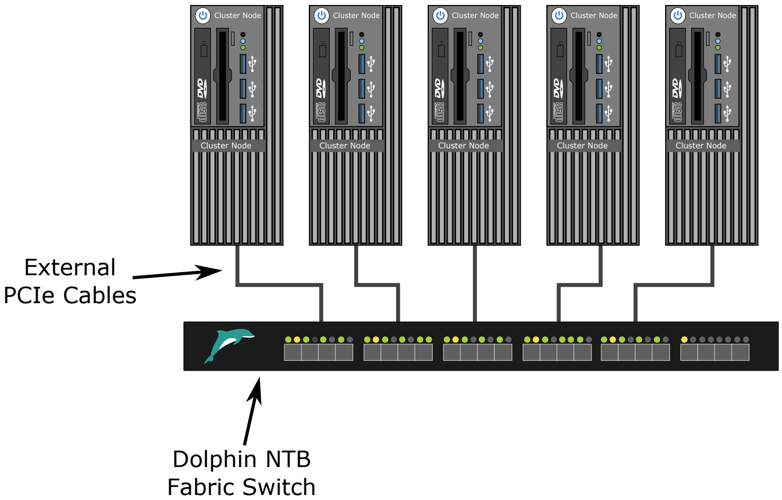


Dolphin makes hardware and software for creating PCIe-interconnected clusters for low latency shared-memory applications

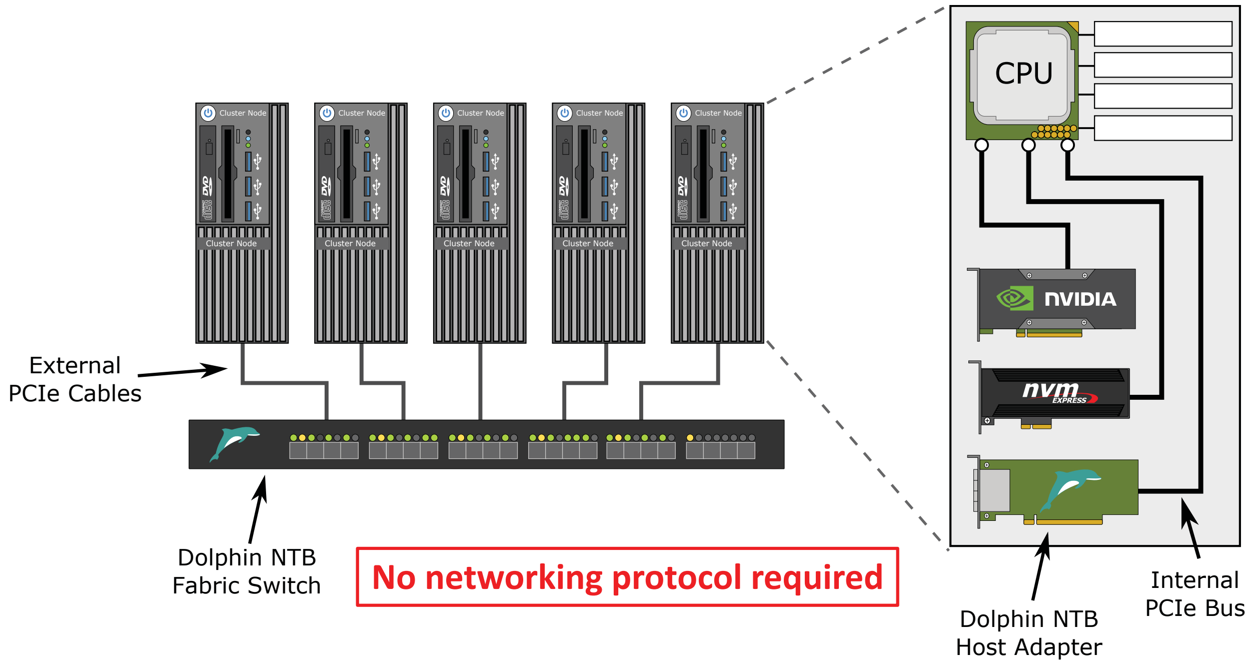


- Up to 60 heterogenous nodes
- Hundreds of nanoseconds **RAM-to-RAM**
- 0.5m up to 15m copper and 100m fibre

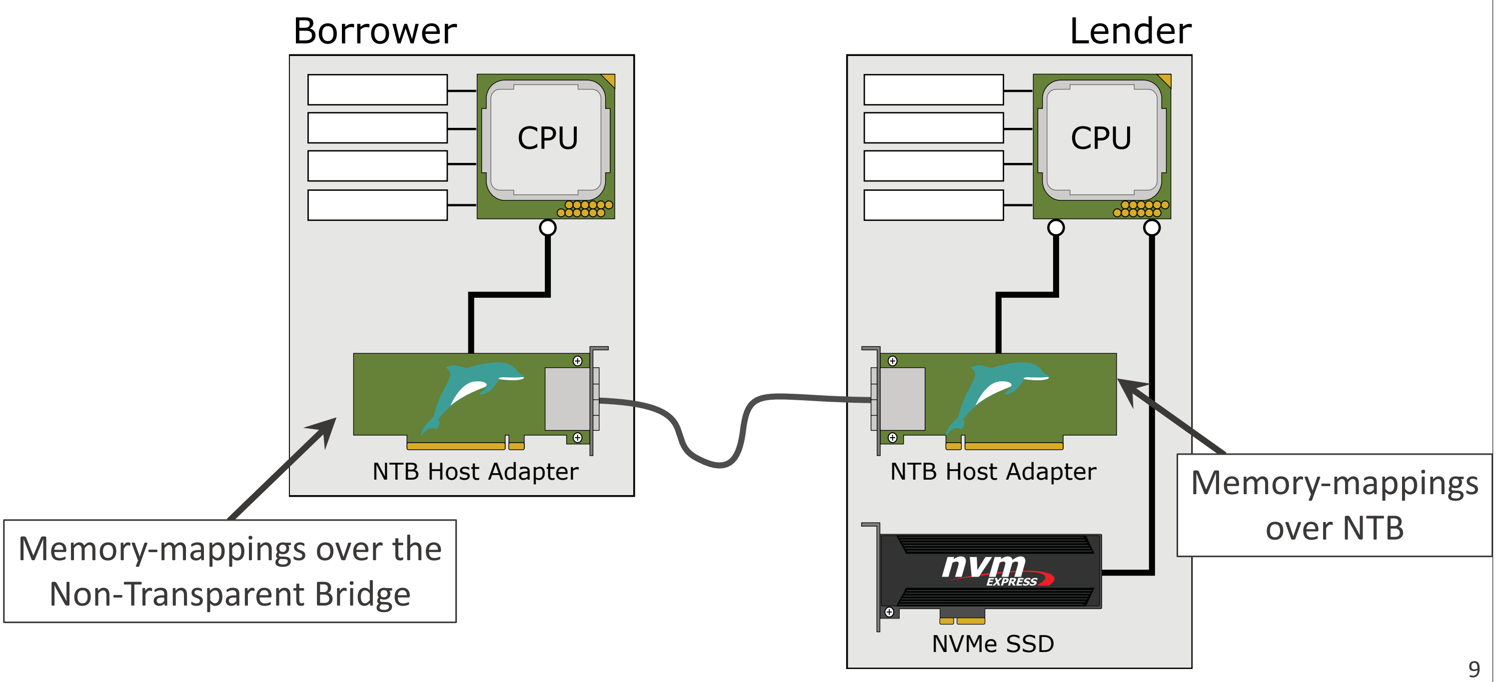
The same PCIe fabric is used both for interconnecting hosts and as the local I/O bus inside each host in PCIe clusters



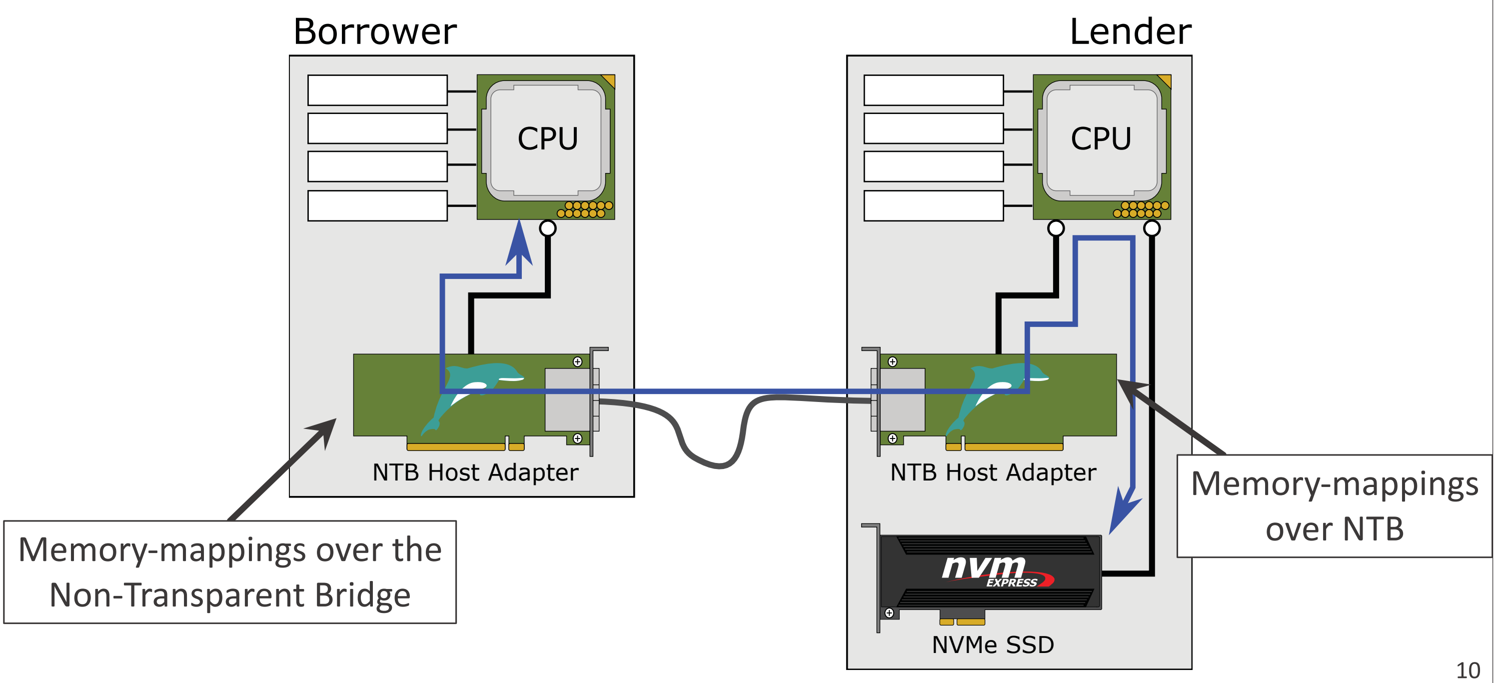
The same PCIe fabric is used both for interconnecting hosts and as the local I/O bus inside each host in PCIe clusters



Device Lending distributes local devices to remote hosts, allowing them to be accessed over native PCIe

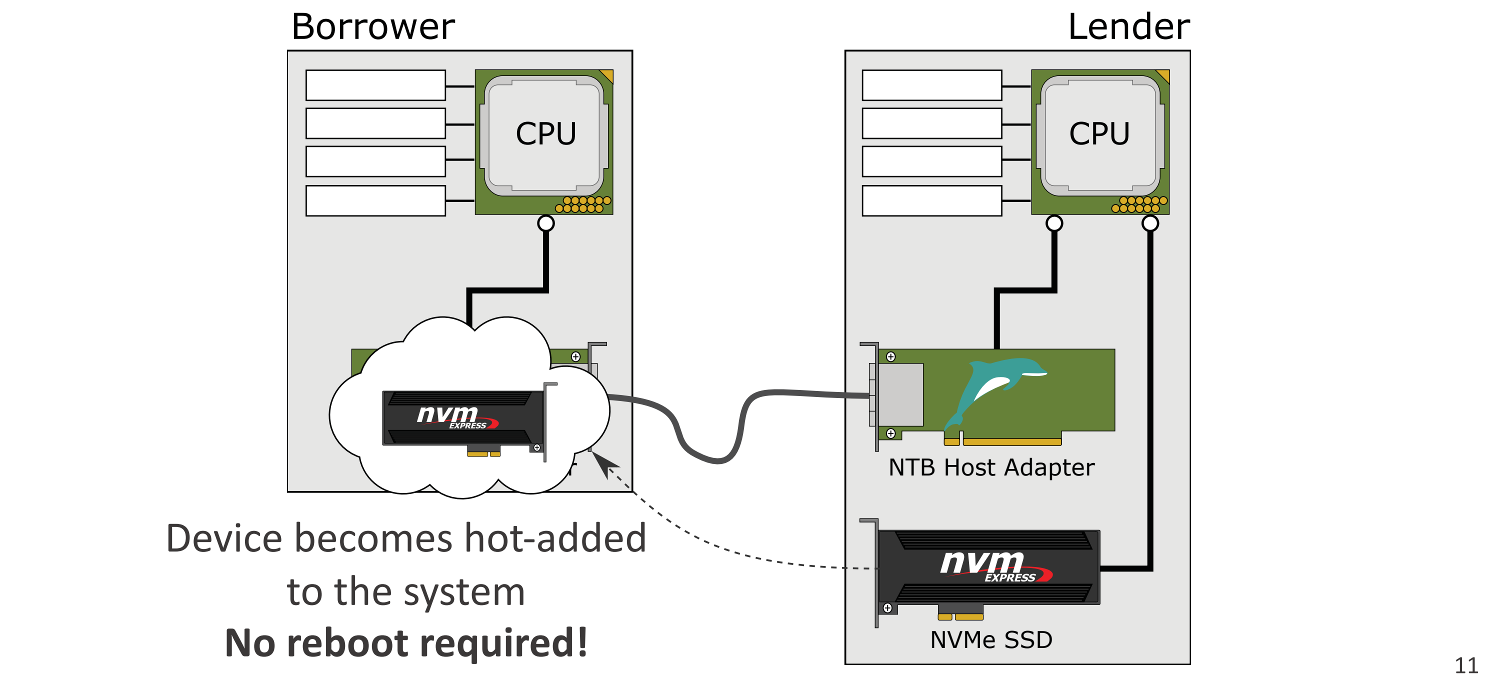


Device Lending distributes local devices to remote hosts, allowing them to be accessed over native PCIe

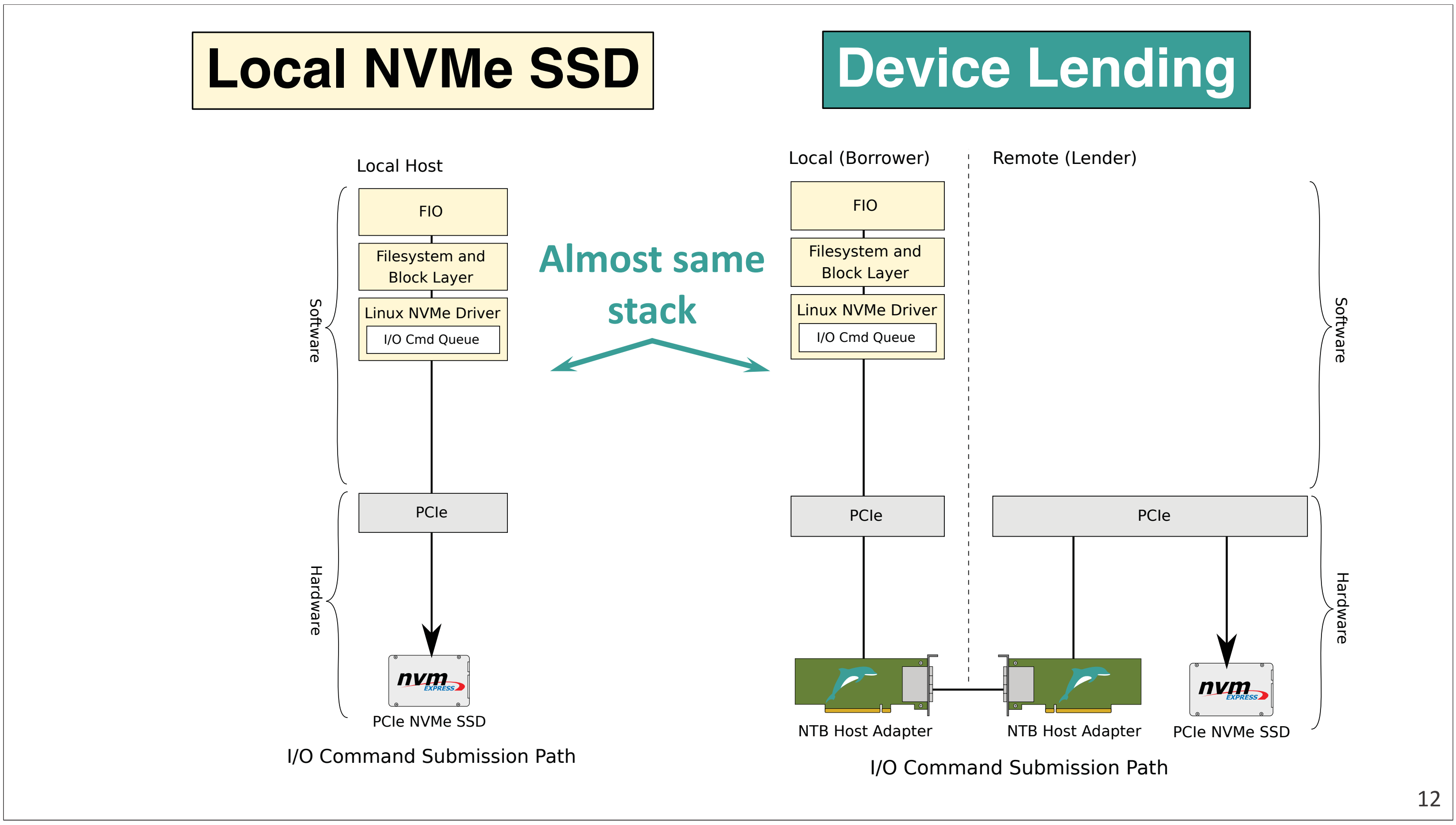


10

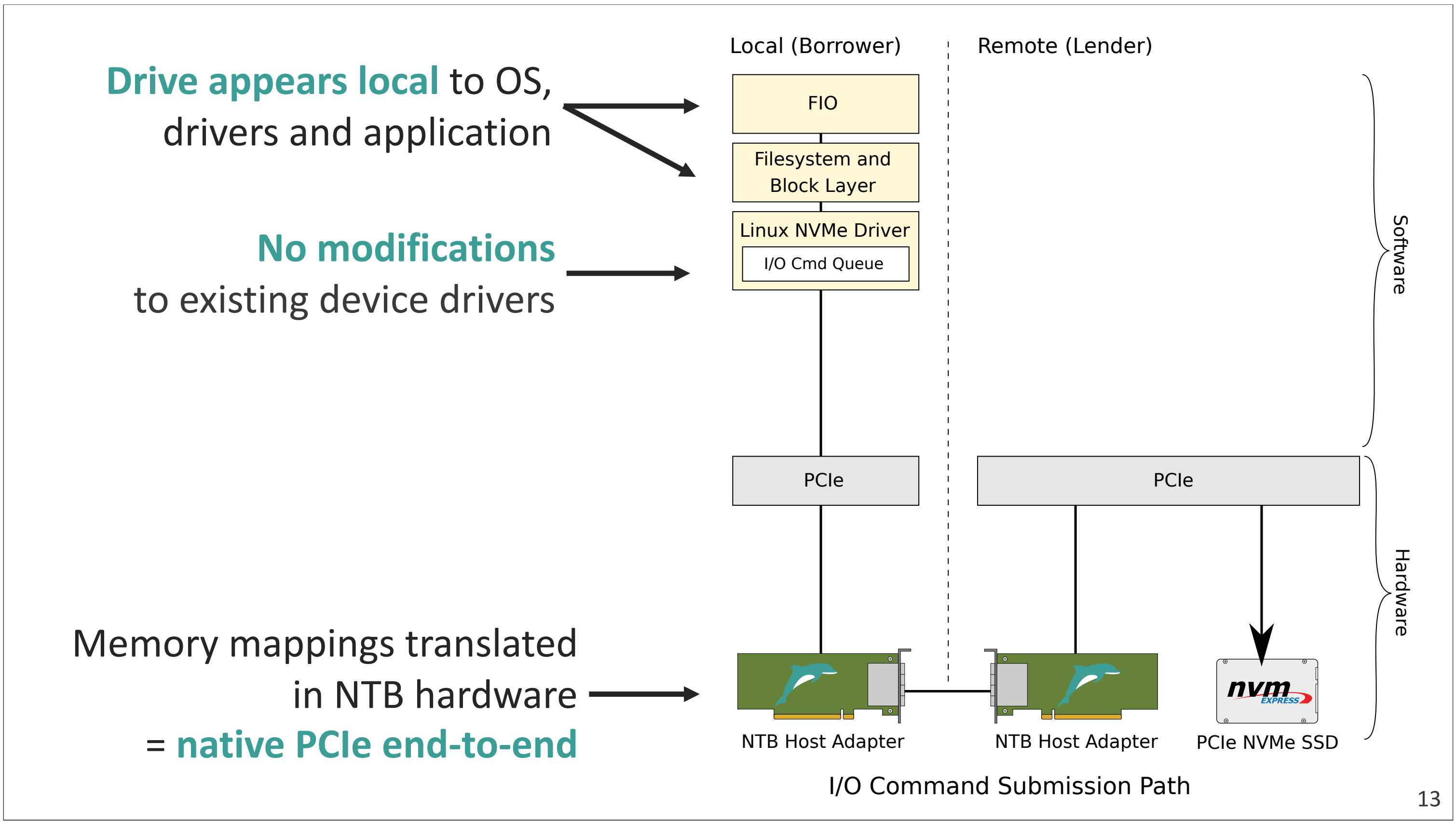
Device Lending distributes local devices to remote hosts, allowing them to be accessed over native PCIe



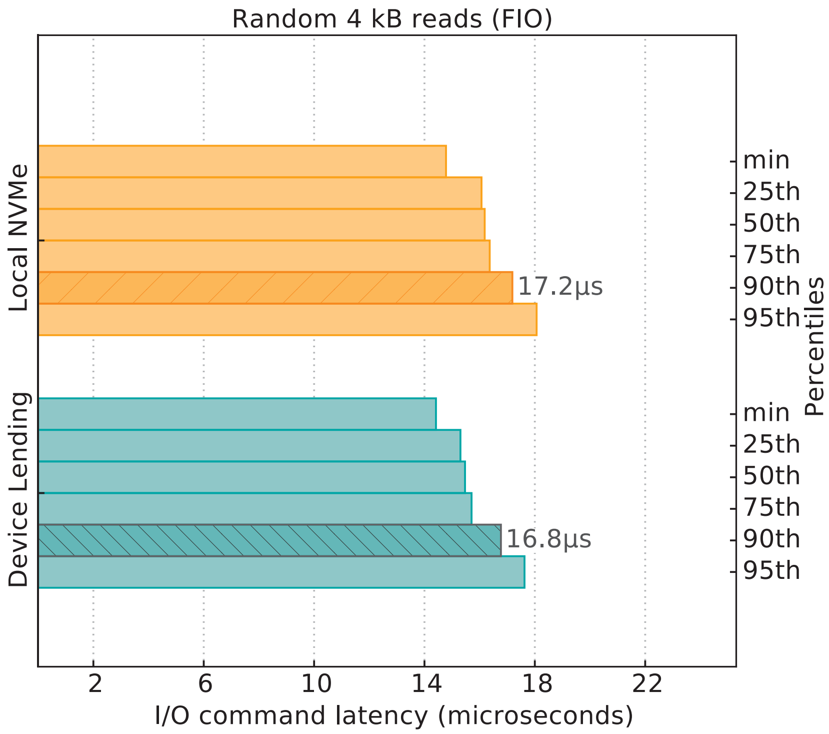
11



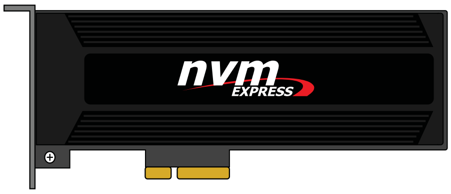
Local NVMe SSD
Device Lending
Local Host
FIO
Filesystem and Block Layer
Linux NVMe Driver
I/O Cmd Queue
Software
PCIe
Hardware
PCIe NVMe SSD
I/O Command Submission Path
Almost same stack
Local (Borrower)
Remote (Lender)
FIO
Filesystem and Block Layer
Linux NVMe Driver
I/O Cmd Queue
PCIe
PCIe
NTB Host Adapter
NTB Host Adapter
PCIe NVMe SSD
Software
Hardware
I/O Command Submission Path
12



Drive appears local to OS, drivers and application
No modifications to existing device drivers
Memory mappings translated in NTB hardware = native PCIe end-to-end
Local (Borrower)
Remote (Lender)
FIO
Filesystem and Block Layer
Linux NVMe Driver
I/O Cmd Queue
PCIe
PCIe
NTB Host Adapter
NTB Host Adapter
PCIe NVMe SSD
Software
Hardware
I/O Command Submission Path
13

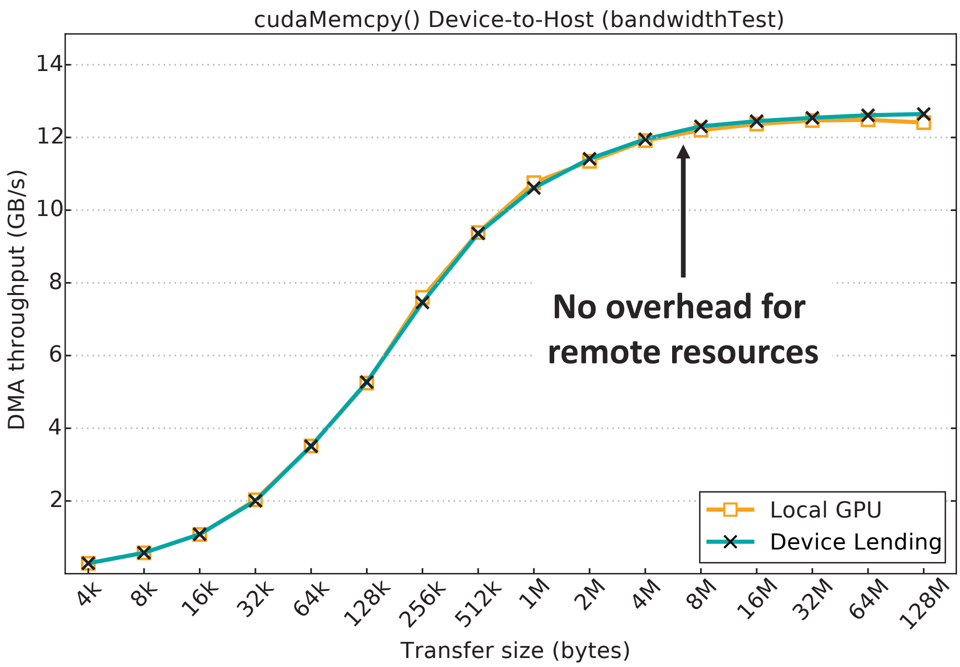


Local vs Remote

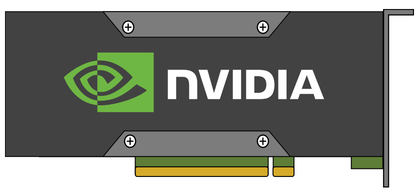


Intel Optane P4800X

Device Lending works for all standard PCIe devices, such as GPUs, NVMe drives, GPUs, FPGAs, and network cards

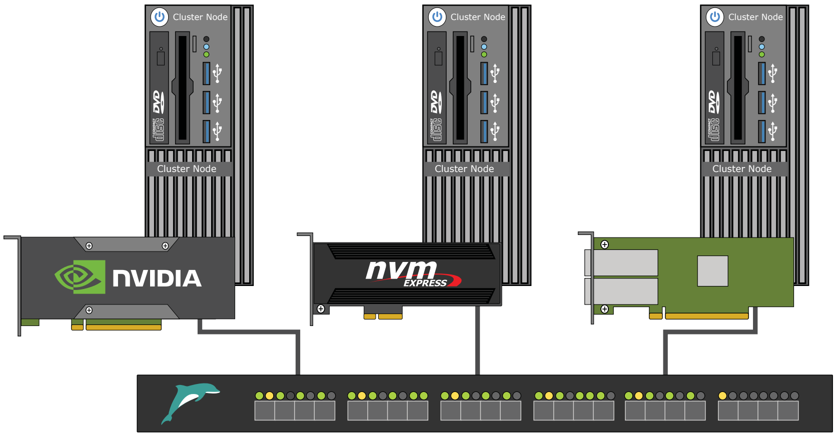


Local vs Remote

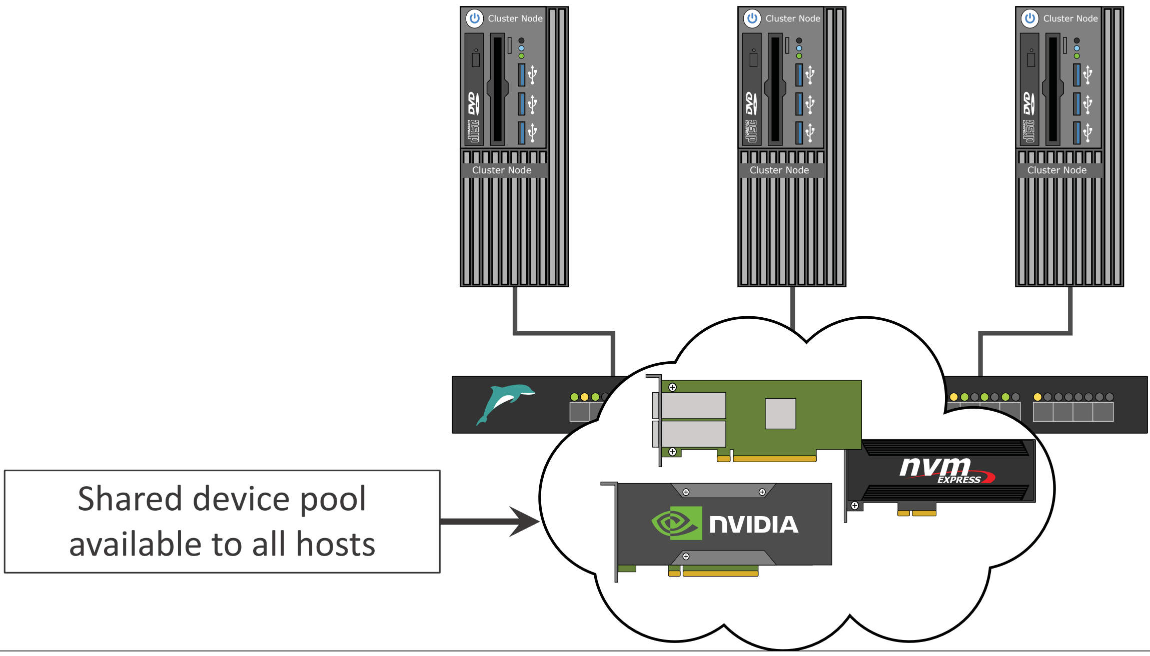


Quadro P4000

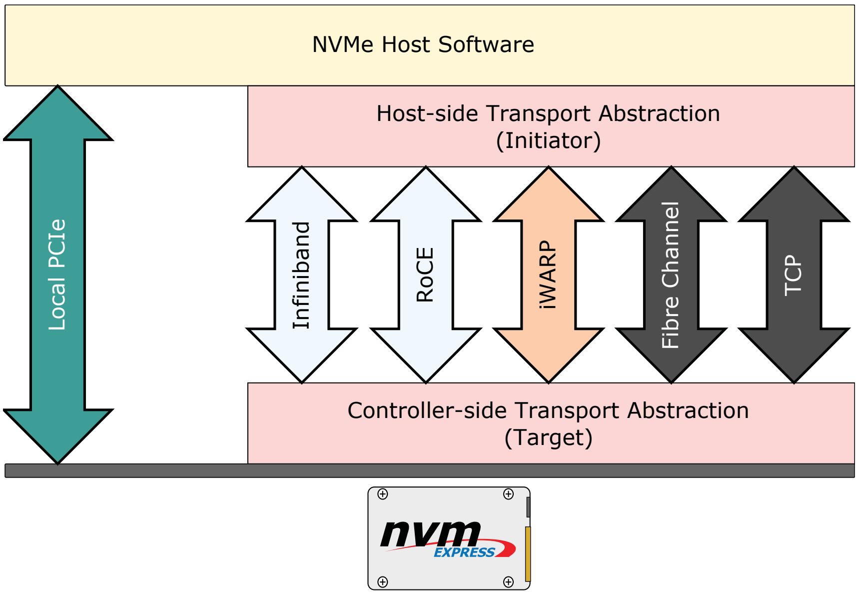
Hosts may lend away their local devices and borrow remote devices, pooling their resources and increasing utilization



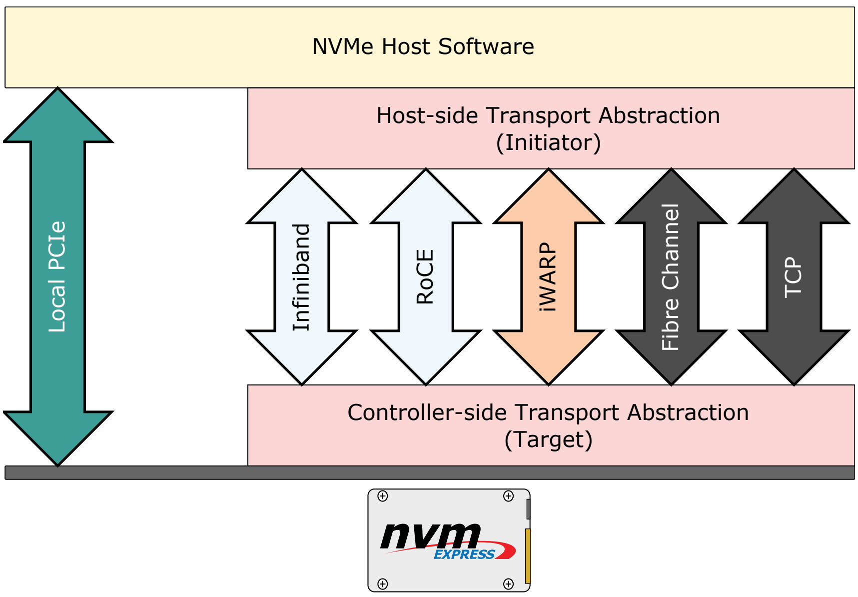
Hosts may lend away their local devices and borrow remote devices, pooling their resources and increasing utilization



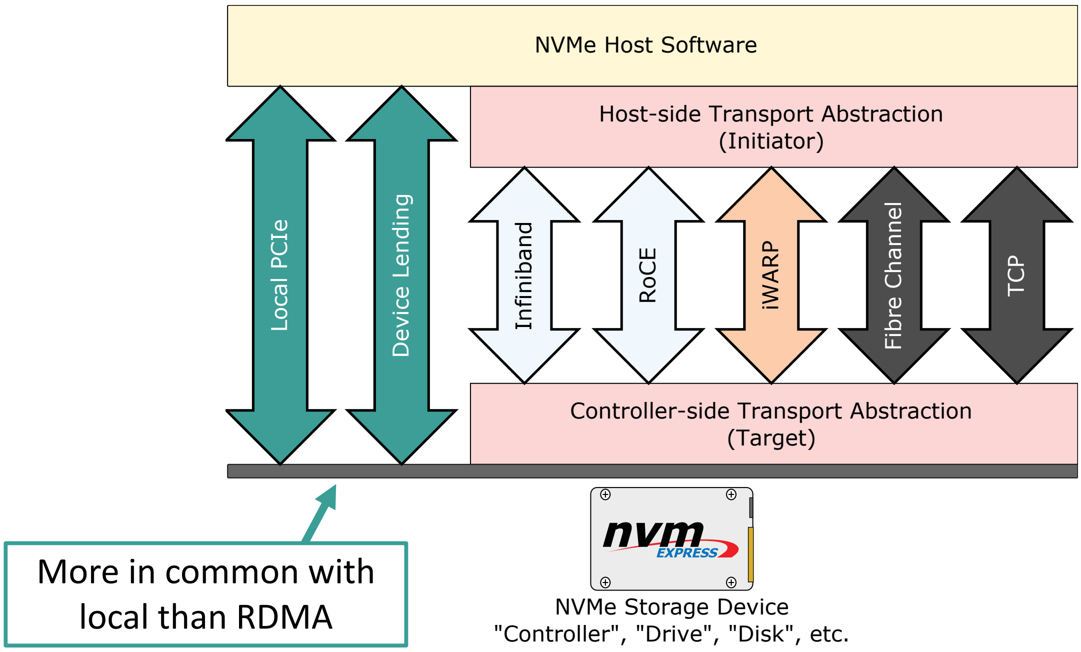
NVMe over Fabrics



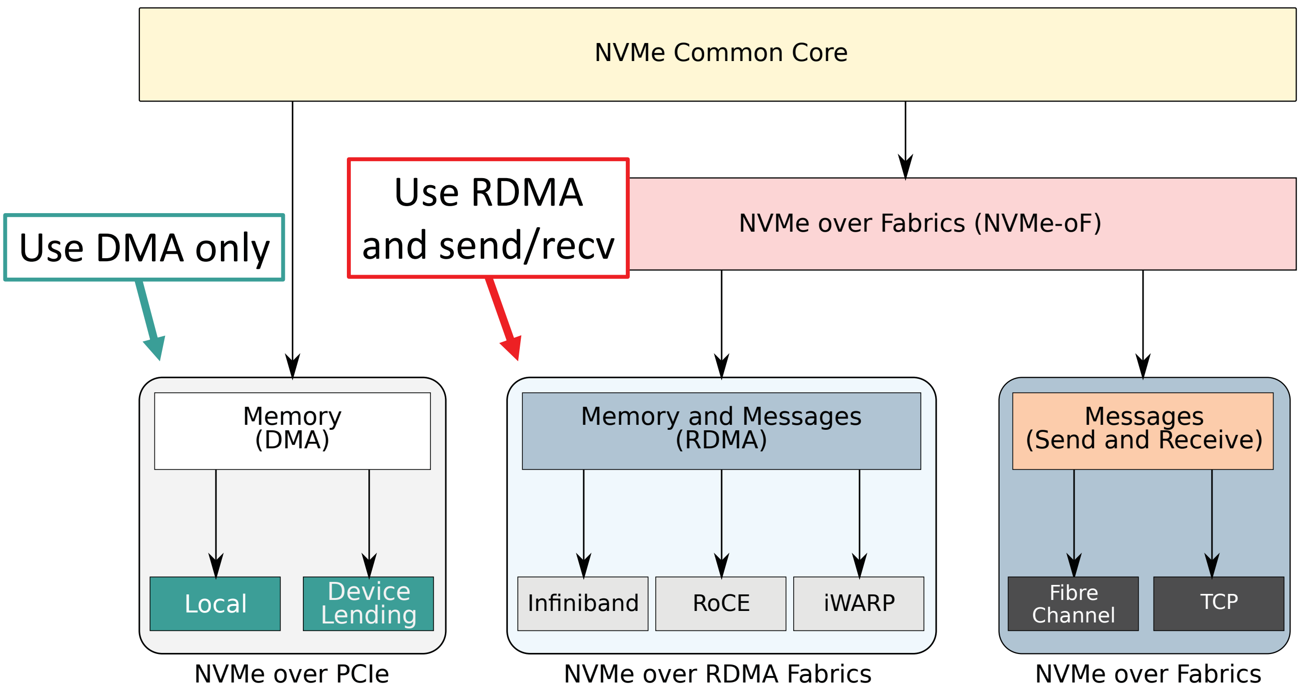
NVMe over Fabrics (NVMe-oF) is the emerging standard for accessing remote storage drives over a network ("fabric")



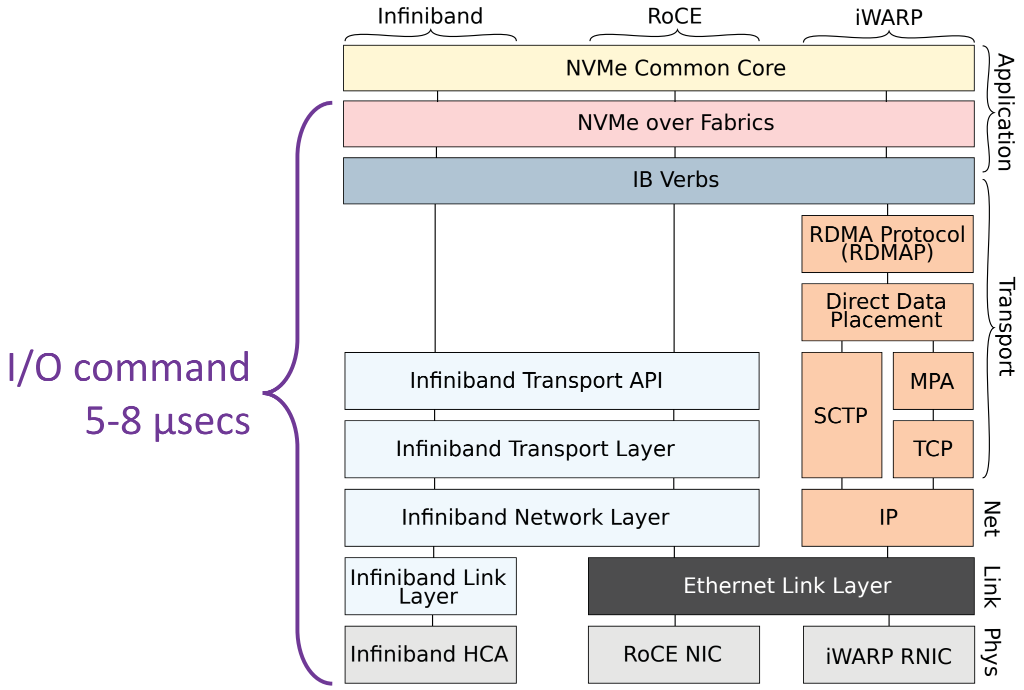
NVMe over Fabrics (NVMe-oF) is the emerging standard for accessing remote storage drives over a network ("fabric")



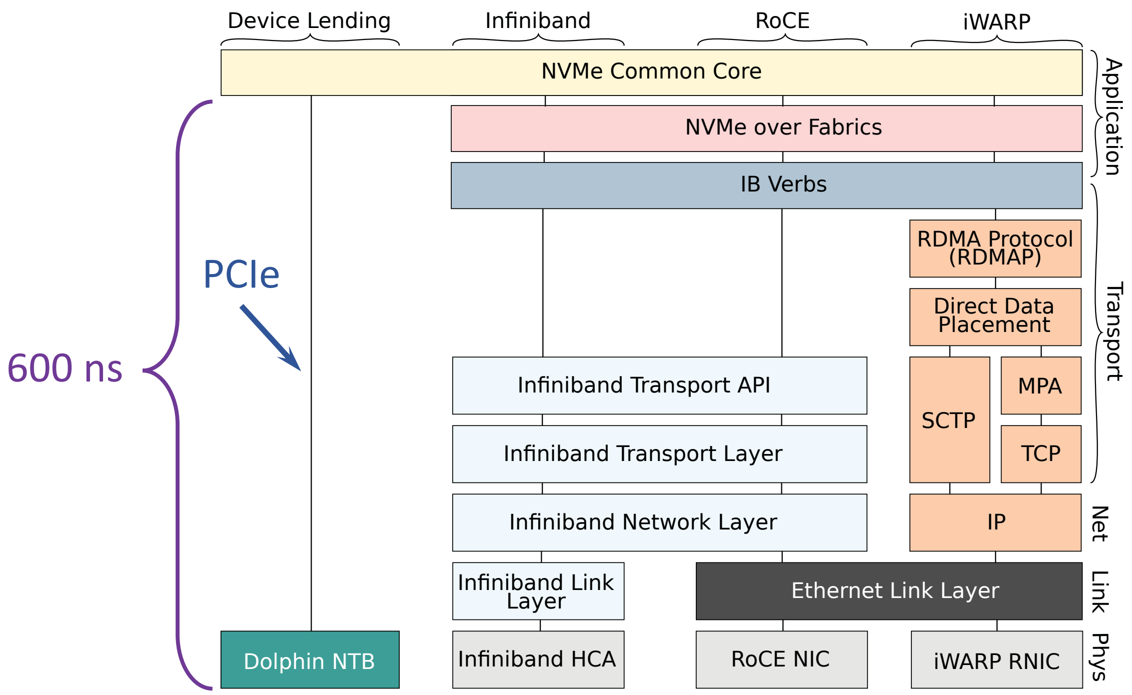
Using a protocol for Remote Direct Memory Access (RDMA), NVMe-oF is able to transfer data using zero-copy transfer methods



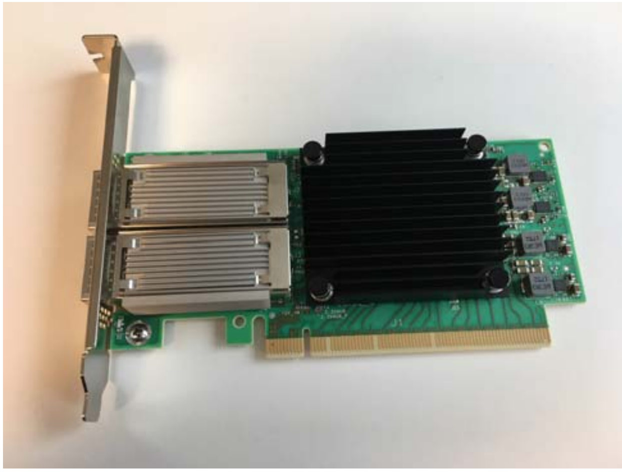
Using a protocol for Remote Direct Memory Access (RDMA), NVMe-oF is able to transfer data using zero-copy transfer methods



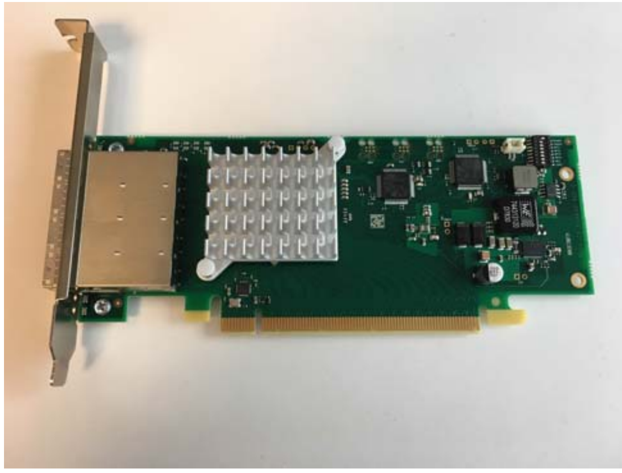
Using Dolphin Device Lending, memory can be accessed directly using native DMA without the need for an RDMA protocol



NVMe-oF vs Device Lending



Mellanox ConnectX-5 EDR

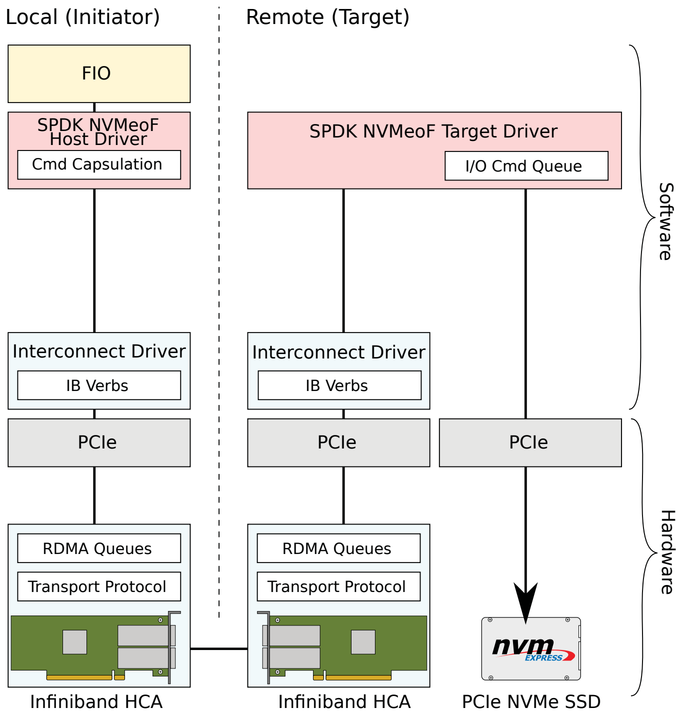


Dolphin PXH830 NTB Adapter

SPDK NVMe-oF

Storage Performance Development Kit

- User-space software library for creating storage applications
- Supports a wide variety of storage drives including NVMe drives
- Has a built-in NVMe-oF stack with Infiniband RDMA support

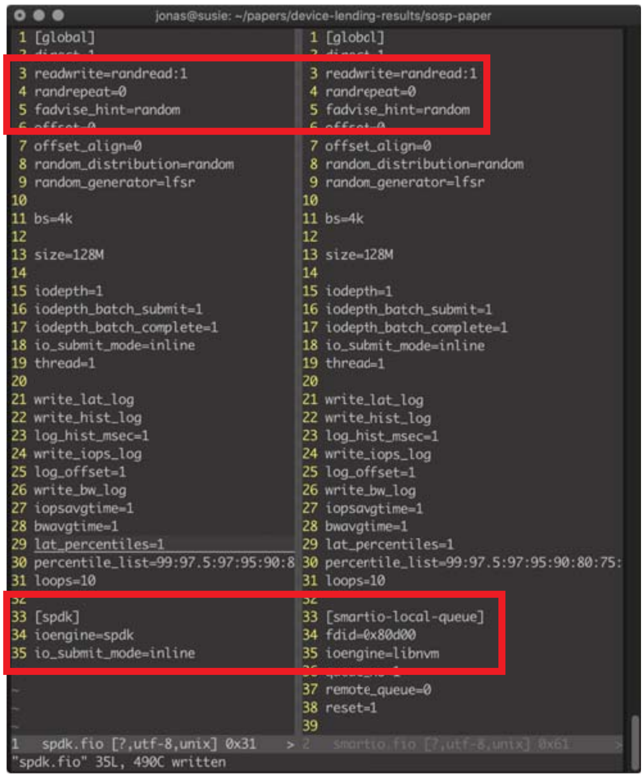


I/O Command Submission Path

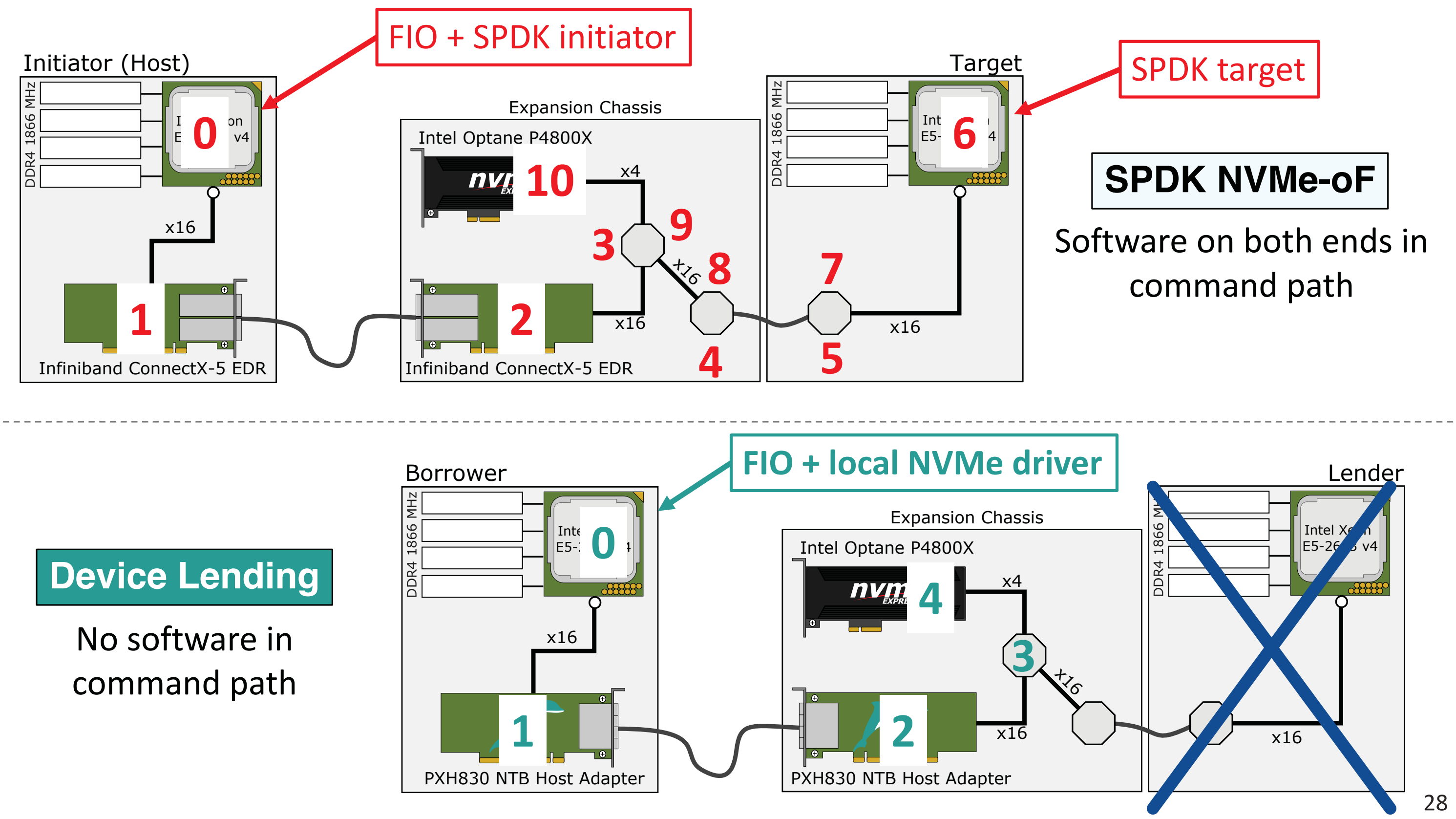
https://github.com/spdk/spdk

Test configuration

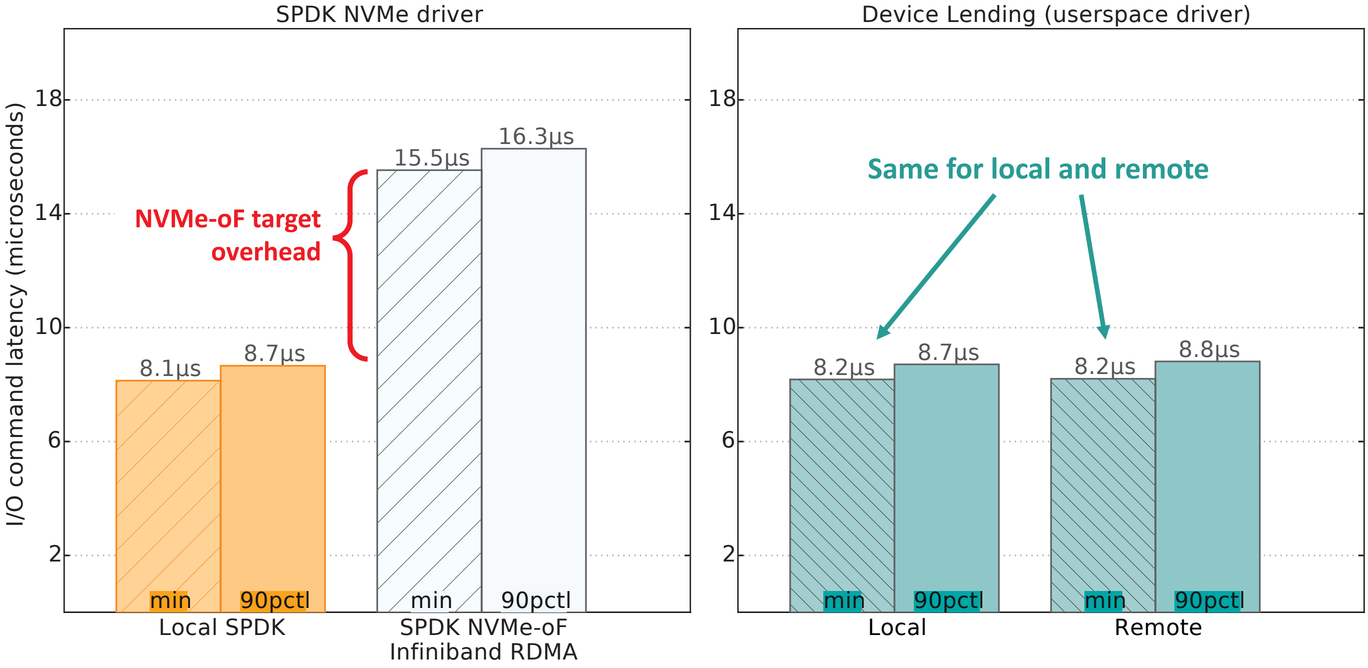
- Latency: 4 kB reads in a random pattern using **FIO 3.13**
- Ubuntu 18.04.2 LTS (4.15 kernel)
- NVMe drivers:
  SPDK vs userspace NVMe driver
- NVMe-oF implementation: SPDK 19.1.1



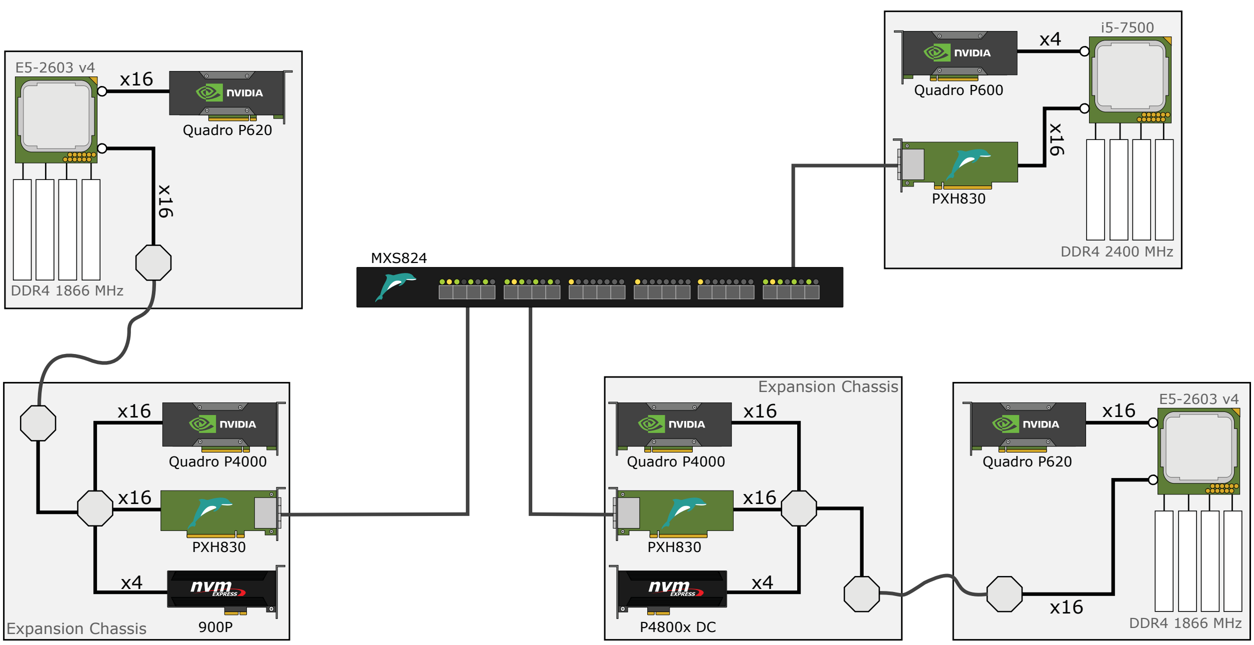

```
https://github.com/axboe/fio
https://github.com/spdk/spdk
https://github.com/enfiskutensykkel/ssd-gpu-dma
```

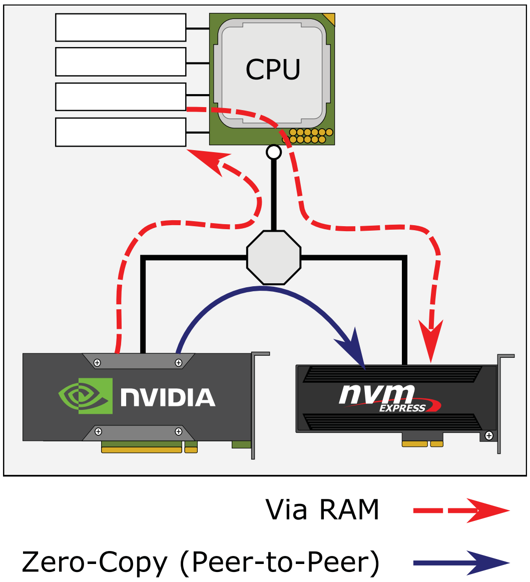
NVMe-oF vs Device Lending



Flexible storage workflows



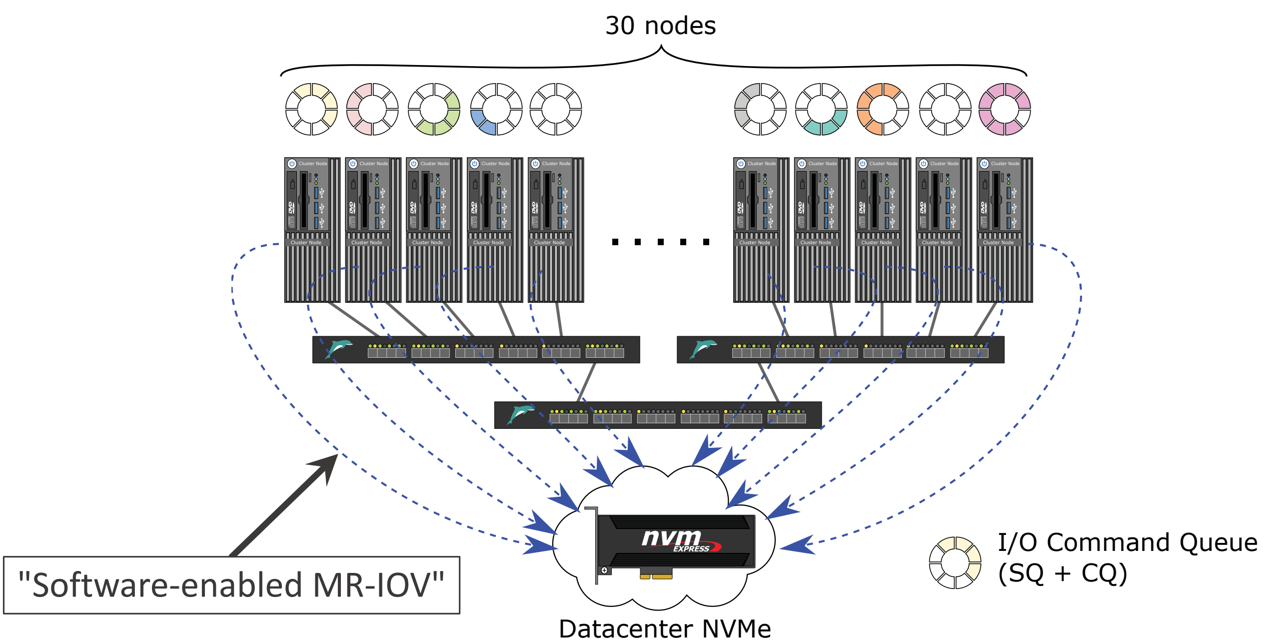
Dolphin NVMe Software Library



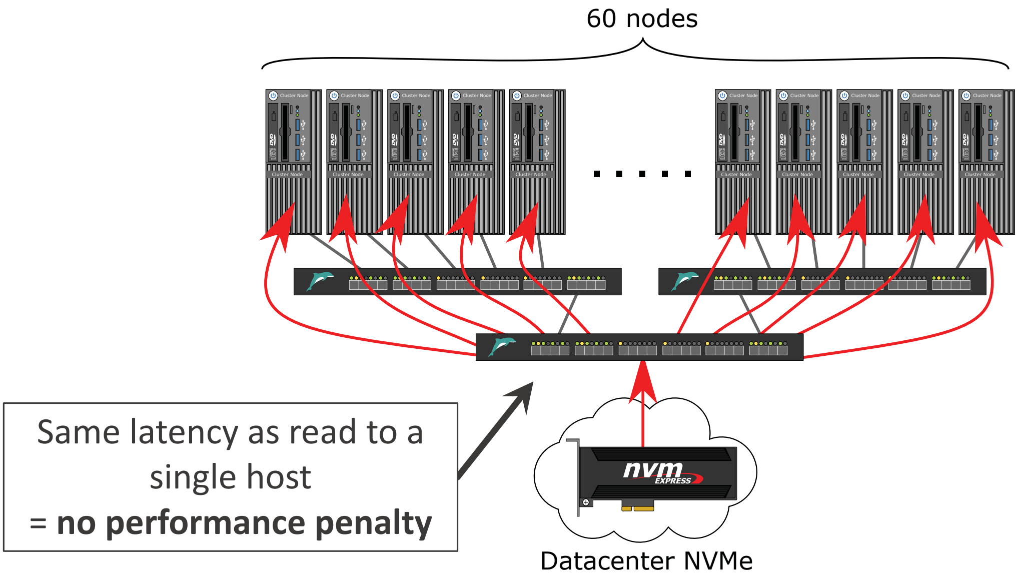
- User-space software driver for creating distributed storage applications
- Provides block-level access to NVMe drives **anywhere in the cluster fabric**
- GPUDirect support for zero-copy read/write to GPU memory
- Can be used in combination with Device Lending (e.g. remote GPUs)

`https://github.com/enfiskutensykkel/ssd-gpu-dma`

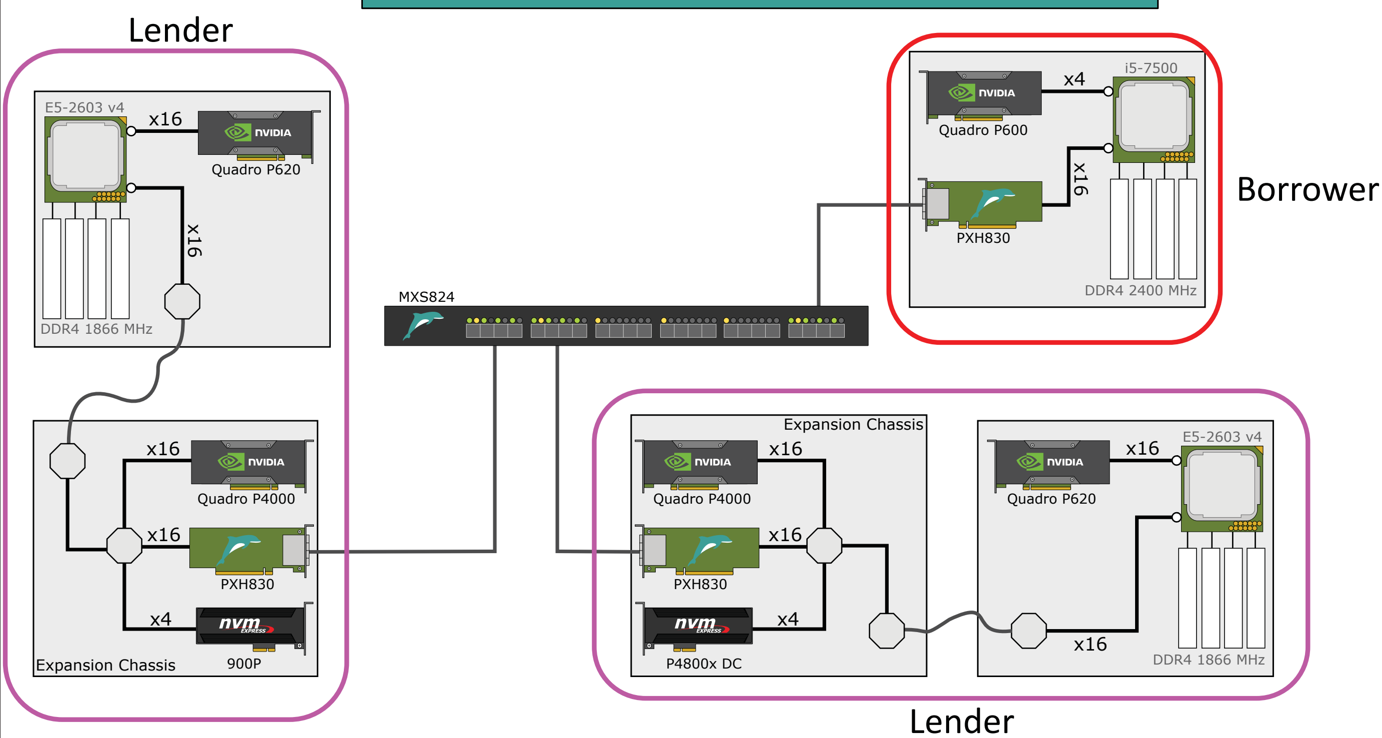
Multiple nodes may share a single function NVMe drive simultaneously by distributing individual I/O command queues

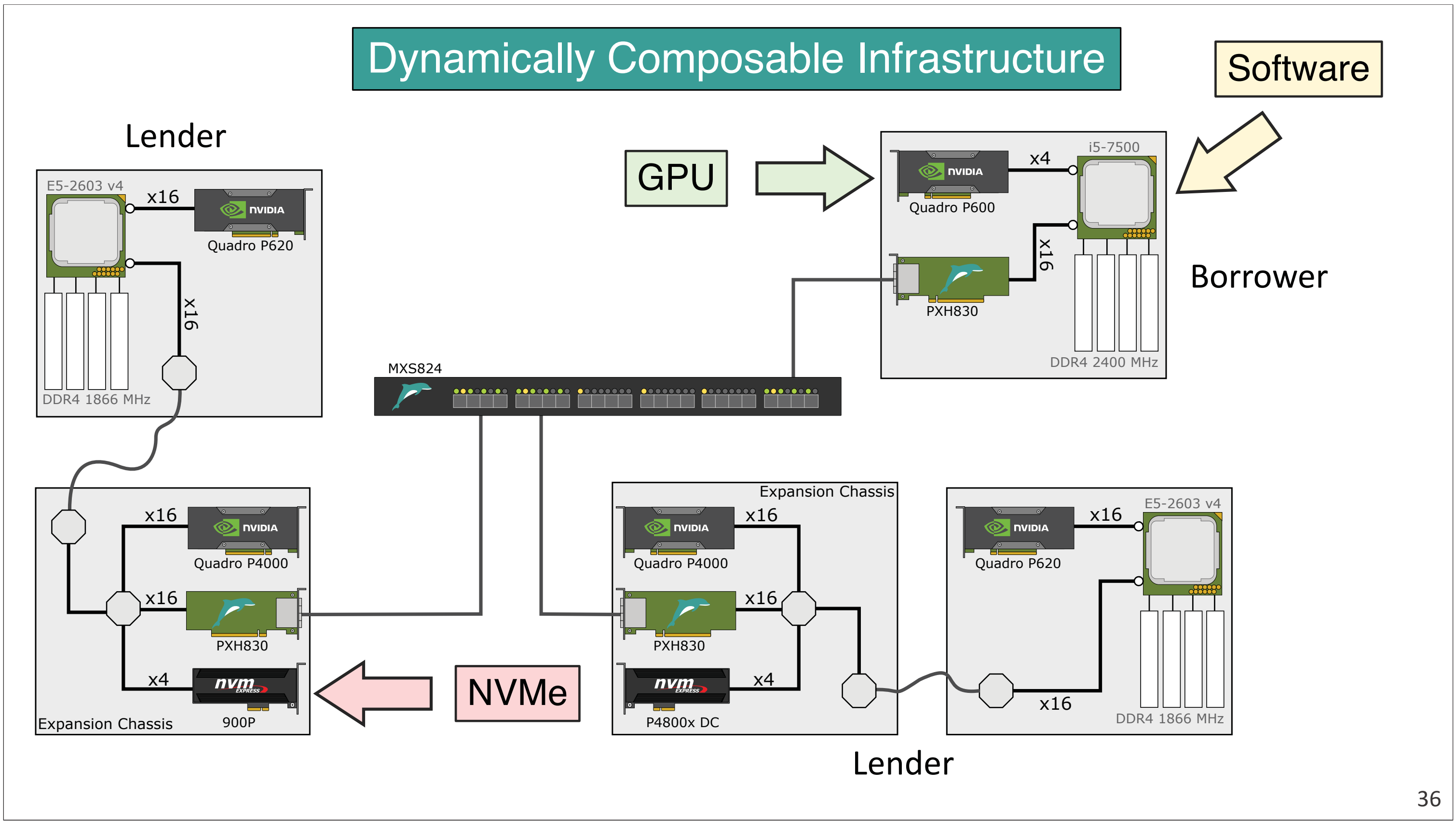


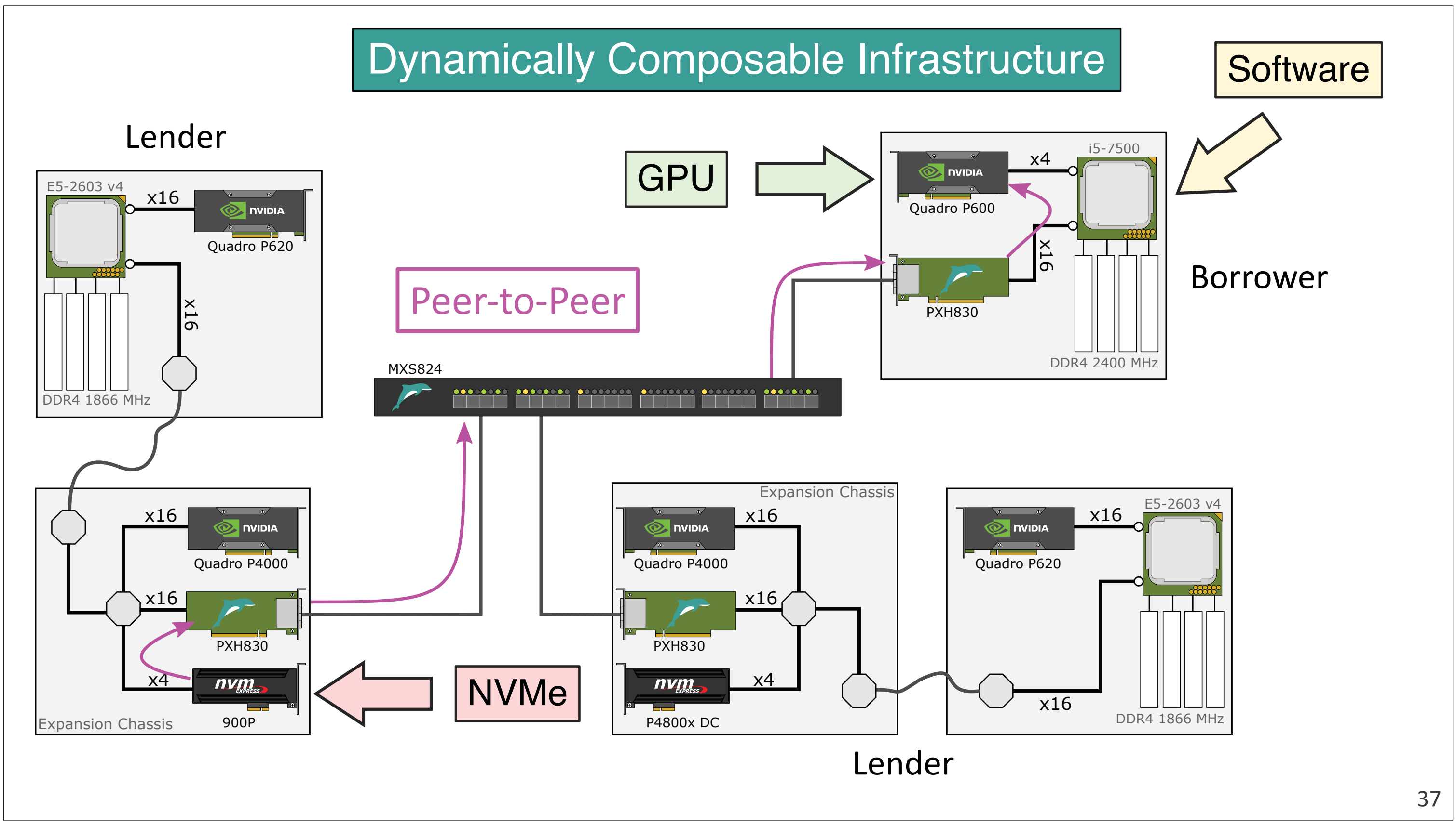
Using PCIe multicast, we can replicate data across multiple nodes in a single read operation

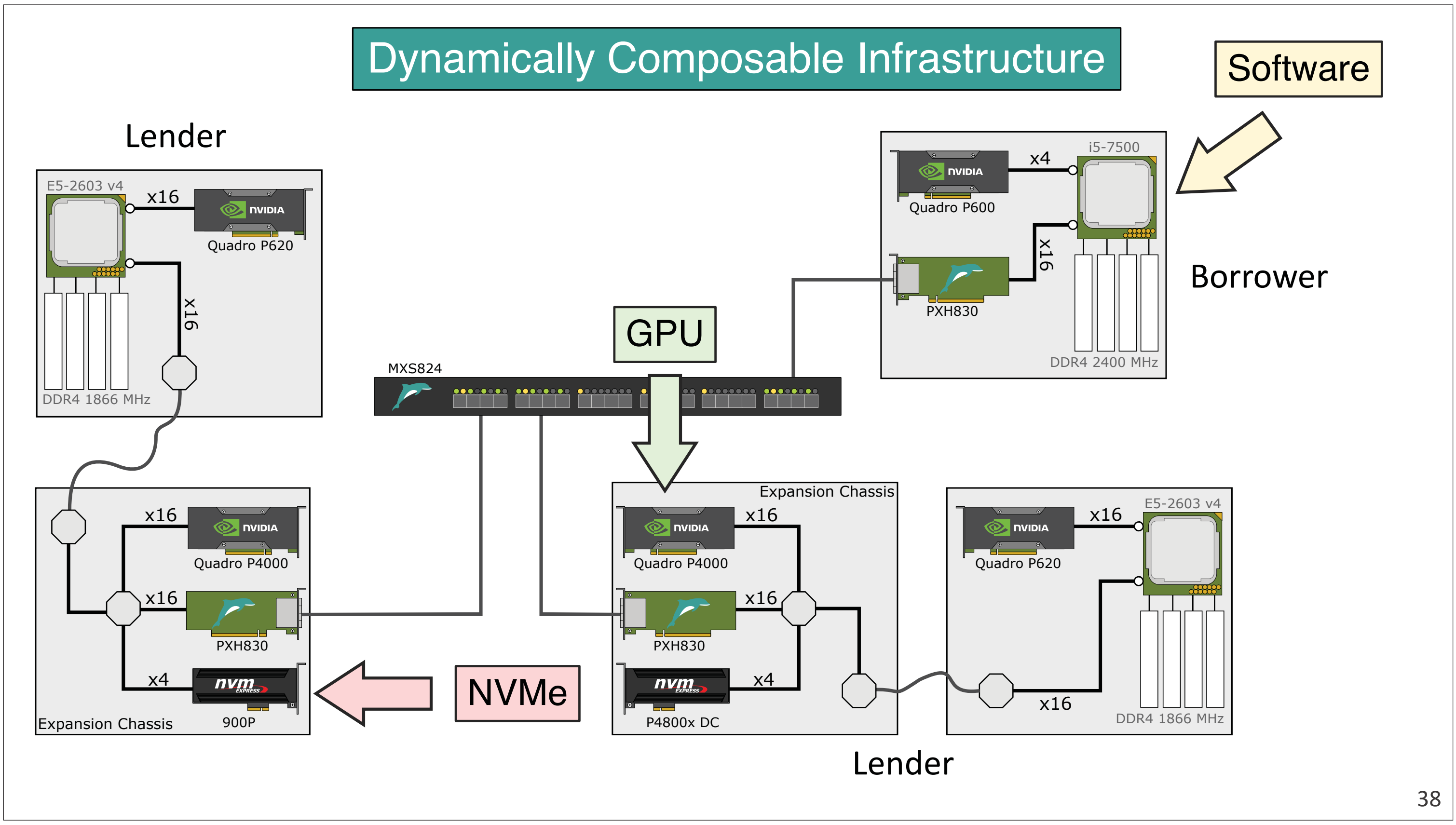


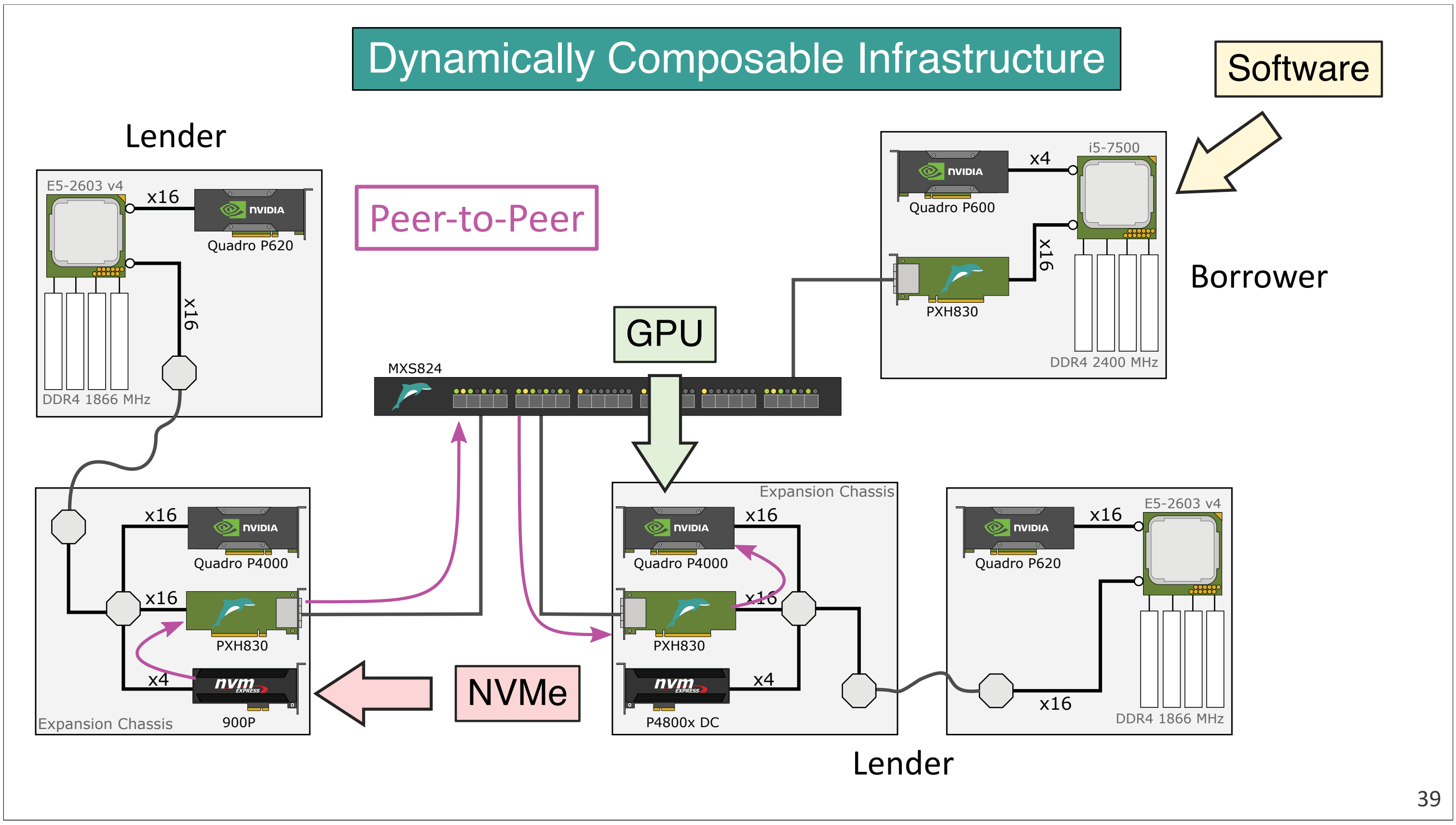
Dynamically Composable Infrastructure

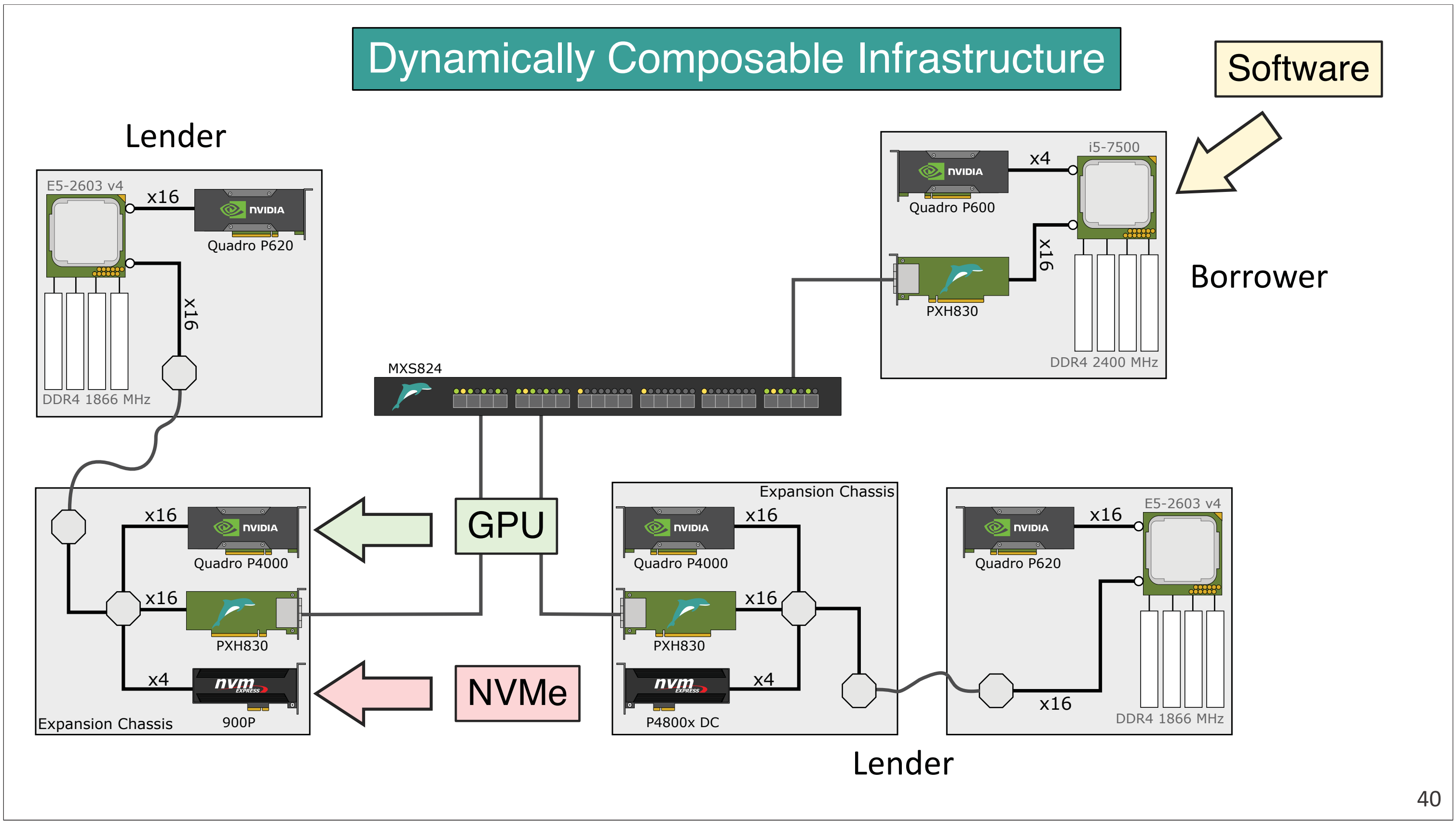


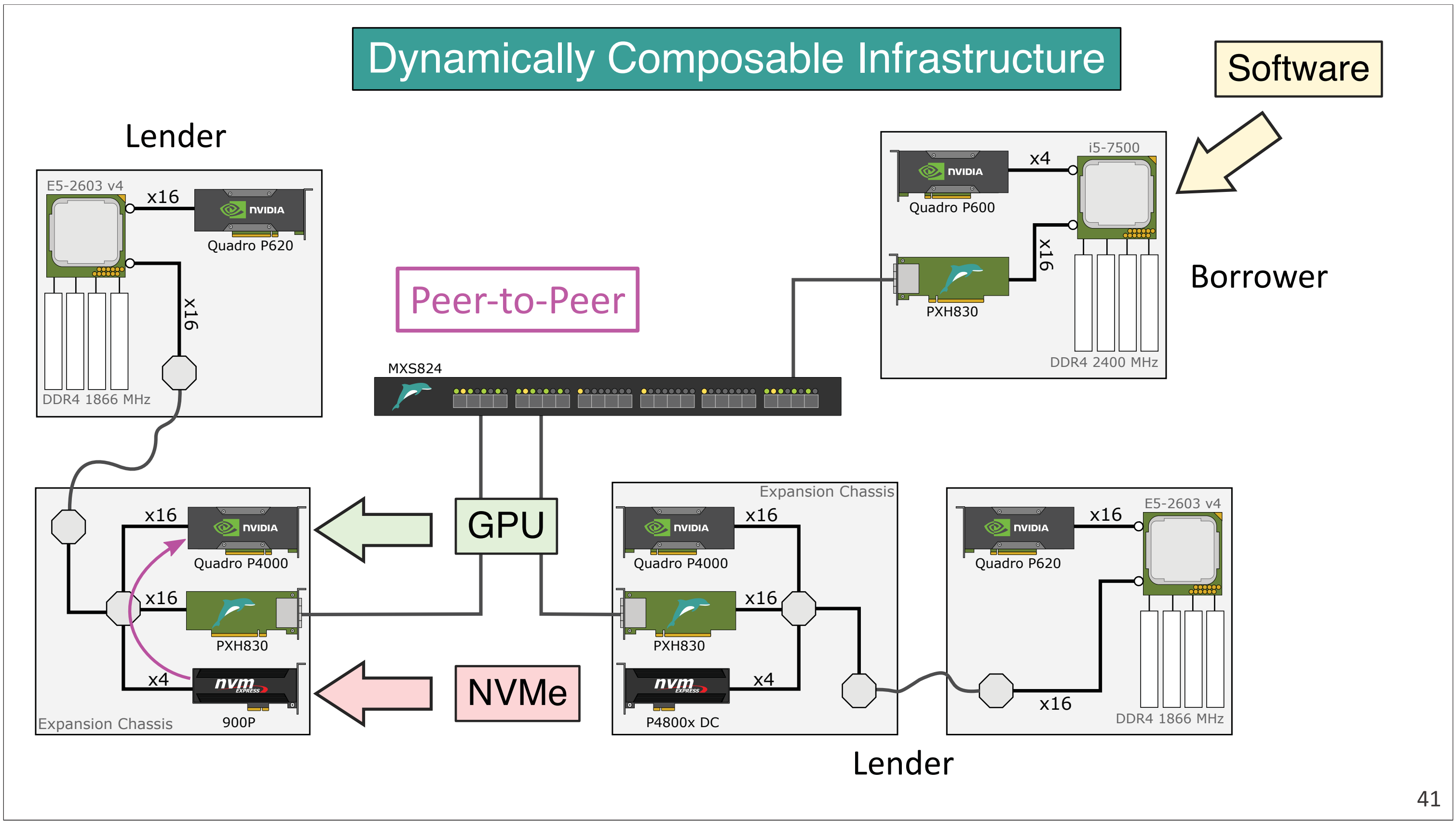

Dynamically Composable Infrastructure
Software
Lender
GPU
Borrower
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
i5-7500
x4
Quadro P600
x16
PXH830
DDR4 2400 MHz
MXS824
x16
Quadro P4000
x16
PXH830
x4
900P
Expansion Chassis
NVMe
Expansion Chassis
x16
Quadro P4000
x16
PXH830
x4
P4800x DC
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Lender
36

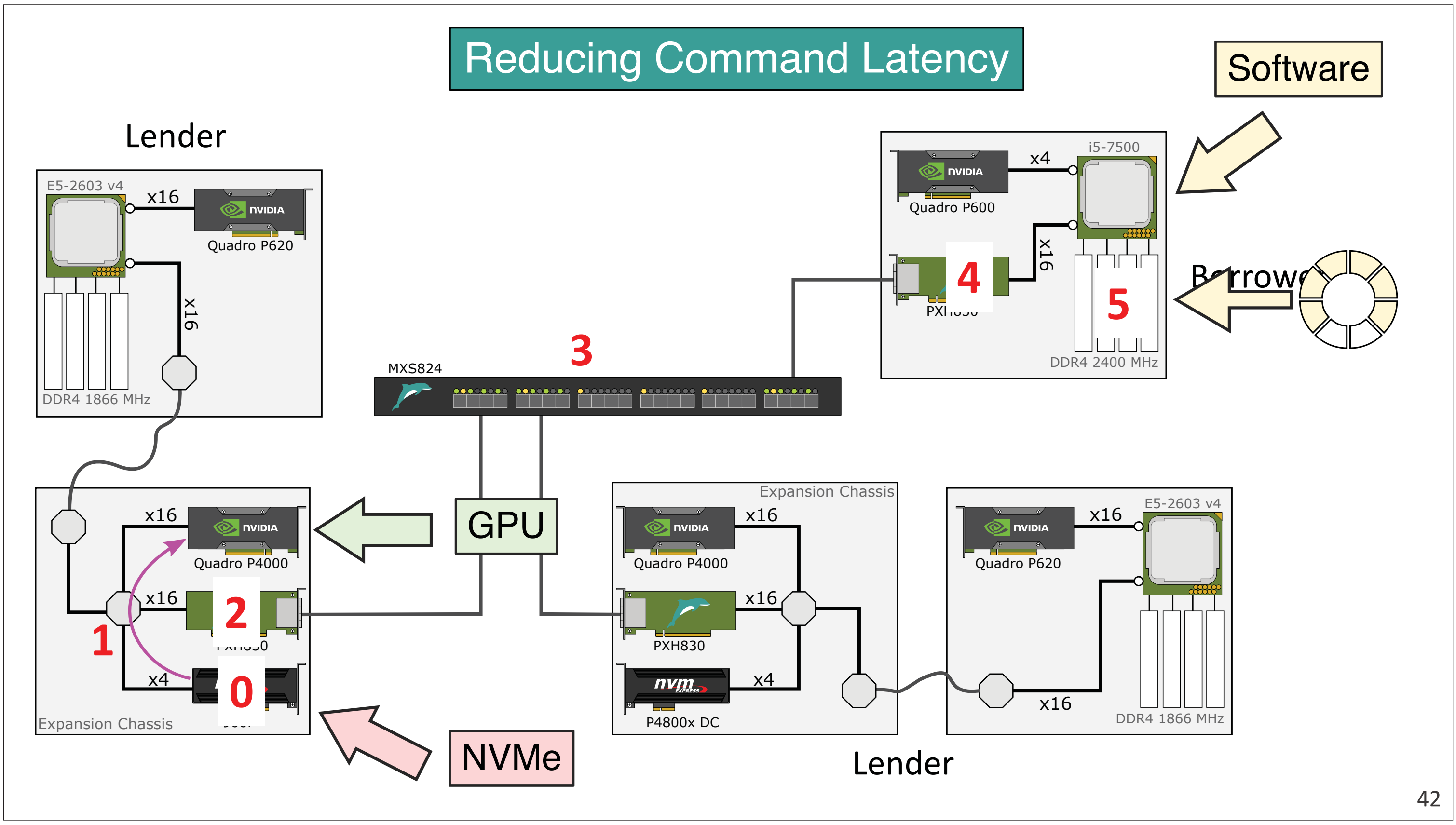

Dynamically Composable Infrastructure
Software
Lender
GPU
Borrower
Peer-to-Peer
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
i5-7500
x4
Quadro P600
x16
PXH830
DDR4 2400 MHz
MXS824
x16
Quadro P4000
x16
PXH830
x4
900P
Expansion Chassis
NVMe
Expansion Chassis
x16
Quadro P4000
x16
PXH830
x4
P4800x DC
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Lender
37

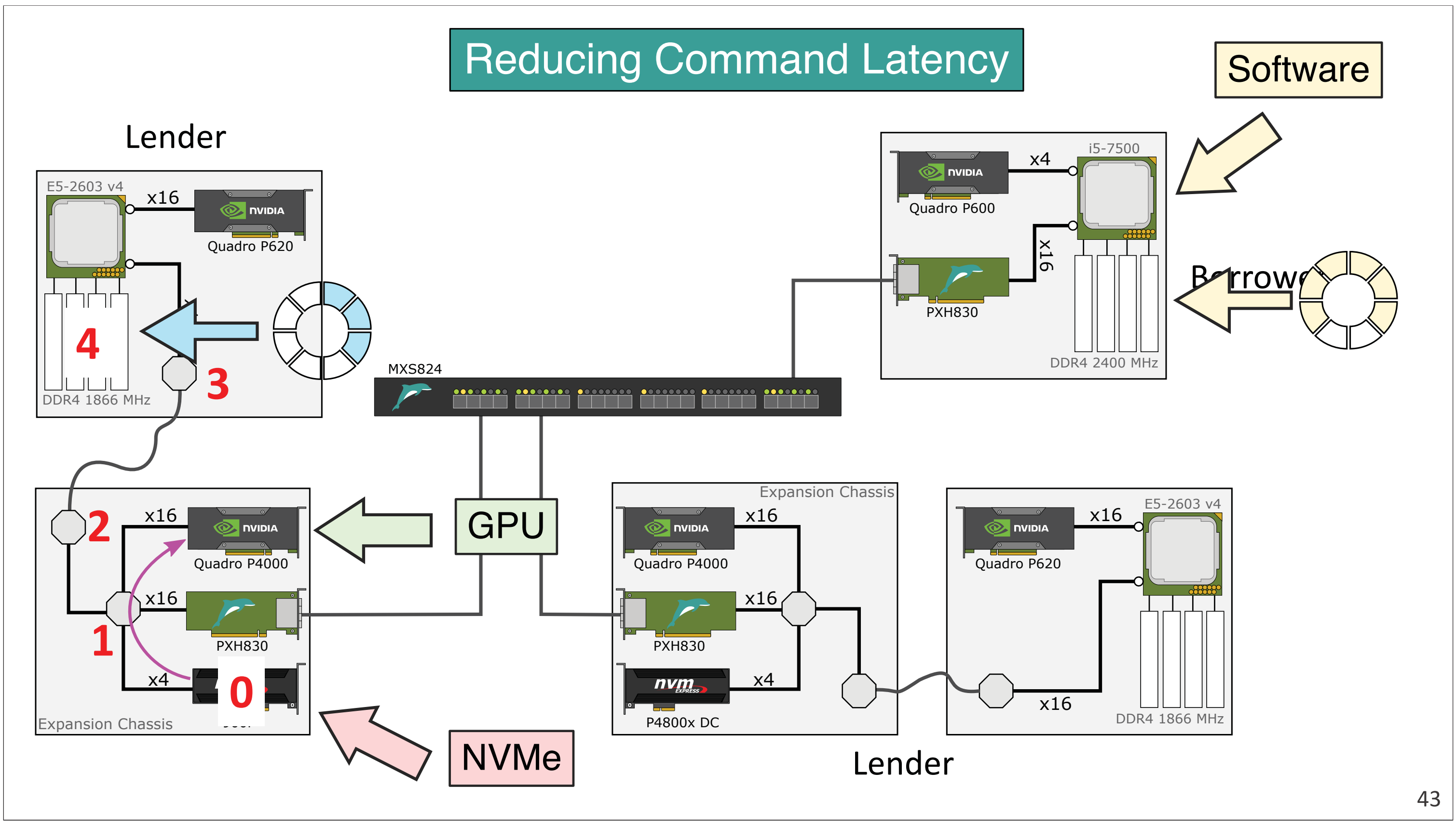

Dynamically Composable Infrastructure
Software
Lender
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Borrower
i5-7500
x4
Quadro P600
x16
PXH830
DDR4 2400 MHz
GPU
MXS824
x16
Quadro P4000
x16
PXH830
x4
900P
Expansion Chassis
NVMe
Expansion Chassis
x16
Quadro P4000
x16
PXH830
x4
P4800x DC
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Lender
38

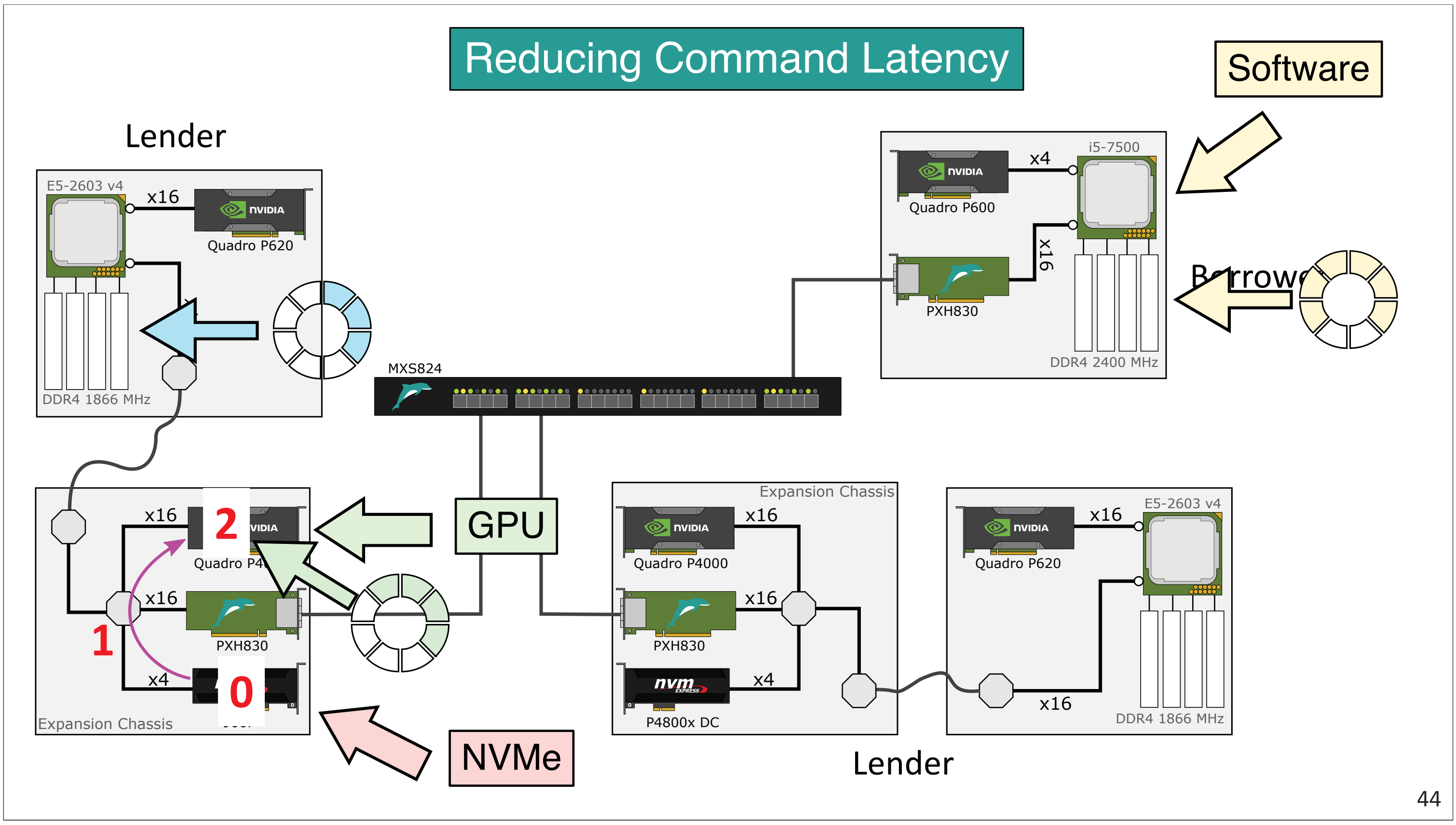

Dynamically Composable Infrastructure
Software
Lender
Peer-to-Peer
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Borrower
i5-7500
x4
Quadro P600
x16
PXH830
DDR4 2400 MHz
GPU
MXS824
x16
Quadro P4000
x16
PXH830
x4
900P
Expansion Chassis
NVMe
Expansion Chassis
x16
Quadro P4000
x16
PXH830
x4
P4800x DC
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Lender
39

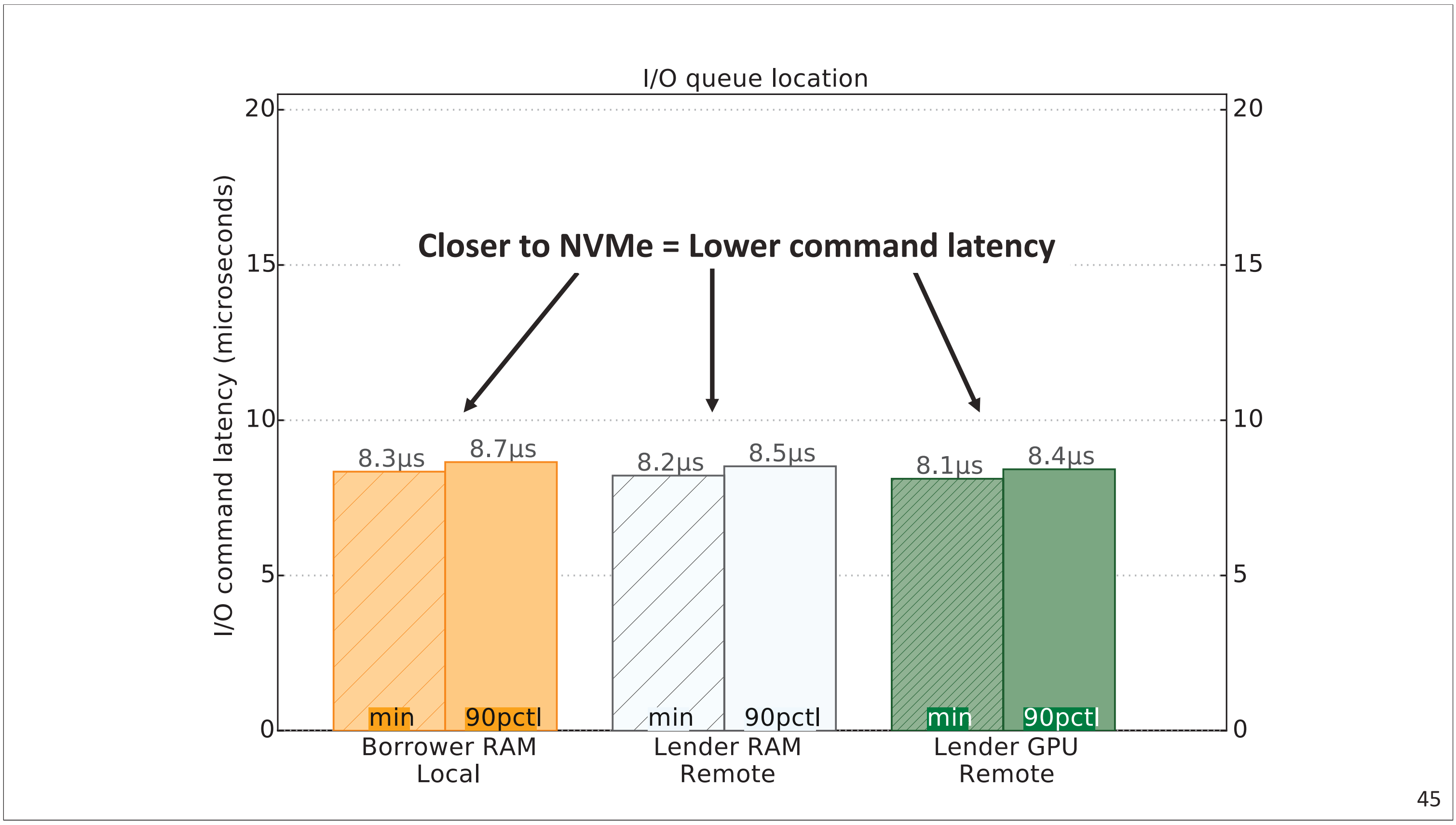

Dynamically Composable Infrastructure
Software
Lender
Borrower
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
MXS824
i5-7500
x4
Quadro P600
x16
PXH830
DDR4 2400 MHz
GPU
NVMe
x16
Quadro P4000
x16
PXH830
x4
900P
Expansion Chassis
Expansion Chassis
x16
Quadro P4000
x16
PXH830
x4
P4800x DC
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Lender
40


Dynamically Composable Infrastructure
Software
Lender
Borrower
Peer-to-Peer
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
MXS824
i5-7500
x4
Quadro P600
x16
PXH830
DDR4 2400 MHz
GPU
NVMe
x16
Quadro P4000
x16
PXH830
x4
900P
Expansion Chassis
Expansion Chassis
x16
Quadro P4000
x16
PXH830
x4
P4800x DC
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Lender
41


Reducing Command Latency
Software
Lender
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
MXS824
3
x4
i5-7500
Quadro P600
x16
4
5
DDR4 2400 MHz
GPU
x16
Quadro P4000
x16
2
1
x4
0
Expansion Chassis
NVMe
Expansion Chassis
x16
Quadro P4000
x16
PXH830
x4
P4800x DC
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Lender
42


Reducing Command Latency
Software
Lender
E5-2603 v4
x16
Quadro P620
4
3
DDR4 1866 MHz
MXS824
x4
i5-7500
Quadro P600
x16
PXH830
DDR4 2400 MHz
GPU
2
x16
Quadro P4000
x16
PXH830
1
x4
0
Expansion Chassis
NVMe
Expansion Chassis
x16
Quadro P4000
x16
PXH830
x4
P4800x DC
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Lender
43


Reducing Command Latency
Software
Lender
E5-2603 v4
x16
Quadro P620
DDR4 1866 MHz
MXS824
i5-7500
x4
Quadro P600
x16
PXH830
DDR4 2400 MHz
Borrower
x16
2
Quadro P4
x16
PXH830
1
x4
0
Expansion Chassis
GPU
NVMe
Expansion Chassis
Quadro P4000
x16
PXH830
x16
x4
P4800x DC
E5-2603 v4
x16
Quadro P620
x16
DDR4 1866 MHz
Lender
44


I/O queue location
Closer to NVMe = Lower command latency
I/O command latency (microseconds)
0
5
10
15
20
8.3µs
8.7µs
8.2µs
8.5µs
8.1µs
8.4µs
min
90pctl
min
90pctl
min
90pctl
Borrower RAM
Local
Lender RAM
Remote
Lender GPU
Remote
45

Questions?

simula

Publications

Flexible Device Compositions and Dynamic Resource Sharing in PCIe Interconnected Clusters
Cluster Computing, 2019

Flexible Device Sharing in PCIe Clusters using Device Lending
ACM ICPP Companion, 2018

Efficient Distributed Storage I/O using NVMe and GPUDirect in a PCIe Network
Presentation S9563, GTC Silicon Valley, 2019

https://github.com/enfiskutensykkel/ssd-gpu-dma

Torben Kling Petersen

Cray

On the Road to ExaScale – New Storage Technologies to Support ExaData

Torben Kling Petersen has worked with high performance computing in one form or another since 1994. After leaving academic life in 2000, he's held technical leadership positions in a number of tech companies (mostly through acquisitions) including Sun Microsystems, Oracle, Xyratex, Seagate, Cray Inc and from January 2020 at HPE.

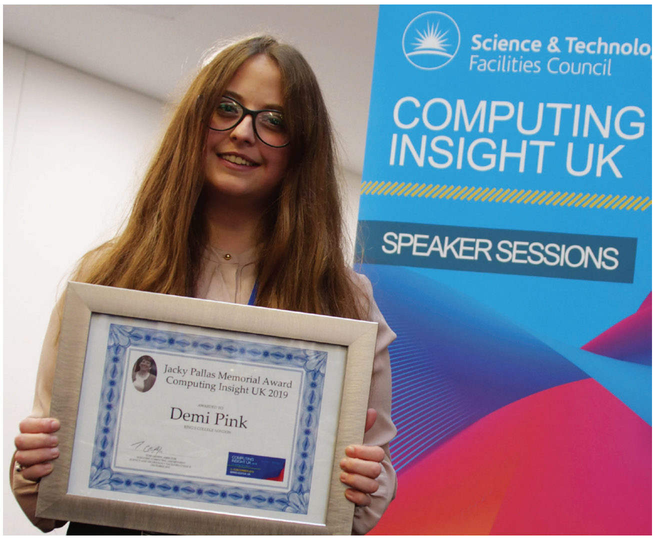
In the various companies, Torben has architected a significant number of HPC and HPC storage systems as well as worked with engineering to bring several new products to market. Torben has authored a large number of white papers and technical articles over the years and have presented at more conferences and events that can be easily listed.

At Cray, Torben currently works as the lead HPC storage architect for strategic engagements in EMEA and APAC. Torben works out of his home in Goteborg, Sweden when not flying all over the world to meet colleagues and customers.

Abstract

Improving data intensive workflows in modern supercomputers by any means possible continues to be a focus of both industry and academic research. And while big steps have been made, most have served to move the I/O bottleneck somewhere else and not actually solve it. With technologies such as persistent memory, next gen NVMe solutions and intelligent tiering software, the goalposts have been moved and traditional approach to the problems, are no longer viable.

With the first pre-exascale and exascale computers being announced, the storage solutions and data acceleration technologies has to match. This talk is intended to provide a view from Cray on new hardware technologies, interconnect methodologies and the enhanced software strategies.



CIUK 2019 JACKY PALLAS MEMORIAL AWARD PRESENTATION

Demi Pink

King's College London

On the Structure of Lipid-Based Nanoparticles for Drug Delivery

Demi Pink is a PhD student in the Department of Physics at King's College London. Whilst she studied for her undergraduate degree in Chemistry at the University of Leicester, she received the OUP Achievement in Chemistry Prize before graduating with 1st Class Honours. Following this, she joined the BBSRC funded London Interdisciplinary Doctoral Training Program where she undertook rotation projects in cell biology and biophysics before beginning her PhD under the supervision of Dr Chris Lorenz and Prof. Jayne Lawrence. Her work uses molecular dynamics simulations and small angle neutron scattering to investigate the self-assembly of lipid-based drug delivery vehicles and their encapsulation of small hydrophobic drug molecules.

Abstract

Solid lipid nanoparticles (SLNs) have a crystalline lipid core which is stabilised in solution by interfacial surfactants. They are considered favourable candidates for future drug delivery vehicles as they are capable of storing and release bioactive molecules. However, when stored over time it is thought that the lipids undergo polymorphic transitions which result in the premature expulsion of the drug molecules. To date, significant experimental studies have been conducted with the aim of investigating the physicochemical properties of SLNs, including their long-term stability, but as-of-yet, no molecular scale investigations have been reported on the behaviours that drive SLN formation and their subsequent polymorphic transitions. Using a combination of small angle neutron scattering (SANS) and all-atom molecular dynamics simulations (MD) we have generated a detailed, atomistic description of the internal structure of an SLN formed from the triglyceride, tripalmitin, and the Brij O10 surfactant. In addition to studying the SLN, we have performed further experiments and molecular-dynamic simulations on the formation of a triolein-based liquid lipid nanoparticle (LLN) which is stabilised by the same Brij O10 surfactant. LLNs are, like SLNs, of interest for their potential applications in drug delivery. This has allowed us to characterise the structure of the LLN in a similar manner to the SLN and to compare the two contrasting nanostructures in order to better understand the relationship between a nanoparticle's internal structure and its role in drug delivery. As well as studying the structure and formation of the nanoparticles, we have characterised and compared the processes involved in the encapsulation and localisation of the steroidal drug, testosterone propionate, by both the SLN and the LLN.

Molecular dynamics and machine learning to study the atomistic structure of drug delivery vehicles

Demi Pink

Computing Insights UK, December 2019

KING'S College LONDON

Introduction

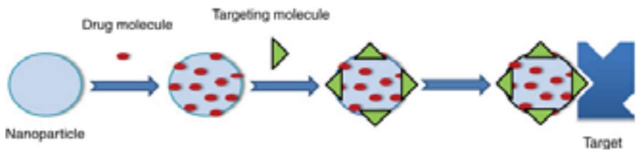
1. Why study drug delivery vehicles?
2. What are solid lipid nanoparticles (SLN)
3. What is Molecule Dynamics (MD)
4. Why HPC is used
5. Some results

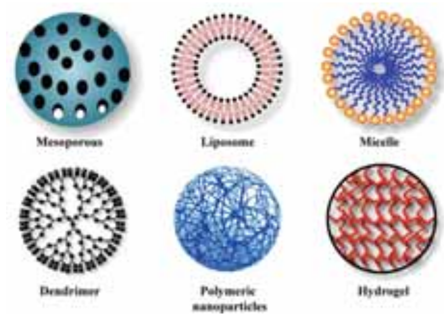
KING'S College LONDON

Drug delivery vehicles

- Many drug molecules are hydrophobic and are poorly soluble.
- DDV improve the solubility of these hydrophobic drug molecules.

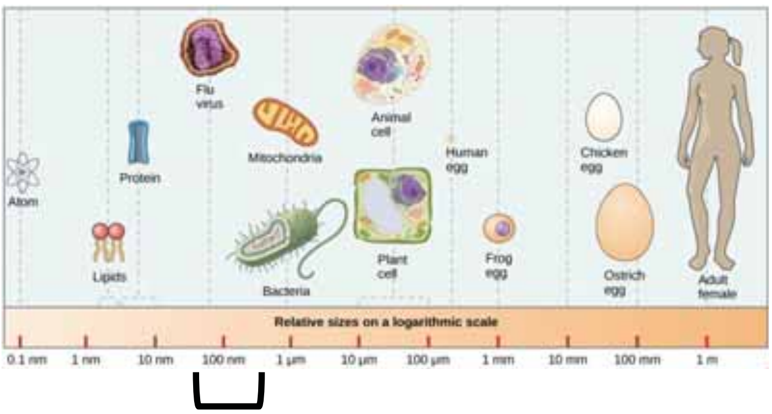
- Targeted drug delivery





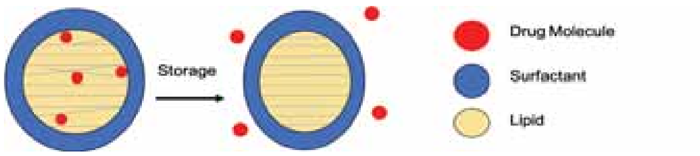
Drug delivery vehicles

Size limitations of experimental methods



50 – 500 nm range

Solid Lipid Nanoparticles (SLN)

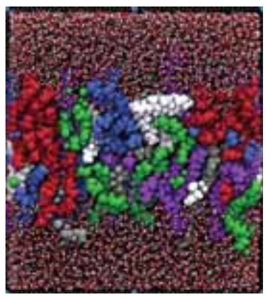


- ➢ Lipid based nanoparticle used in drug delivery
- ➢ Lipids are solid at room temperature allowing drug molecules to be trapped amongst the solid lipids
- ➢ Experimental method of preparation impacts location of drug within the nanoparticle

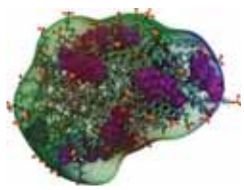
- ➢ **Simulate a SLN**
- ➢ Understand how structure might be impacting drug localisation

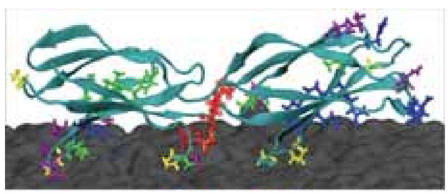
KING'S College LONDON

Molecular dynamics



- Computational simulation technique
- 1950/1960s
- Requires:
  1. Forcefield
  2. Package to run the simulation
  3. Time





KING'S College LONDON

Molecular dynamics

- Requires:
    1. Forcefield
        - → All of information about atoms and bonds needed to calculate the intra and intermolecular forces
        - → Choice of forcefield depends on what you are simulating

Molecular dynamics

1. Forcefield

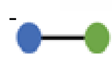
→ Bonded and non-bonded parameters

→ Bonded:

$$U_{bond} = k_{bond}(l - l_o)^2 \qquad U_{angle} = k_{angle}(\theta - \theta_o)^2 \qquad U_{dihedral} = \sum_{i=1}^{5} A_n \cos^{n-1}(\phi)$$

→ Non-bonded:

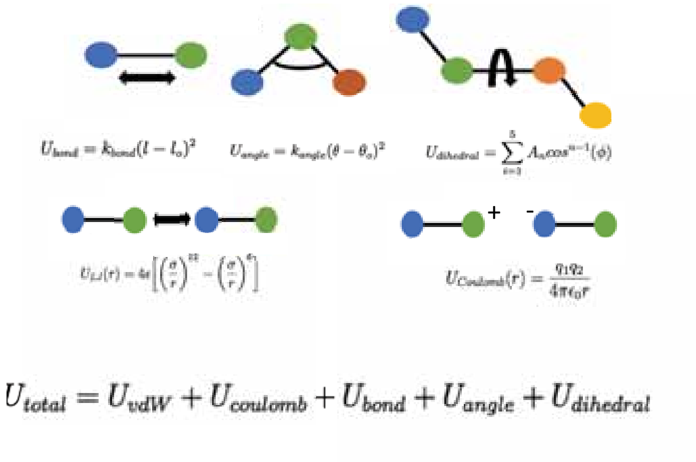
$$U_{LJ}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$



$$U_{Coulomb}(r) = \frac{q_1 q_2}{4\pi\epsilon_0 r}$$

Molecular dynamics

1. Forcefield

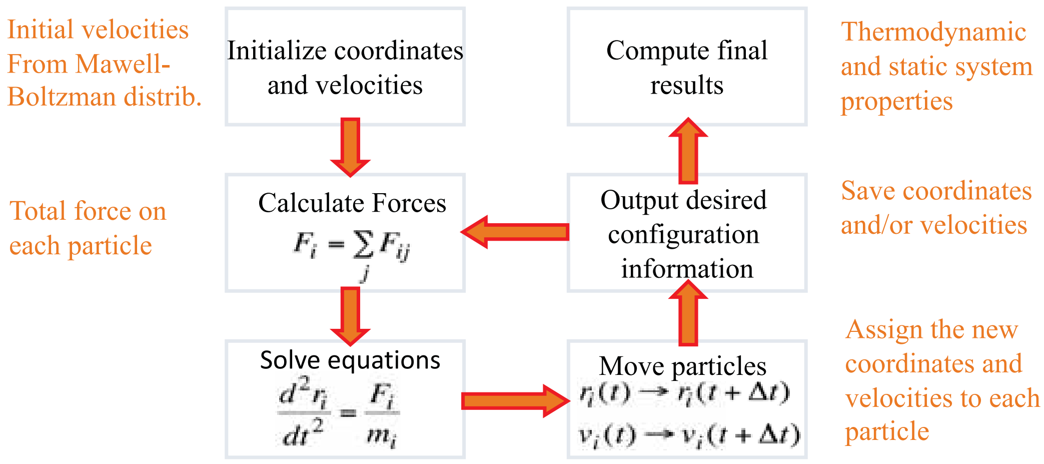


Molecular dynamics

- Requires:
  1. Forcefield
  2. Package to run the simulation
     - → Runs MD algorithms using forcefield values.
     - → GROMACS, NAMD LAMMPS and AMBER
     - → Choice of simulation package often depends on your forcefield

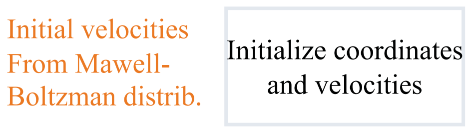
Molecular dynamics

- Requires:
  2. Package to run the simulation



Molecular dynamics

- Requires:
  2. Package to run the simulation



Starting structure

Molecular dynamics

- Requires:
  2. Package to run the simulation

Initial velocities From Mawell-Boltzman distrib.

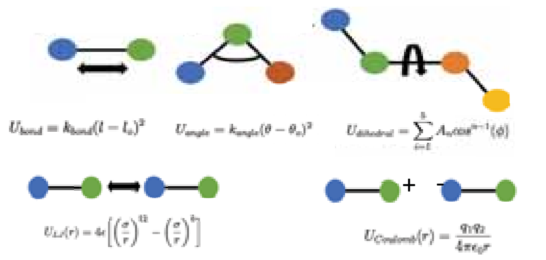
Initialize coordinates and velocities

↓

Total force on each particle

Calculate Forces

$$F_i = \sum_j F_{ij}$$



$$U_{total} = U_{vdW} + U_{coulomb} + U_{bond} + U_{angle} + U_{dihedral}$$

Molecular dynamics

- Requires:
  2. Package to run the simulation

Initial velocities From Mawell-Boltzman distrib.

Initialize coordinates and velocities

↓

Total force on each particle

Calculate Forces

$$F_i = \sum_j F_{ij}$$

↓

Solve equations

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i}$$

$$F_i = m_i a_i$$

$$F_i = m_i \frac{d^2 x_i}{dt^2}$$

Integration algorithm

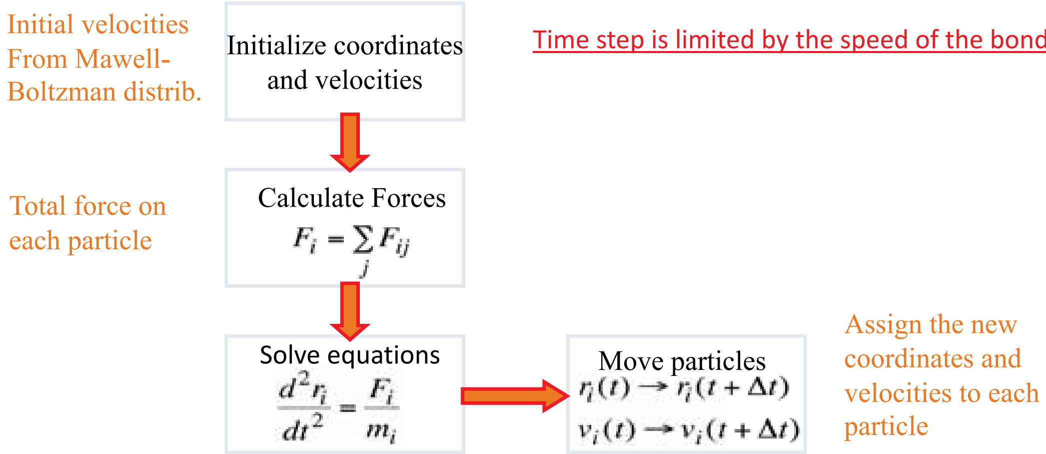
KING'S College LONDON

Molecular dynamics

- Requires:

  2. Package to run the simulation

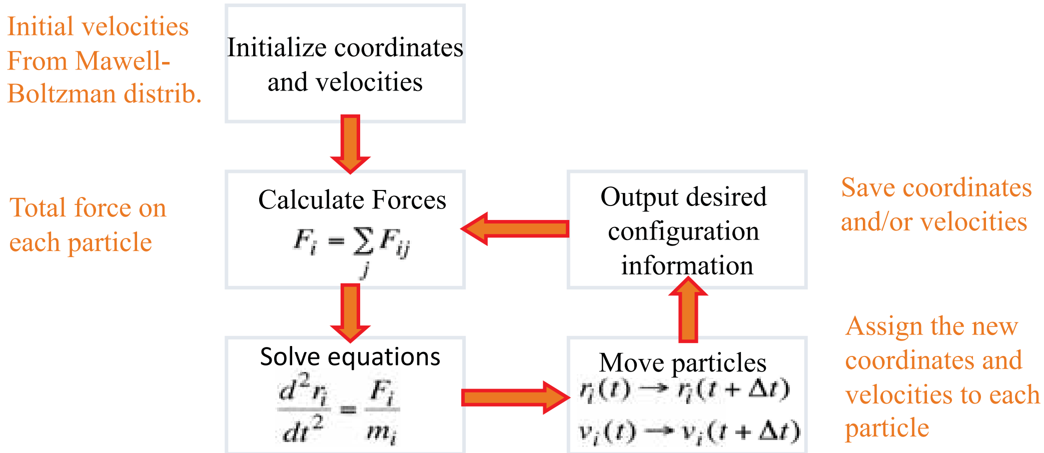
<u>Time step is limited by the speed of the bonds.</u>



Molecular dynamics

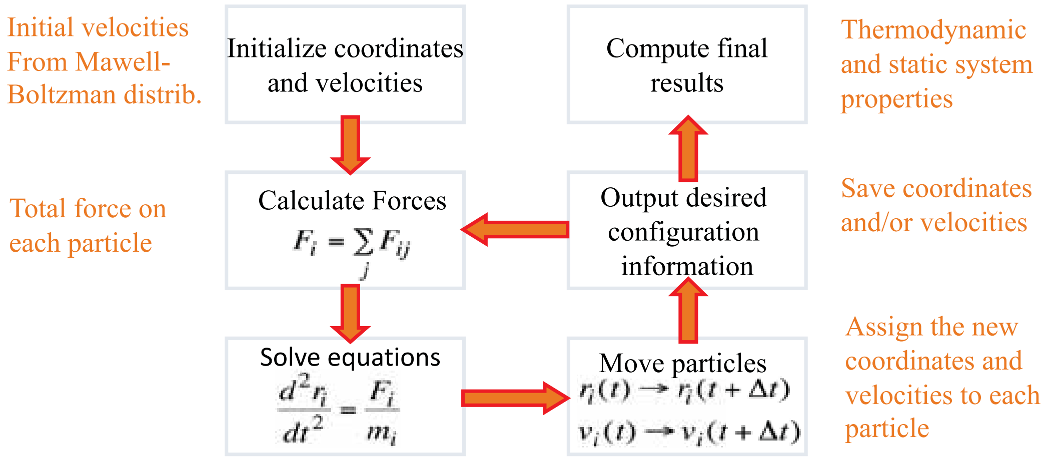
- Requires:

  2. Package to run the simulation



Molecular dynamics

- Requires:
  2. Package to run the simulation



KING'S College LONDON

Molecular dynamics

- Requires:
  1. Forcefield
  2. Package to run the simulation

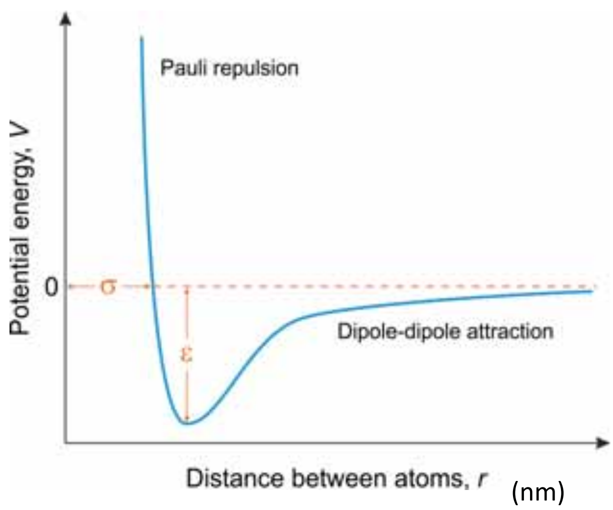
KING'S College LONDON

Molecular dynamics

- Requires:
  1. Forcefield
  2. Package to run the simulation
  3. Time
     - → Limiting factor in most simulations

Molecular dynamics

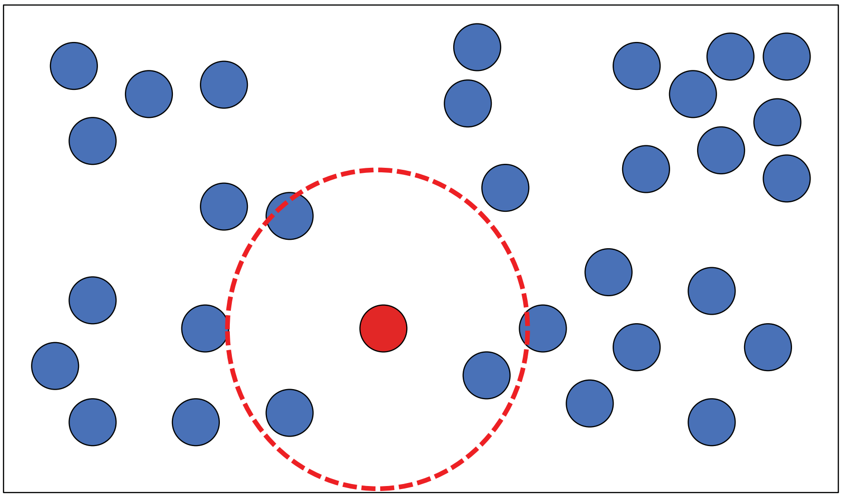
- Requires:
  3. Time
     - → Evaluating the forcefield
     - → Lennard-Jones forces act very short range
     - → Calculating the forces past a certain distance is a waste of computational effort
     - → Cut-off distance, r_c



Molecular dynamics

Neighbour Lists

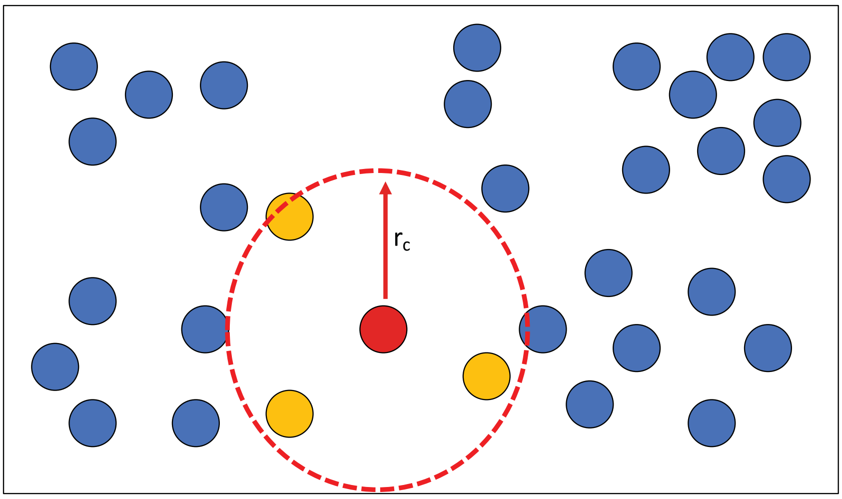
- In order to increase computational efficiency, simulations take advantage of neighbour lists when calculating non-bond interactions



Molecular dynamics

Neighbour Lists

- Lists the atoms that are within the cut-off distance of another atom.



Molecular dynamics

- Requires:
    3. Time

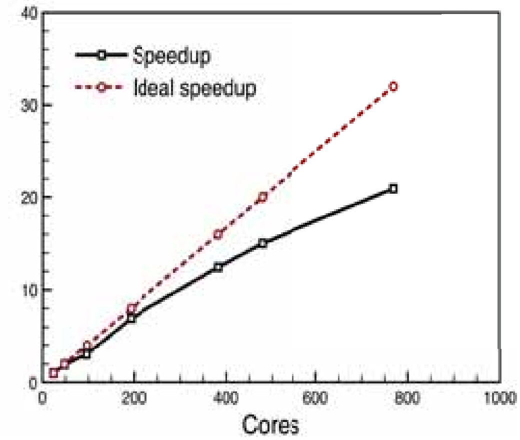
→ Evaluating the forcefield

→ Whilst the number of atoms per processor is > 1000, the speed up scales linearly with the number of processors.

→ 24 cores: 0.58 ns/day → 207 days

→ 360 cores: 8.7 ns/day → 14 days

→ Domain decomposition



Molecular dynamics

Domain decomposition

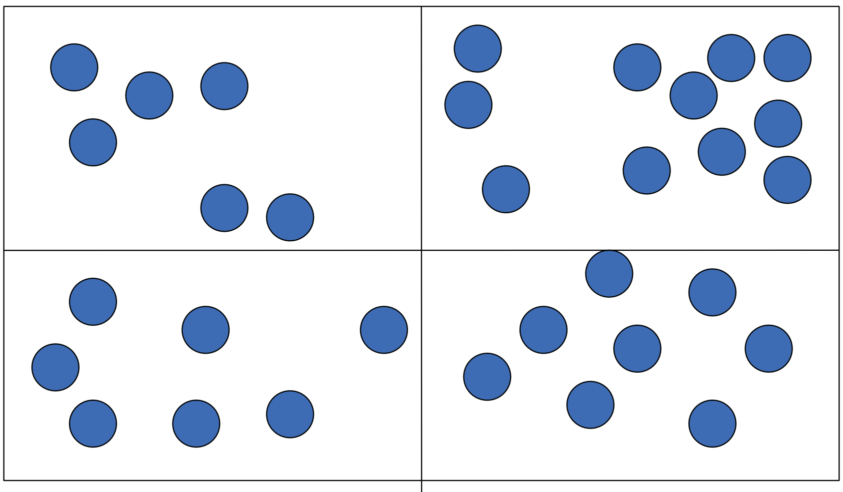
(Ex. 4 processors → 32 atoms)

Cell 1: 6 atoms

Cell 2: 11 atoms

Cell 3: 7 atoms

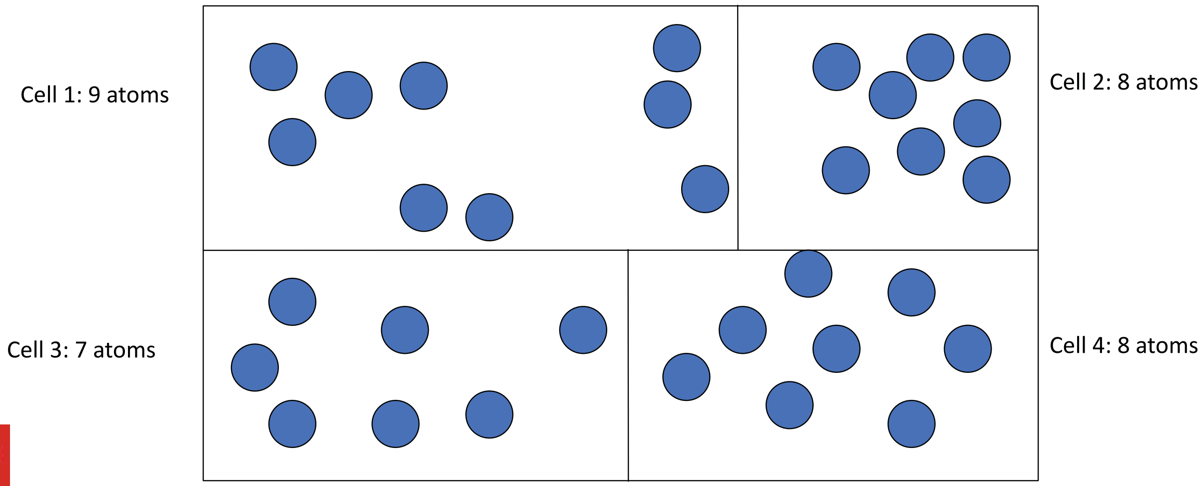
Cell 4: 8 atoms



Molecular dynamics

Dynamic load balancing

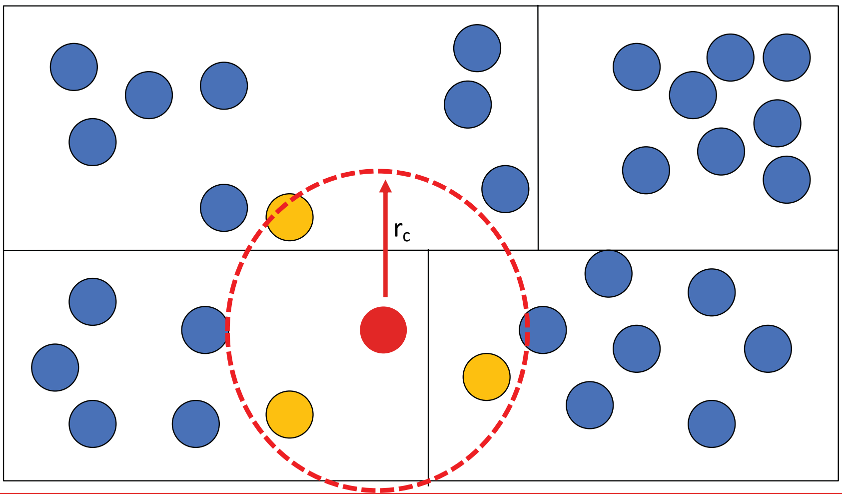
Divide up system so each processor does almost the same work (load balancing)



Molecular dynamics

Neighbour Lists

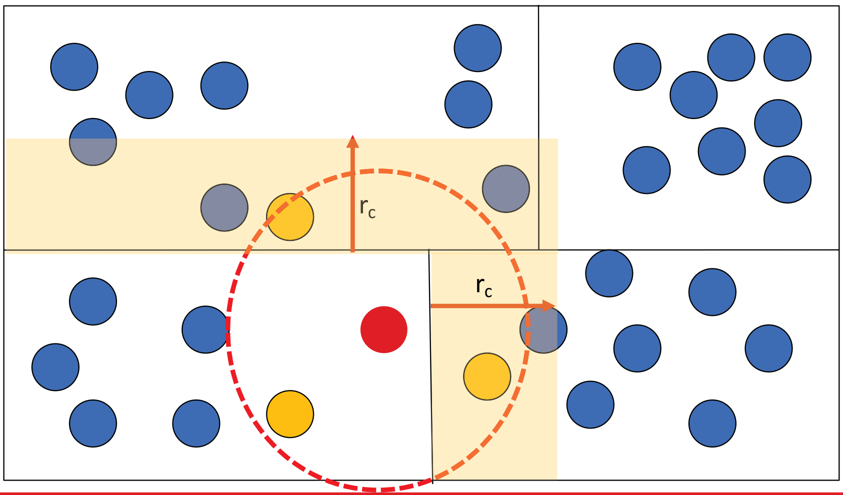
- In order to increase computational efficiency, simulations take advantage of neighbour lists when calculating non-bond interactions



Molecular dynamics

Ghost region

- Processor knows the co-ordinates of atoms within the 'ghost' cut-off region, limits communication between processors.



Molecular dynamics

Summary

→Molecular dynamics used to study dynamic evolution of atoms and molecules in a system over time

→Does this by calculating the forces on each particle and solving newton's second law of motion

→HPC rapidly speeds up these simulations through domain decomposition which parallelises the force calculations

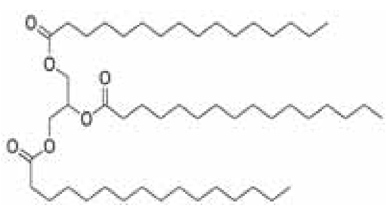
Simulation set-up

	LIPID MOLECULES	SURFACTANT MOLECULES	WATER MOLECULES	FINAL TEMP (K)
LA	229	0	92146	353
SA	229	0	92146	310
SLN	229	650	135031	310

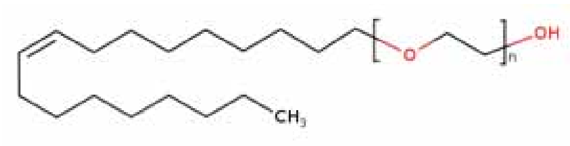
1. 229 lipids placed in a box, heated, left to self-assemble and then cooled.
2. 650 surfactant molecules added to the box and allowed to equilibrate
3. Resulting SLN was analysed.
4. 45 drug molecules added to the box.
5. SLN + drug system was analysed.

KING'S College LONDON

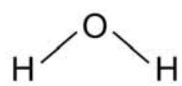
Simulation set-up



Lipid: Tripalmitin



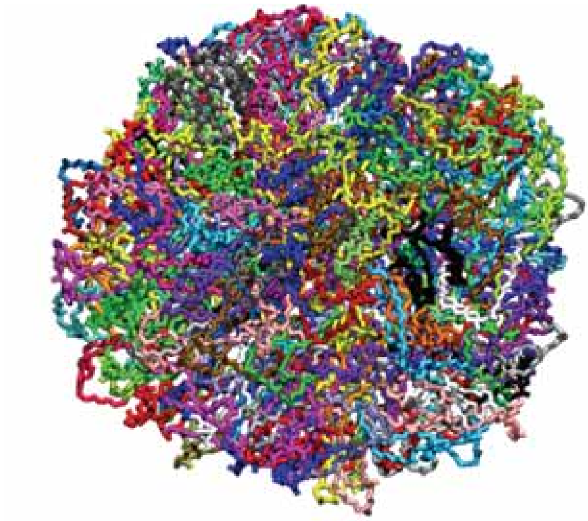
Surfactant: Brij O10



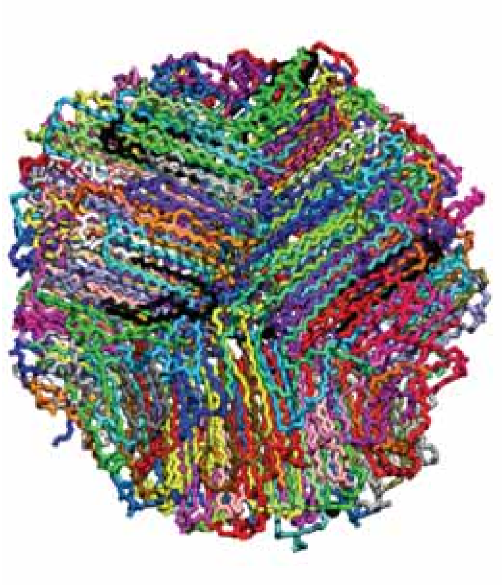
Solvent: Water

KING'S College LONDON

LA & SA

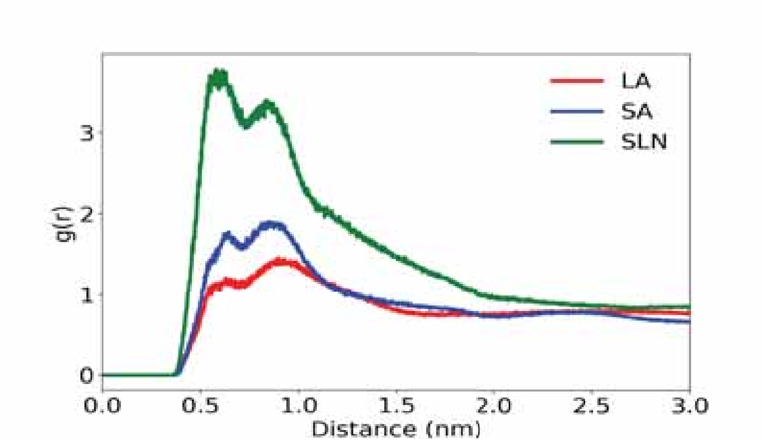


Liquid Aggregate

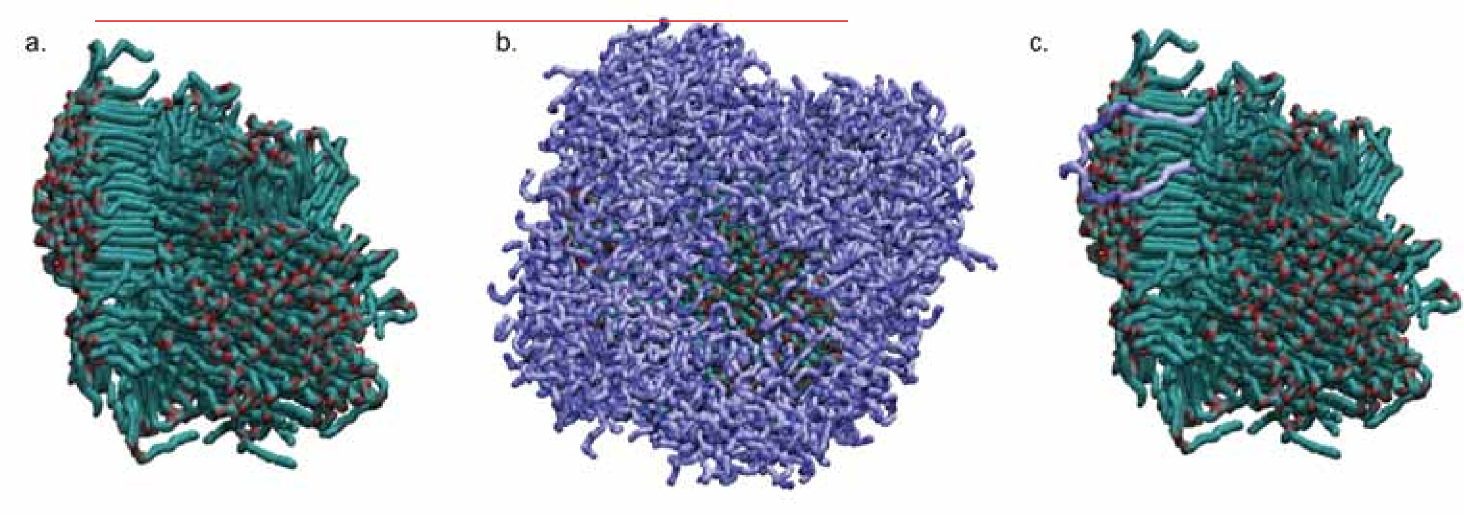


Solid Aggregate

Radial Distribution Function (RDF)

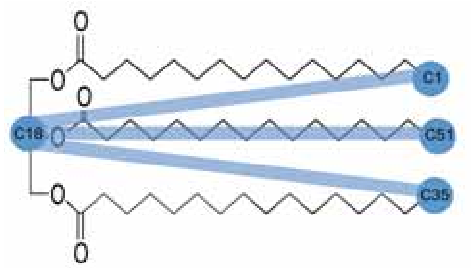


Solid Lipid Nanoparticle

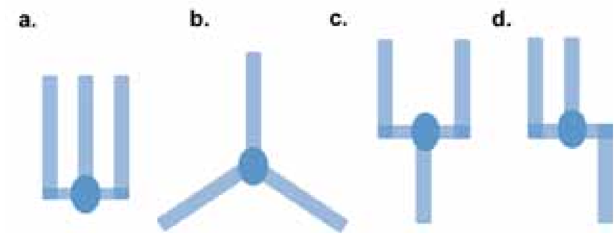


KING'S College LONDON

Lipid Structure



Tripalmitin triglyceride used as the lipid in all simulations. Relevant carbons are labelled and three vectors have been defined.



Lipids are classified as one of 4 conformations.
a. trident, b. propeller, c. tuning fork and d. chair

Propeller : $75^\circ \leq \theta_{C18-C1} \leq 165^\circ$ & $75^\circ \leq \theta_{C18-C35} \leq 165^\circ$ & $75^\circ \leq \theta_{C18-C51} \leq 165^\circ$

Tuning Fork : $0^\circ \leq \theta_{C18-C1} \leq 75^\circ$ & $95^\circ \leq \theta_{C18-C35} \leq 180^\circ$ & $95^\circ \leq \theta_{C18-C51} \leq 180^\circ$

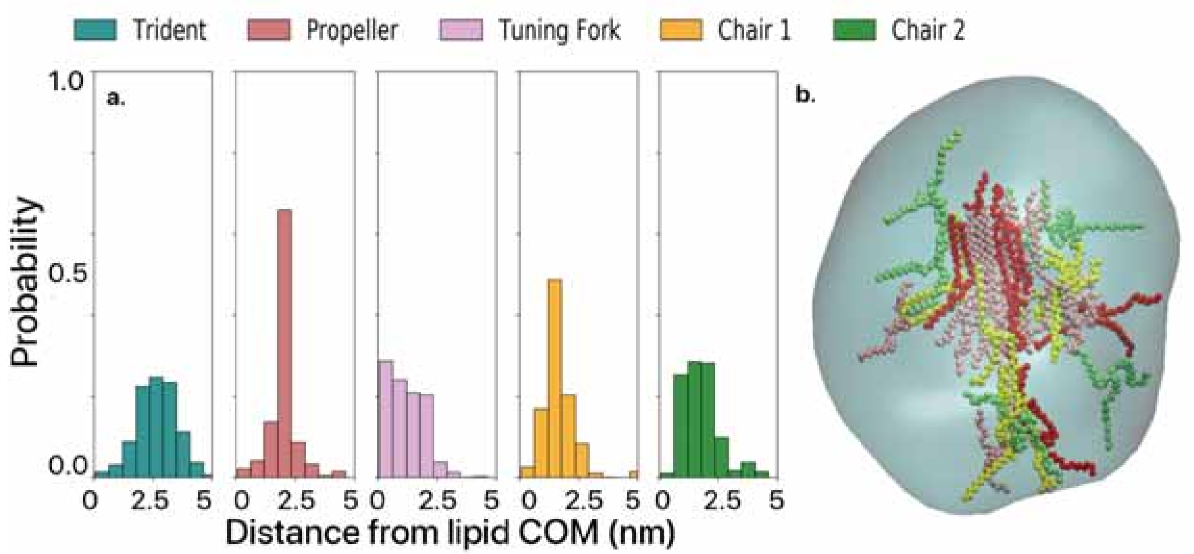
Chair : $95^\circ \leq \theta_{C18-C1} \leq 180^\circ$ & $0^\circ \leq \theta_{C18-C35} \leq 75^\circ$ & $95^\circ \leq \theta_{C18-C51} \leq 180^\circ$

or $95^\circ \leq \theta_{C18-C1} \leq 180^\circ$ & $95^\circ \leq \theta_{C18-C35} \leq 180^\circ$ & $0^\circ \leq \theta_{C18-C51} \leq 75^\circ$

Trident : All other angle combinations

KING'S College LONDON

Lipid Distribution



Angle values

Propeller : $75° \leq \theta_{C18-C1} \leq 165°$ & $75° \leq \theta_{C18-C35} \leq 165°$ & $75° \leq \theta_{C18-C51} \leq 165°$

Tuning Fork : $0° \leq \theta_{C18-C1} \leq 75°$ & $95° \leq \theta_{C18-C35} \leq 180°$ & $95° \leq \theta_{C18-C51} \leq 180°$

Chair : $95° \leq \theta_{C18-C1} \leq 180°$ & $0° \leq \theta_{C18-C35} \leq 75°$ & $95° \leq \theta_{C18-C51} \leq 180°$

or $95° \leq \theta_{C18-C1} \leq 180°$ & $95° \leq \theta_{C18-C35} \leq 180°$ & $0° \leq \theta_{C18-C51} \leq 75°$

Trident : All other angle combinations

Self-organising maps

Self organizing maps are unsupervised artificial neural networks that reduce a series of input vectors and generate a low dimensional representation of the input space.

1. Generate a matrix of a specific size. The weights of nodes in the matrix are initalised.

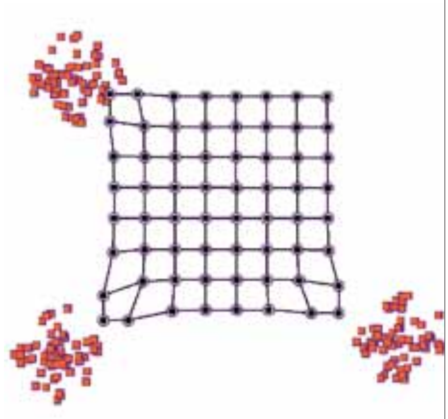
2. A vector is chosen at random from the set of training data.

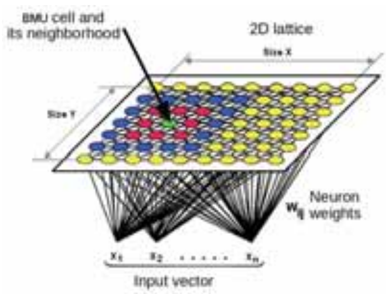
3. Find the similarity between the input vector and the map's node's weight vector, the most similar node is the Best Matching Unit (BMU) and 'wins' the vector.

4. Update the weight vectors of the nodes in the neighbourhood of the BMU (including the BMU itself) by pulling them closer to the input vector.

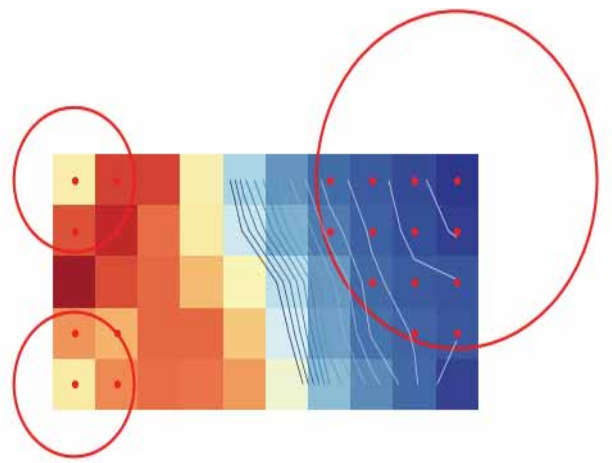
5. The closer a node is to the BMU, the more its weights get altered and the farther away the neighbour is from the BMU, the less it learns.

6.Repeat step 2 for N iterations.

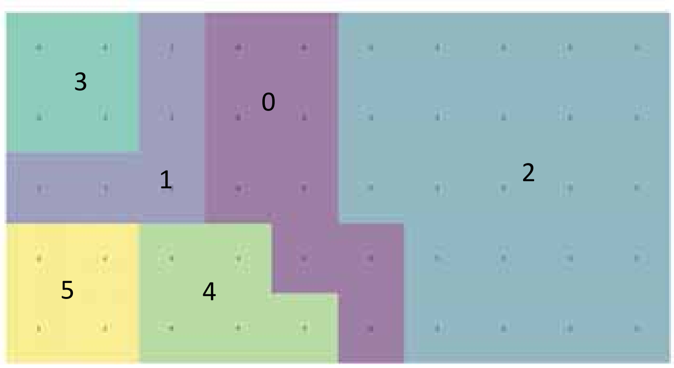




Self-organising maps

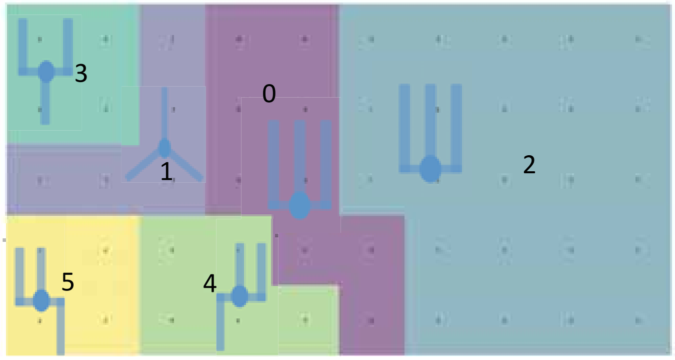


K-means



Self-organising maps

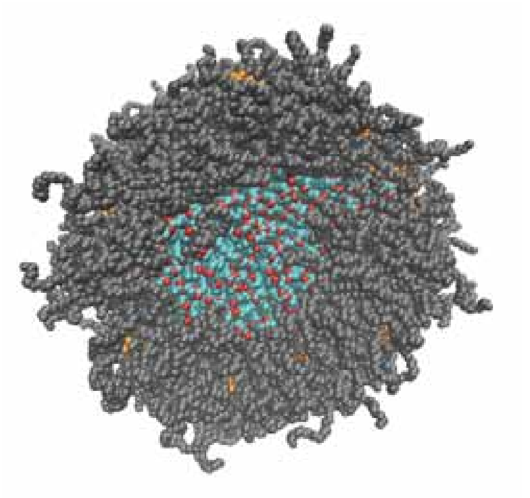
K-means

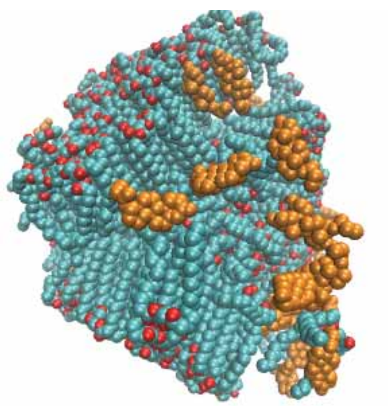


	Propeller	Tuning fork	C1	C2	Trident
Cluster 0	0.00	0.00	0.05	0.00	16.76
Cluster 1	44.80	2.32	0.39	0.00	4.60
Cluster 2	0.00	0.00	0.00	0.00	70.49
Cluster 3	1.94	97.68	0.00	0.03	0.99
Cluster 4	7.97	0.00	0.00	98.09	4.77
Cluster 5	45.29	0.00	99.56	1.88	2.40

KING'S College LONDON

SLN & Drug





KING'S College LONDON

Conclusions

- HPC is vital in molecular dynamic simulations
- Simulations can be used to accurately replicate the properties of solid lipid nanoparticles.
- Self-organising maps can be used to study the structure of flexible molecules
- The method of Solid lipid nanoparticle preparation impacts the localization of drug within the nanoparticle
  → Lipid crystallizes when surfactant is introduced before drug preventing penetration

KING'S College LONDON

Acknowledgements

- Lorenz Lab

Dr Chris Lorenz
Natasha Rhys
Rob Ziolek
Mohamed Al-Badri
Jirawat Assawkhajornsak
Paul Smith
Hrachya Ishkhanyan
Adam Suhaj
Sze May Yee

Prof. Jayne Lawrence –
University of Manchester

archer

MMM MATERIALS AND MOLECULAR MODELLING HUB

KING'S College LONDON

Yvan Fournier

EDF

Readying an Industrial CFD Code for Pre-Exascale

Yvan Fournier obtained a “diplôme d’ingénieur” (equivalent to a Master) in aeronautics in 1994 from École Centrale Paris. He has worked as a researcher at EDF R&D since 1998, and is currently a principal research engineer, working on various HPC, pre and post-processing, and software engineering aspects of EDF’s CFD in-house codes, mainly code_saturne (code-saturne.org). His current interests include in-situ mesh improvement and post-processing, distributed algorithms, software engineering, and high performance computing. Past interests also include CFD modeling of cooling towers and PWR fuel assemblies.

Abstract

EDF make use of CFD for an increasing use of applications, many of which require high resolution and complex geometries, in particular for simulating the flow in a full power plant reactor, with additional expectations for sensitivity analysis and uncertainty quantification. This can only be achieved by developing software for Exascale and optimising it for new architectures and new workflows.

This presentation will show what has been carried out to develop a massively parallel toolchain, from mesh modification to in-situ visualisation and analysis, not forgetting improvements in the resolution of the equations. New discretisation schemes are also being developed in the code and will be outlined, along some application examples.

Evolution of the *Code_Saturne* tool

Readying an industrial CFD code for pre-Exascale

Code_Saturne development team[1]

2019 @ EDF Lab Chatou

[1]Fluid Mechanics, Energy and Environment Department
EDF R&D, Chatou, France
saturne-support@edf.fr

Outlines

EDF R&D Strategic Priorities

CONSOLIDATE AND DEVELOP COMPETITIVE AND ZERO-CARBON PRODUCTION MIXES

- → Consolidate the nuclear assets of the Group and build its future
- → Control and anticipate environmental impacts
- → Contribute to the success of renewable energy projects and prepare tomorrow's technologies
- → Ensure a flexible articulation in the nuclear and renewable mix

DEVELOP AND TEST NEW ENERGY SERVICES FOR CLIENTS

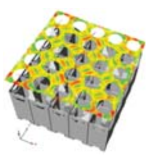
- → Develop new offers for our customers
- → Promote new uses of electricity
- → Develop offers for cities and territories
- → Develop energy efficiency services

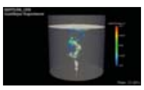
PAVE THE WAY FOR ELECTRIC SYSTEMS OF THE FUTURE

- → Optimize the life of network infrastructure to contribute to the success of smart meter projects
- → Contribute to the success of smart meters project
- → Develop advanced management tools for electrical systems to integrate intermittent energy
- → Develop local energy solutions and integrate into the overall system

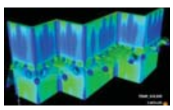
Code Develpment at EDF R&D (1)

- **Computational Fluid Dynamics**
  - general usage single phase CFD, plus specific models
    - *Code_Saturne*
      - property of EDF, open source (GPL)
      - https://code-saturne.org
  - multiphase module, esp. water/steam
    - NEPTUNE_CFD
      - property of EDF/CEA/AREVA/IRSN



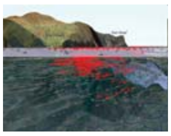


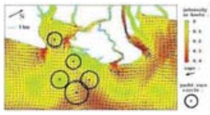
- **Thermal diffusion in solids and radiative transfer**
  - SYRTHES
    - property of EDF, open source (GPL)
    - http://rd.edf.com/syrthes



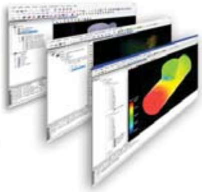
- **Structural Mechanics**
  - General usage
    - *code_aster*
      - property of EDF, open source (GPL)
      - http://www.code-aster.org
  - Rapid dynamics
    - Europlexus
      - property of CEA, EDF

Code Develpment at EDF R&D (2)

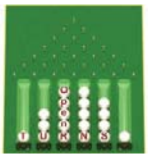




- **Free Surface Flows**
  - TELEMAC system
    - Many partners, mostly open source (GPL, LGPL)
    - http://www.opentelemac.org



- **Integration Platform**
  - SALOME platform
    - CAD, meshing, post-processing, code coupling
    - Utility libraries, Workbench
    - property of EDF/CEA/OpenCascade, open source (LGPL)
    - http://www.salome-platform.org

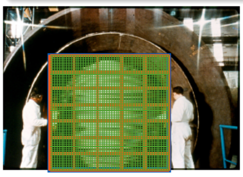


- **Uncertainty and reliabilty analysis**
  - Open TURNS
    - property of EDF/CEA/Phimeca, open source (LGPL)
    - www.openturns.org
- **and many others**
  - Neutronics, electromagnetism, Materials
  - component codes, system codes
  - ...

Three scales of modelling

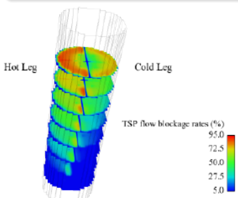
System scale

- $0D$ modelling
- global mass/momentum/energy balances
- correlations
- the boilers, the vessel, ...



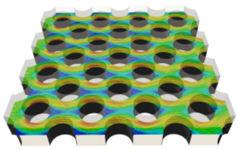
Component scale

- $1D$, $2D$, ($3D$) modelling
- mass / momentum / energy balances of mixture of fluid and solid
- correlations and porous approach for the core or the boilers



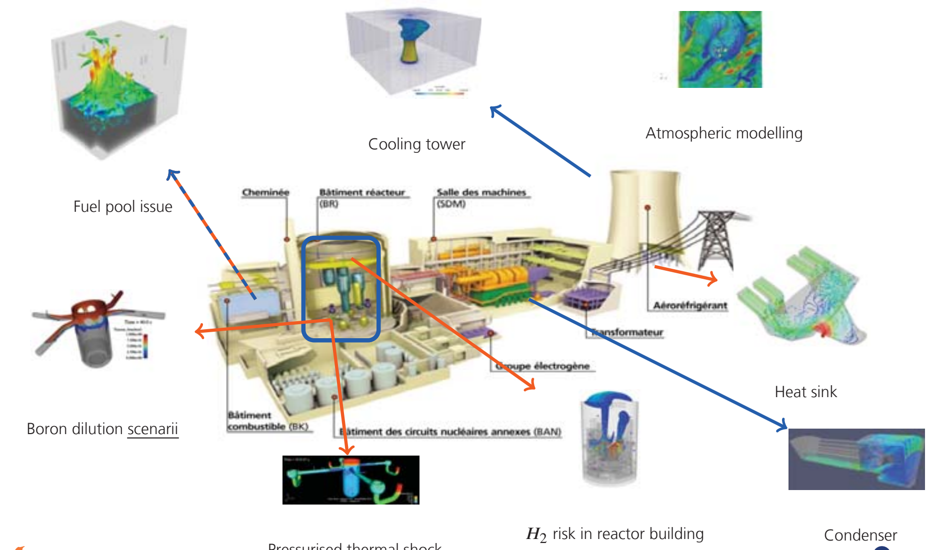
local CFD scale

- $3D$ local modelling
- explicit representation of solids
- local mass/momentum/energy balances of the fluid



Computational Fluid Dynamics in Nuclear Power Plants

Some applications for safety or design



Open-source CFD software : *Code_Saturne*

development under Software Quality Assurance

open-source :
`www.code-saturne.org`

→ Transparency

→ Co-development

→ (Academic) partnerships

→ Linux (workstations to clusters), also Mac and Windows

Software Quality Assurance

→ Production versions every 2 years

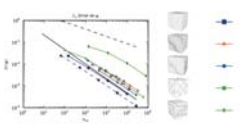
→ Version control with Git - Sources on GitHub

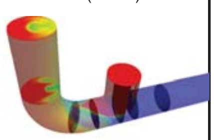
`https://github.com/code-saturne/code_saturne.git`

→ Nightly partial validation

→ 20 verification cases - 1673 runs (v6.0)

→ 67 validation cases - 781 runs (v6.0)





Focus on versioning policy

Release of version X.Y.Z

→ Production version every 2 years (Long Term Support).
Undergoes full Verification and Validation (V&V) procedure

→ Intermediate version every 6 months.
Nightly tests during the development phase ensure code quality.

→ Corrective versions (patches) when needed.
To make sure the users are provided with bug fixes and ports.

Data setup compatibility rules

→ Patches (z version) never break compatibility.
→ GUI parameters file (xml) automatically updated.
→ Manual update of user sources (support in Doxygen).

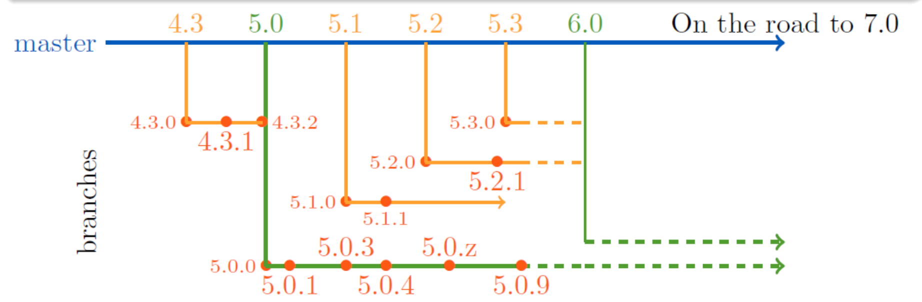
Focus on versioning policy

Release of version X.Y.Z

→ Production version every 2 years (Long Term Support).
Undergoes full Verification and Validation (V&V) procedure

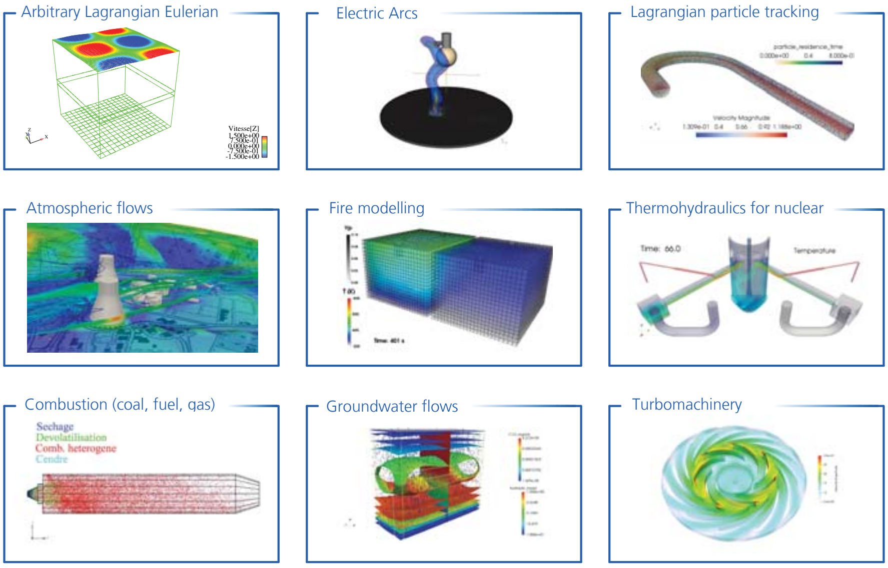
→ Intermediate version every 6 months.
Nightly tests during the development phase ensure code quality.

→ Corrective versions (patches) when needed.
To make sure the users are provided with bug fixes and ports.

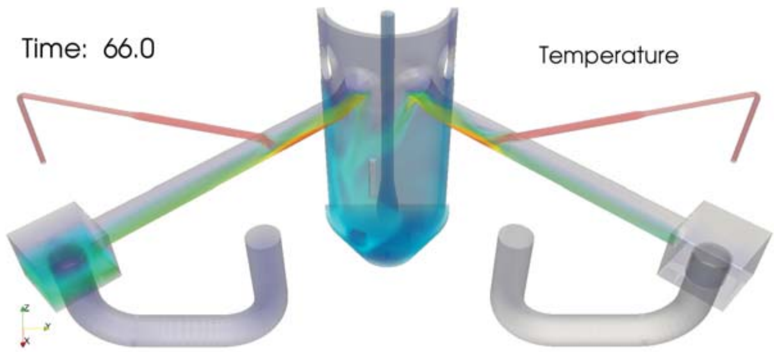


EDF *Code_Saturne* dev. team EDF open-source CFD software [15/65]

Multiphysics solvers gathered in *Code_Saturne*



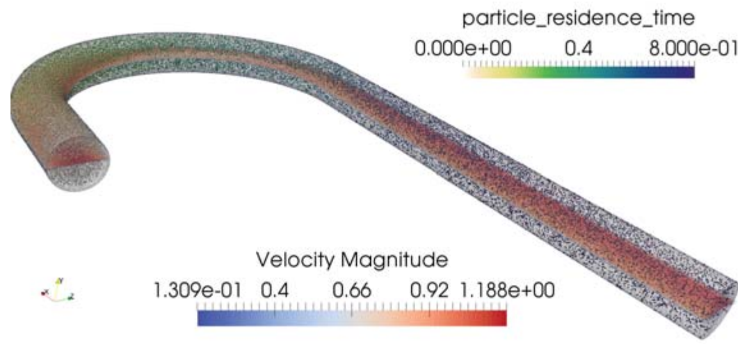
Thermohydraulics for nuclear applications



$$\left\{\begin{array}{r}\dfrac{\partial \rho}{\partial t} + \text{div}\, \rho \underline{\overline{u}} = 0 \\ \dfrac{\partial \rho \underline{\overline{u}}}{\partial t} + \underline{\text{div}}\, \left(\underline{\overline{u}} \otimes \rho \underline{\overline{u}}\right) = -\underline{\nabla} \overline{P} + \underline{\text{div}}\, \left(\mu \left(\underline{\underline{\nabla}}\, \underline{\overline{u}} + \underline{\underline{\nabla}}\, \underline{\overline{u}}^T\right)\right) + \rho \underline{g} - \underline{\text{div}}\, \left(\rho \overline{\underline{u}' \otimes \underline{u}'}\right)\end{array}\right.$$

$$C_p \left(\frac{\partial \rho \overline{T}}{\partial t} + \text{div}\, \left(\overline{T} \rho \underline{\overline{u}}\right)\right) = \text{div}\, \left(\lambda \underline{\nabla} \overline{T}\right) - C_p \text{div}\, \left(\rho \overline{T' \underline{u}'}\right)$$

Lagrangian particle tracking



- → Simulation of polydispersed particle-laden turbulent flow
- → Moments/PDF (Euler/Lagrange) approach
- → Frozen field, one-way or two-way coupling
- → Dedicated models for particle heat transfer, droplets evaporation, particle deposition

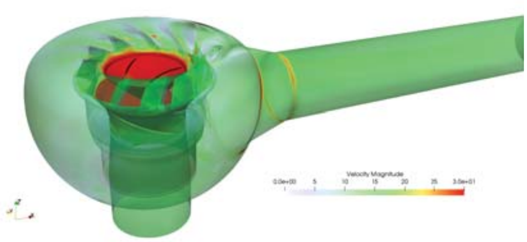
$$d\underline{X}_p = \underline{U}_p \, dt$$

$$d\underline{U}_p = -\frac{1}{\rho_f}\nabla \overline{P}_f \, dt - \frac{\underline{U}_p - \overline{\underline{U}}_f}{T_L} dt + \sqrt{C_0 \varepsilon_f}\, d\underline{W}$$

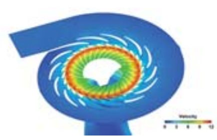
where

$$T_L = \frac{1}{\frac{1}{2} + \frac{3}{4} C_0} \frac{k_f}{\varepsilon_f}$$

Turbomachinery



Temperature at a safety pump suction



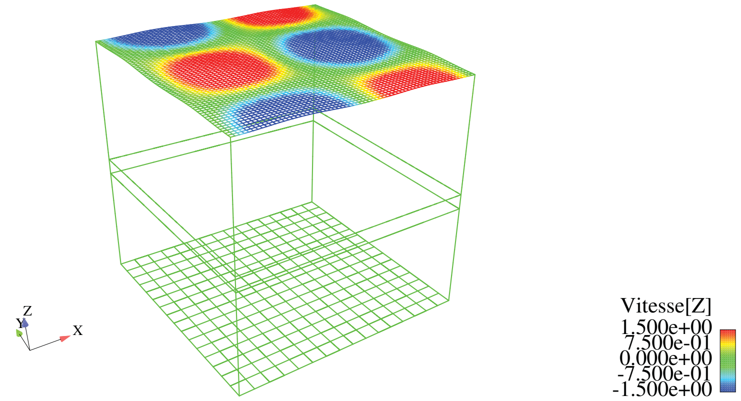
Flow field in a Francis99 turbine

Q-criteria around the MEXICO wind turbine

Application examples

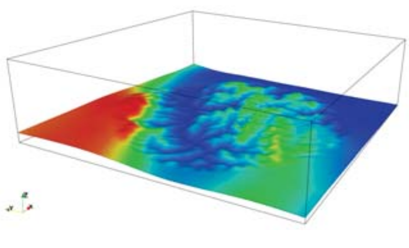
- → Reactor Coolant Pump : performance studies, hydraulic loads as input of mechanical calculations (*code_aster*)
- → Safety pumps : thermal transient, lagrangian particle tracking toward guiding and sealing systems
- → Renewable energy : hydraulic turbines, wind turbines

Arbitrary Lagrangian Eulerian (ALE)

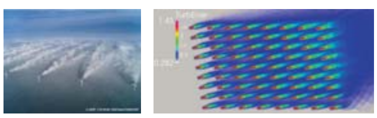


- → Solve mesh displacement with mesh deformation imposed at the boundaries
- → Or impose the displacement of any node
- → Fluid-structure interaction
- → Free surface modelling (no wave break)

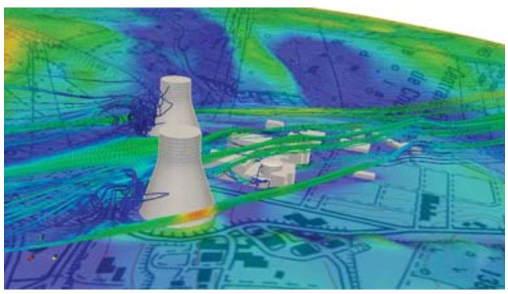
Atmospheric flows



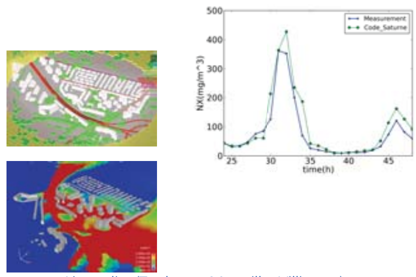
Wind potential estimates on complex terrain



Wake effects on an offshore wind farm (WRAPP)



Atmospheric dispersion, and high-speed winds

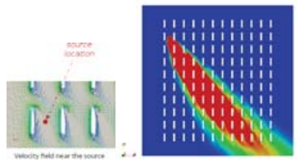


Air quality (Toulouse, Marseille, Villiers,...)

Lagrangian stochastic modelling

pollutant atmospheric dispersion

MUST (Mock Urban Setting Test) campaign (idealized city)

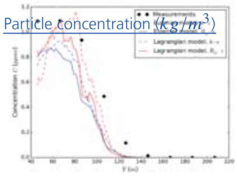


Model equations (Simplified Langevin model, cf. Pope 2000)

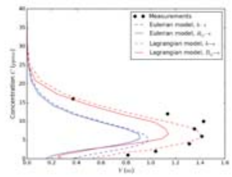
$$d\underline{X}_p = \underline{U}_p \, dt$$
$$d\underline{U}_p = -\frac{1}{\rho_f} \nabla \overline{P_f} \, dt - \frac{\underline{U}_p - \overline{\underline{U}_f}}{T_L} dt + \sqrt{C_0 \varepsilon_f} \, d\underline{W}$$

where

$$T_L = \frac{1}{\frac{1}{2} + \frac{3}{4} C_0} \frac{k_f}{\varepsilon_f}$$



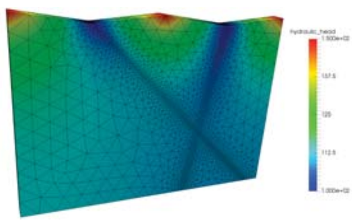
Particle concentration (kg/m^3)



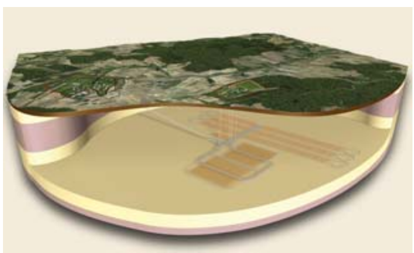
Concentration profiles (ppmv)

Groundwater flows

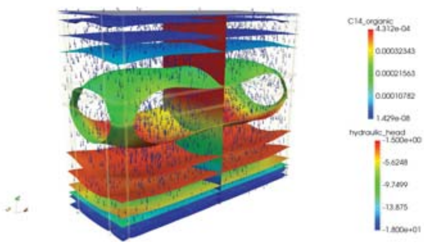
- → Richards equation is solved (Darcy law injected in mass equation)
- → Species mass fractions transport
- → Heterogeneous and anisotropic permeability
- → Large meshes : several hundred million cells
- → Possible long physical period : up to a million years (1 time step ≈ 1000 years)



2 geological fractures (lower permeability)



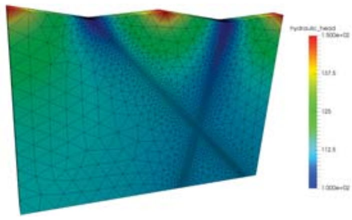
source www.andra.fr, CIGEO project



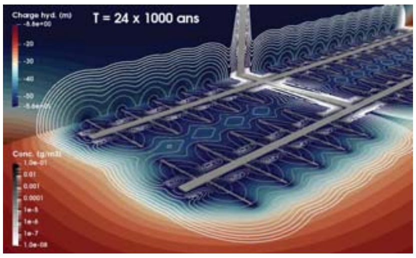
Iso-surfaces of C^{14} on a storage site

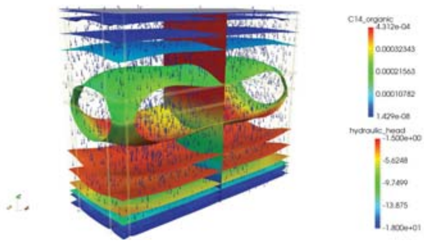
Groundwater flows

- → Richards equation is solved (Darcy law injected in mass equation)
- → Species mass fractions transport
- → Heterogeneous and anisotropic permeability
- → Large meshes : several hundred million cells
- → Possible long physical period : up to a million years (1 time step ≈ 1000 years)



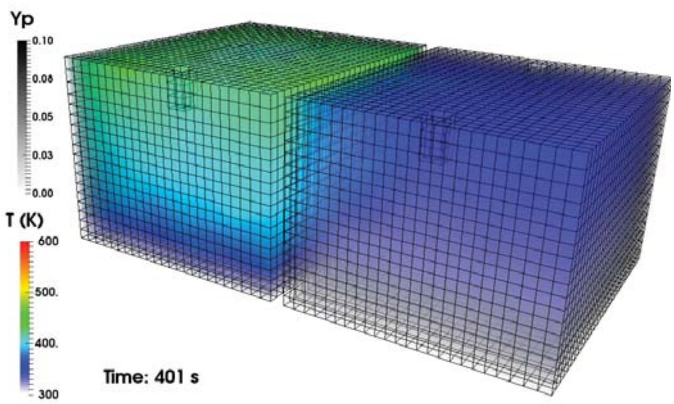
2 geological fractures (lower permeability)





Iso-surfaces of C^{14} on a storage site

Fire modelling

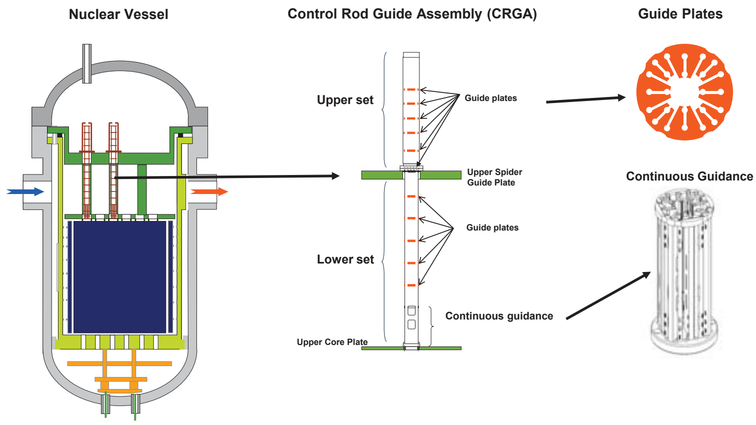


Dodecane fire

- → Weakly compressible algo. for gas mixture
- → Free inlet
- → Soot models
- → Radiative transfer models (DOM and P1)

Application : Turbulent Flow through PWR guide card

Context



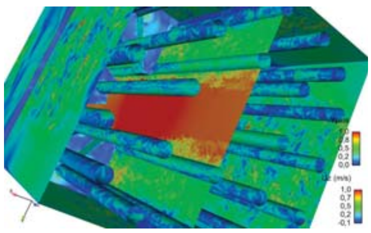
Observed fact : control rod vibration with large displacements

→ No fluid structure interaction for the moment

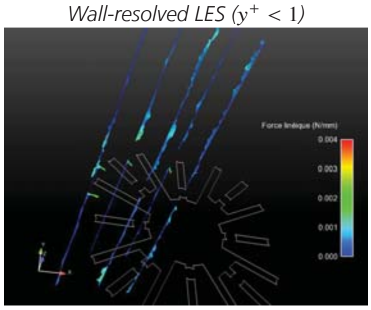
Application : Turbulent Flow through PWR guide card

Numerical set-up

- → Wall-resolved LES on one plate
- → Re = 10,000 (bulk velocity $0.2 m/s$)
  (Industrial Reynolds number : 400,000)
- → 1.1 billion cells (450x28 Intel Xeon E5 2.4Ghz cores, 12 million CPU hours)
- → Mesh extruded from 2 meshes with the extrusion option available in *Code_Saturne* (3 minutes on 50 to 300 cores)
- → 1 million time-steps, of which 400,000 time-steps are provided for mechanical calculations
- → 500 Gb of `ASCII` data



Wall-resolved LES ($y^+ < 1$)



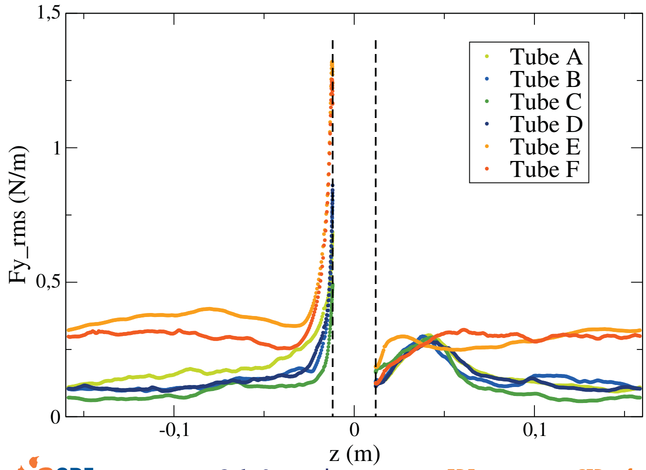
Linear force calculated at each time-step

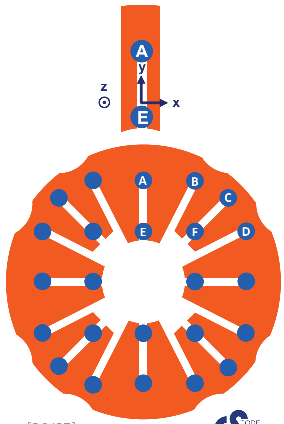
Application : Turbulent Flow through PWR guide card

Some insights on the hydraulic forces

Fluctuating pressure force according to rod location

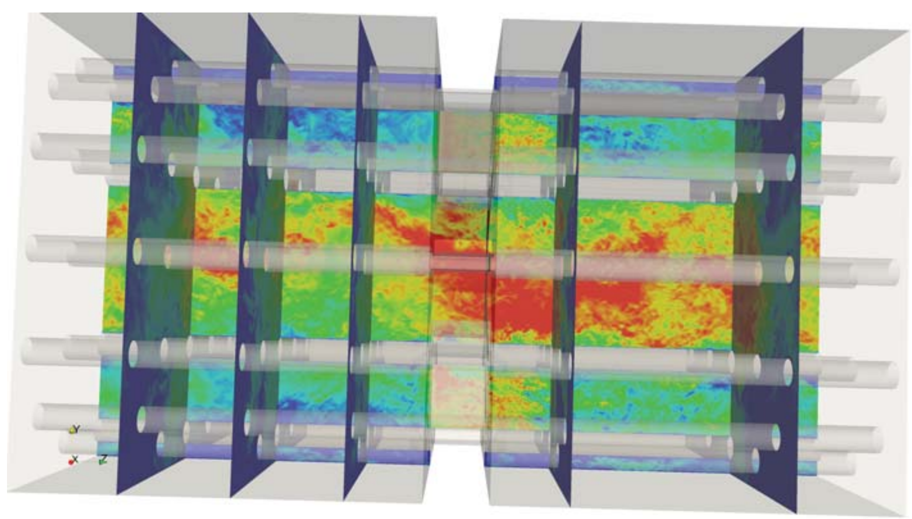
- → Rods E and F can be distinguished from others by a higher level of fluctuations
- → This is coherent with on-site observations where the highest level of wear is observed for rods E and F





High fidelity simulation to get dynamic load on structures

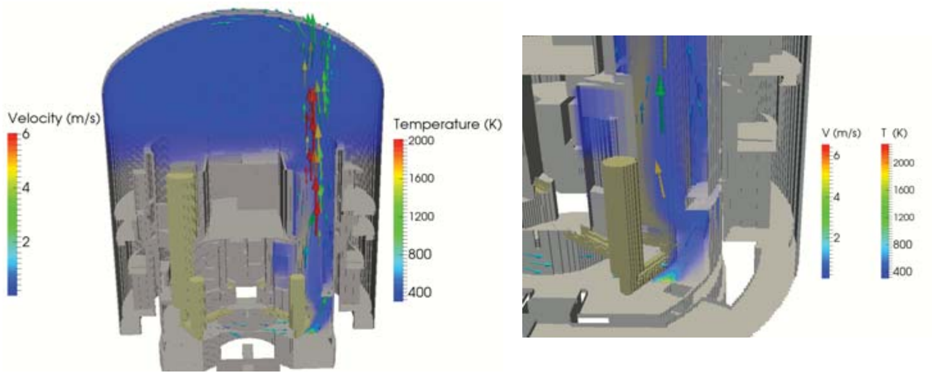
Billion cell LES to get pressure load on rods : on 450×24 proc



Industrial study

Modelling of a fictitious fire in the EPR Reactor Building

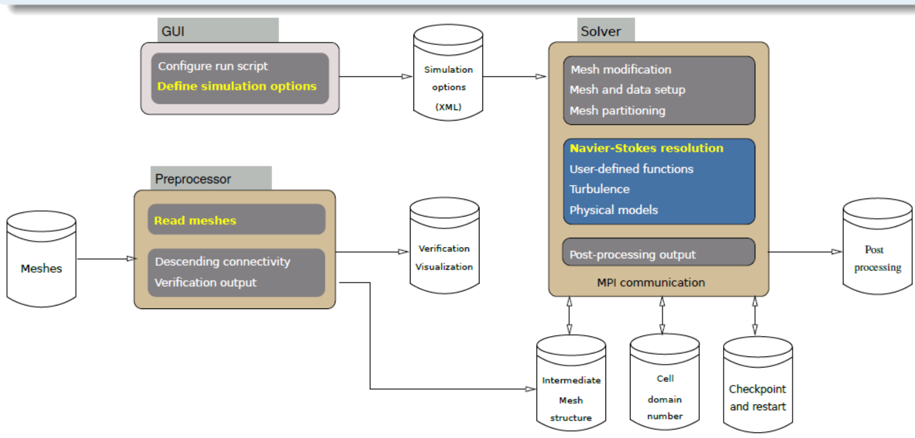
- $50L$ oil pool fire at Reactor Coolant Pumps (RCP) at the bottom of the RCP
- 250 targets studied : cables, captors, electrical cabinets, valves, doors, ...
- $20cm$ cells (10 M cells)
- 1 day of calculation on 392 cores



Code_Saturne toolchain

Reduced number of tools

- Managed by a common `code_saturne` Python script
- Natural separation between interactive and potentially long-running parts
  - Front-end (GUI + Preprocessor) and back-end (solver) designed to separate serial-only and parallel parts
  - Preprocessor could be moved or added to back-end to separate “light” and “heavy” parts of execution.



Hybrid parallelism

Two-level parallelism...

... to take a better benefit of modern CPU architecture :

→ Distributed memory based on **MPI**

→ Shared memory based on **OpenMP**

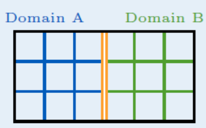
Parallelism (distributed memory) and periodicity

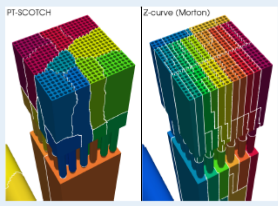
BASED ON DOMAIN PARTITIONING USING MPI

Domain / Mesh partitioning :

→ external libraries : **METIS, SCOTCH**

→ internal **space-filling curve** algorithm (Hilbert, Morton).



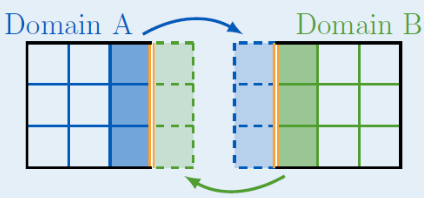


Communications between sub-domains :

Classical method using **ghost cells** :

→ sharing **faces** with other domains cells

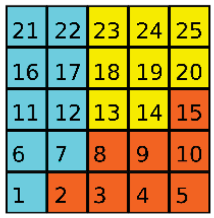
→ sharing **vertices** with other domains cells (extended neighborhood for gradients)

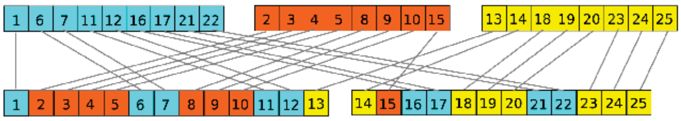


TRUE GEOMETRIC PERIODICITY based on SAME method (NOT a boundary condition !)

Global numbering : basics

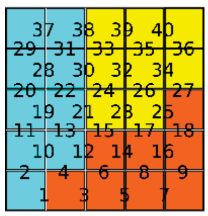
- Use of global numbering
  - We associate a global number to each mesh entity
    - A specific C type (cs_gnum_t) is used for this. An unsigned long integer (64-bit) is necessary for larger meshes
    - Currently equal to the initial (pre-partitioning) number
- Allows for partition-independent single-image files
  - Essential for restart files, also used for postprocessing output
  - Shared file MPI-IO possible does not require indexed datatypes
- Redistribution on n blocks
  - n blocks ≤ n cores, block size and stepping may be adjusted for performance or constraints
  - Inefficient for halo exchange, but allow for simpler data structure related algorithms with deterministic performance bounds
    - Owning rank determined simply by global number, allows rendez-vous type algorithms
    - Similar to "assumed partition" algorithm

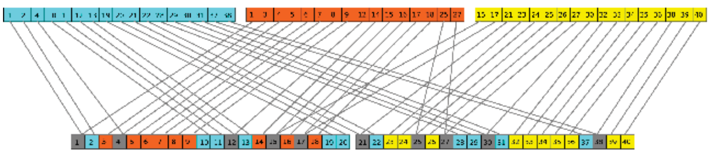




Global numbering : matching

- Conversely, simply using global numbers allows reconstructing neighbor partition entity equivalents mapping
  - Allows automatic identification of "Interfaces"
  - Matching between faces or vertices on parallel boundaries
  - Used for parallel ghost cell construction from initially partitioned mesh with no ghost data
- Switch from one representation to the other currently uses `MPI_Alltoall` and `MPI_Alltoallv`, but we may switch to a more "sparse" algorithm such as CrystalRouter
  - Not an issue under 16000 cores, not critical at 64000





All to all algorithms : legacy

- Redistribution example
  - some parts replaced by wrappers or utility functions in real code

```
block_size = global_count / size;
if (global_count % size > 0)
  block_size += 1;
send_count = malloc(n_elts*sizeof(int));
recv_count = malloc(n_elts*sizeof(int));
send_shift = malloc(n_elts*sizeof(int));
recv_shift = malloc(n_elts*sizeof(int));

/* Count number of values to send to each process */
/*------------------------------------------------*/

for (rank = 0; rank < size; rank++)
  send_count[rank] = 0;

for (i = 0; i < n_elts; i++)
  send_count[(global_num[i] - 1) / block_size] += 1;

MPI_Alltoall(send_count, 1, MPI_INT, recv_count, 1, MPI_INT,
             comm);

send_shift[0] = 0;
recv_shift[0] = 0;

for (rank = 1; rank < size; rank++) {
  send_shift[rank] = send_shift[rank - 1] + send_count[rank -1];
  recv_shift[rank] = recv_shift[rank - 1] + recv_count[rank -1];
}

n_ent_recv = recv_shift[size - 1] + recv_count[size - 1];

recv_global_num = malloc(n_ent_recv*sizeof(cs_gnum_t));
recv_order = malloc(n_ent_recv*sizeof(cs_lnum_t));

MPI_Alltoallv(new_global_num, send_count, send_shift, CS_MPI_GNUM,
              recv_global_num, recv_count, recv_shift, CS_MPI_GNUM,
              comm);
```

```
/* Do work */
…

/* Return reverse (processed) info */

MPI_Alltoallv(recv_global_num, recv_count, recv_shift, CS_MPI_GNUM,
              new_global_num, send_count, send_shift, CS_MPI_GNUM,
              comm);

/* Free memory */

free(recv_order);
free(recv_global_num);
free(send_count);
free(recv_count);
free(send_shift);
free(recv_shift);
```


All to all algorithms : API

- Redistribution example
  - data movement is hidden
    - allows runtime choice of algorithm (MPI_Alltoall, Crystal Router)
    - loses some count (previously MPI_Alltoall-based) reuse opportunities

```
block_size = global_count / size;
if (global_count % size > 0)
  block_size += 1;

dest_rank = malloc(n_elts*sizeof(int));

for (i = 0; i < n_elts; i++)
  dest_rank[i] = (global_num[i] - 1) / block_size;

/* Create distributor and associate data elements */

d = cs_all_to_all_create(n_part_elts,
                         0,          /* flags */
                         NULL,       /* dest_id */
                         dest_rank,
                         comm);

cs_all_to_all_transfer_dest_rank(d, &dest_rank);

cs_gnum_t *b_data
  = cs_all_to_all_copy_array(d,
                             CS_GNUM_TYPE,
                             1,
                             false,  /* reverse */
                             new_global_num,
                             NULL);
```

```
/* Get number of elements in "receiving" distribution */

n_block_elts = cs_all_to_all_n_elts_dest(d);

/* Do work */
…

/* Return reverse (processed) info */

cs_all_to_all_copy_array(d,
                         CS_GNUM_TYPE,
                         1,
                         true,  /* reverse */
                         b_data,
                         new_global_num);

BFT_FREE(b_data);

cs_all_to_all_destroy(&d);
```

All to all algorithms : performance

- the new all to all API is not completely deployed yet
  - about 80 places in the code where this needs to be done
  - work in progress, requires caution (60%done)
- sample results on Blue Gene/Q (using preliminary API)
  - using 12.8 million cell benchmark test

n_ranks	Alltoall(v) time (s)	CrystalRouter time (s)
128 (16x8)	0.03	0.45
256 (16x16)	0.02	0.26
512 (32x16)	0.012	0.20

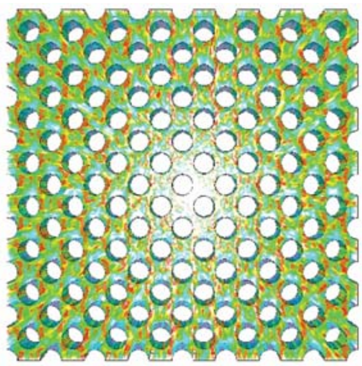
- 2019 results on Intel Xeon cluster with OFA
  - using 725 milion cell LES case

n_ranks	Alltoall(v) time (s)	CrystalRouter time (s)
10500 (300x35)	809,9 (496 calls)	
12250 (350x35)		348,7 (492calls)

Scalability of *Code_Saturne*

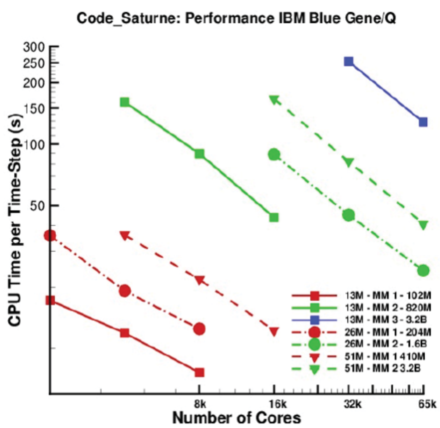
→ Scalability as a function of mesh size

At 65 000 cores and 3,2 billion cells, about 50 000 cells / core



Experiment of Simonin and Barcouda

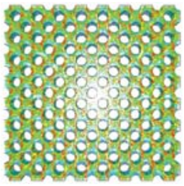
2-D section : 100,040 cells ; 3rd direction : 128 layers -> 13M cells

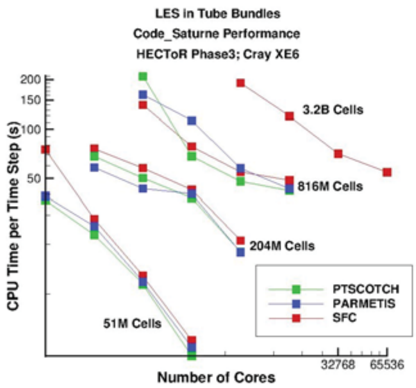


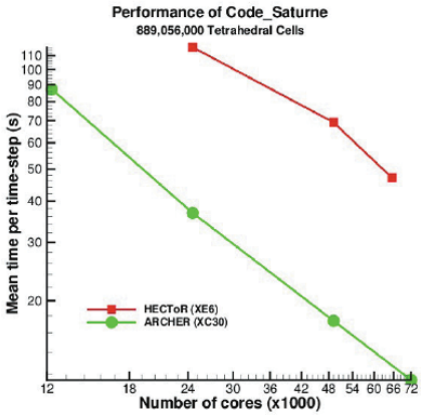
Scalability of *Code_Saturne*

→ Scalability as a function of mesh size, here, comparing partitioning options

→ Best scalability usually observed on Blue Gene and Cray machines (degrades faster on typical machines)



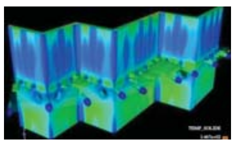




Parallel code coupling

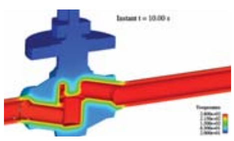
→ **Parallel n to p coupling** using "Parallel Location and Exchange" sub-library

- Uses MPI
- Successor to / refactoring of FVM (also used in CWIPI)
- Core communication in PLE, rest moved to code
- Fully distributed (no master node requiring global data)
- Also used by BSC's ALYA code)



→ SYRTHES (conjugate heat transfer)

- Coupling with (parallel) SYRTHES 4 on industrial study
- Scaling degrades above a few hundred MPI ranks, so box-tree neighbor search from mesh joining will be used in the future



→ *Code_Saturne*/ *Code_Saturne* coupling

- RANS / LES

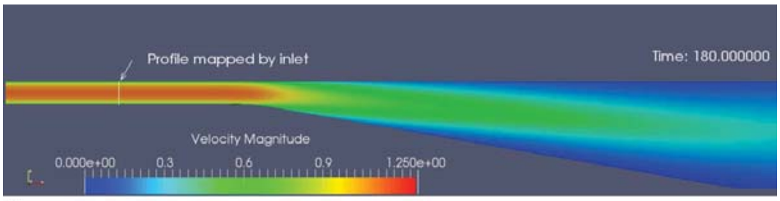
  Different turbulence models and time stepping fixed overlapping domains
- Turbo machinery (alternative to mesh joining method)

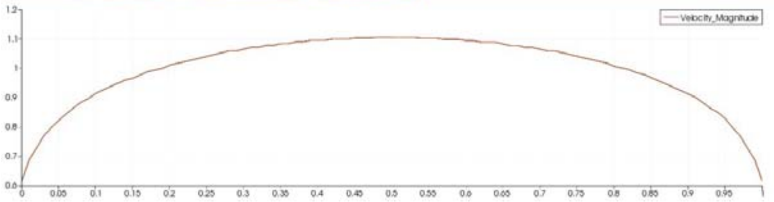
  Same turbulence model and time step, moving sub domain

→ Other coupling modes : MedCoupling support

Parallel code coupling

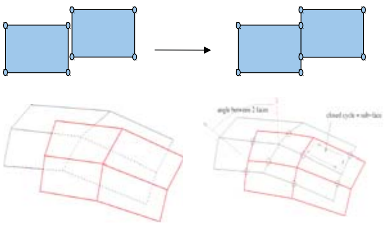
→ Coupling with self also allows "mapped inlet boundary conditions"
- point values at inlet mapped to cells inside domain
  allows good profiles even with short inlets
  several rescaling options available
- works in parallel
  partition-independent
- Fully distributed (no master node requiring global data)

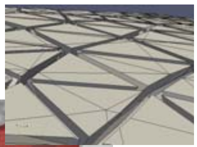


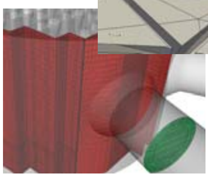


Parallel mesh joining

- Build a distributed global face visibility map
  - Build distributed octree-like structure of face bounding boxes
    - Built bottom-up
    - Coarser tree built first for load balance
  - Faces may intersect if bounding boxes do
- Redistribute faces based on their global numbers
  - Copies of faces visible to a given face also sent to its owning rank
- Determine intersections of face edges
  - Subdivide edges along those intersections
- Merge vertices
  - Build "chains" of vertices within merging distance, reduce local merging distance if this leads to excessive merging (subdiving chains), then merge all vertices in a same chain
  - All ranks must take the same decision regarding merging a shared vertex
- Reconstruct sub-faces
  - Close shortest loops of edges on approximate face surfaces
  - Merge identical sub-faces: 2 boundary faces with 1 adjacent cell merge into interior faces with 2 adjacent cells

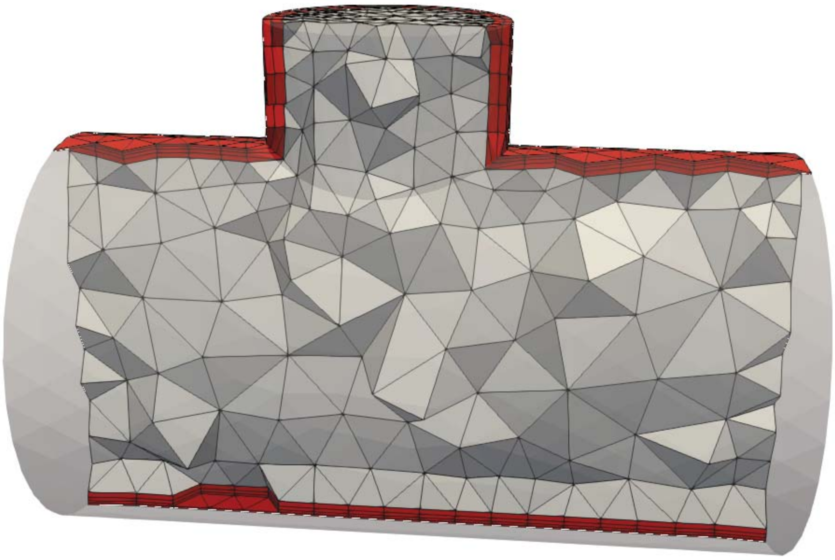






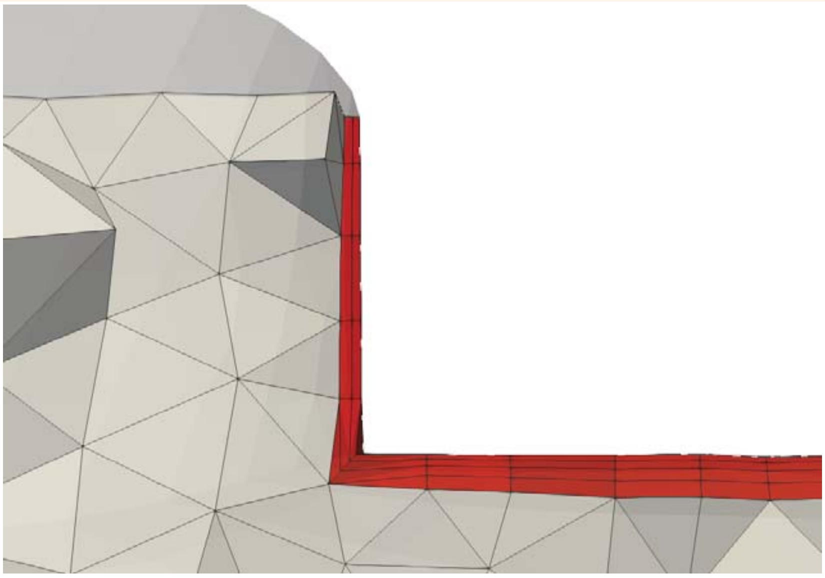
Preprocessing : automatic insertion of wall-layer cells

using CDO vertex based ALE solver



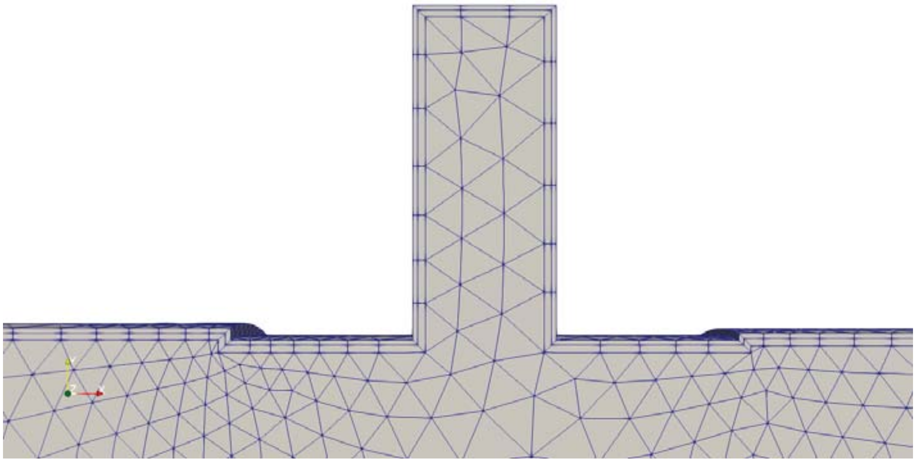
Preprocessing : automatic insertion of wall-layer cells

using CDO vertex based ALE solver



Preprocessing : automatic insertion of wall-layer cells

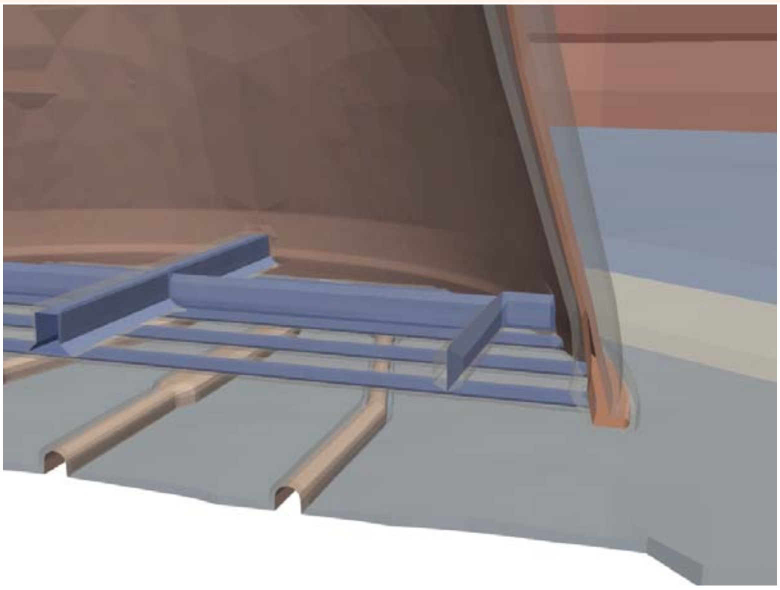
using CDO vertex based ALE solver



see user source file `cs_user_mesh-modify.c`.

Preprocessing : automatic insertion of wall-layer cells

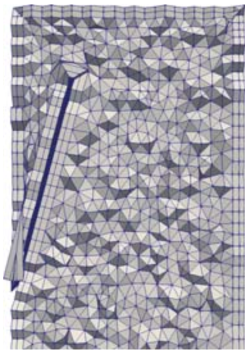
using CDO vertex based ALE solver

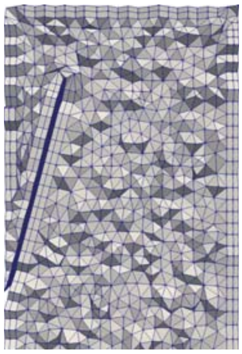


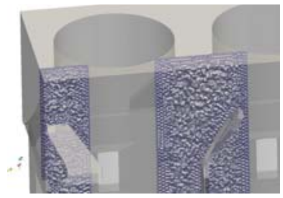
Preprocessing : automatic insertion of wall-layer cells

Fix features in sharp angles

A solution is to test for negative volumes while deforming the mesh, and locally limit the extrusion on adjacent boundaries (removing one extrusion layer at vertices of those cells). This is done iteratively until no negative volume cells are produced.







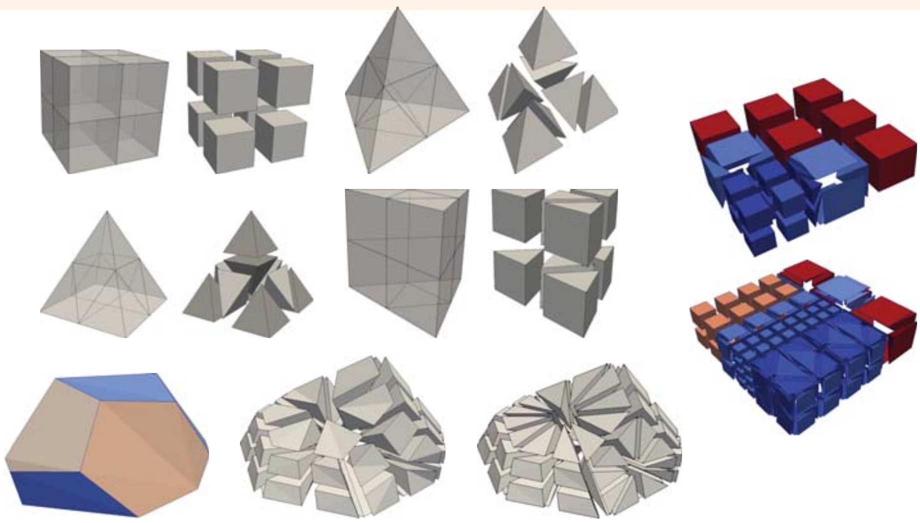
CALIFS

before after

Also add optional cell volume ratio limiter to reduce the extrusion near cells that would be excessively flattened or entangled.

Preprocessing : Add mesh refinement engine

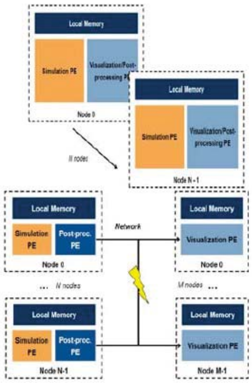
for any polyhedral, load balancing currently handled through complete repartitioning, in collaboration with STFC

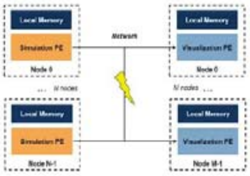


see function `cs_user_mesh-modify.c`.

In-situ postprocessing

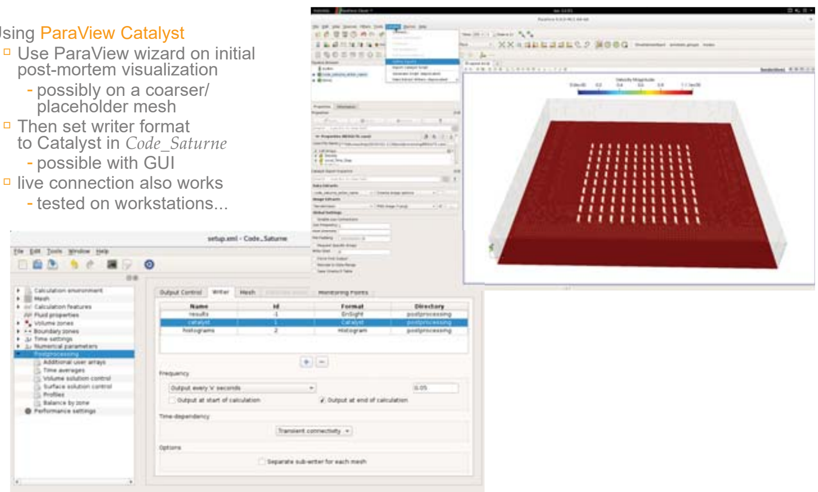
- Compute power progresses faster than storage
  - Avoid storing unnecessary data
- In-situ often used as an “umbrella” term for in-situ/in-transit/coprocessing
  - tightly coupled vs loosely coupled approaches
- Several in-situ visualization or postprocessing tools exist
  - Catalyst (from ParaView)
  - libsim (from VisIt)
    - also VTK-based, well established
  - ADIOS (more a staging system for “in-transit” simulations)
  - Sensei (very similar API to Catalyst, sort of “Generalization” allowing use of at least Catalyst, libsim, ADIOS
  - ASCENT (next generation, lightweight)
- Many other more specialized tools are being developed
  - Most publications seem to be about visualization





In-situ postprocessing : usage

- Using ParaView Catalyst
  - Use ParaView wizard on initial post-mortem visualization
    - possibly on a coarser/ placeholder mesh
  - Then set writer format to Catalyst in *Code_Saturne*
    - possible with GUI
  - live connection also works
    - tested on workstations...

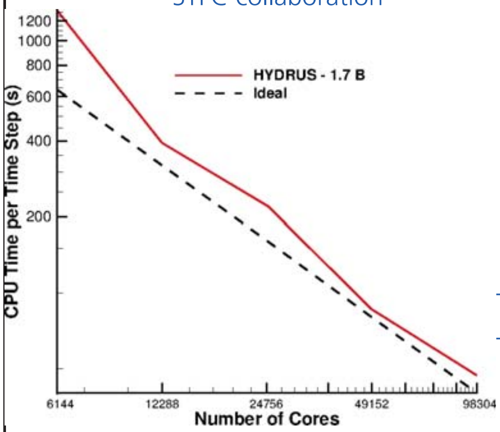


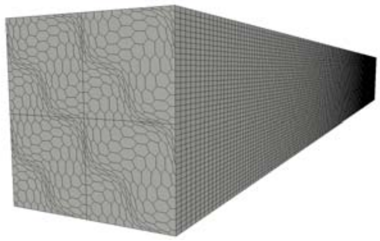
Stress tests on HPC capabilities of CDO schemes

Groundwater flow applications

ARCHER CRAY supercomputer (UK)

STFC collaboration





→ 1.7 Billion polyhedral cell mesh

→ Nearly optimal speedup up to 89% of the machine

☺ Enhanced two-level MPI/OpenMP parallelism sucessfully tested on Intel Xeon Phi MIC (Many Integrated Core)

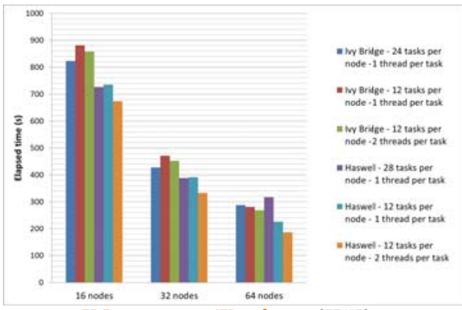
Hybrid parallelism : basics

- Since 2015, hybrid parallelism using OpenMP is in a working state
  - built by default since version 5.0 (2017)
  - Requires mesh renumbering
    - Also useful for cache behavior
    - Cache effects even more important under OpenMP, to avoid false sharing, but also try not to saturate bandwidth
    - Both internal (in progress) and external (IBM library) renumberings are possible, and may be compared
  - some subsets of the code have better OpenMP scaling
  - some subsets do not use OpenMP

- OpenMP debugging still very painful
  - Valgrind DRD or clang or gcc ThreadSanitizer help
    - very high overhead for DRD
    - requires compiling gcc with specific option (-disable-linux-futex)
    - worse with recent versions of gcc (false positives between threaded sections, probably due to OpenMP not keeping a thread pool instead of forking/joining for better performance)
    - Archer project (based on clang) might help
  - Writing C code with loop local variable definitions helps avoid missing “private” qualifiers
    - no equivalent in Fortran to our knowledge
    - alternative would be to use a tasking/data model more similar to MPI, but this would be fragile

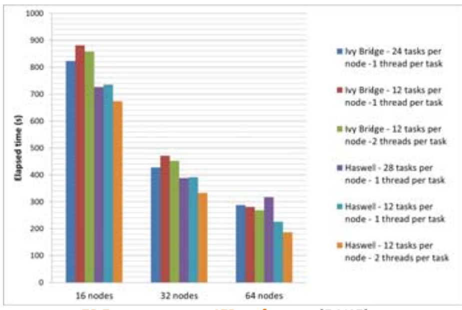
Hybrid parallelism : performance

- Bandwidth is the main limitation (mostly memory-bound performance characteristics)
- On several clusters, best results observed with 2 threads per task, MPI for all the rest
  - example on 51-million cell tube bundle test case
  - recommandation: test with both 1 and 2 OpenMP threads
    - use more threads only in case of MPI memory usage issues when memory per core is limited
  - 2 x Intel Xeon® Processor E5-2680 v4 (Haswell) /node: 14 cores (28 HT), 2 QPI links, 4 channels
    - at 24 tasks/node, 3 tasks/channel; at 28 tasks/node, 3 or 4 tasks/channel
  - Since 2016, with other performance improvements in parts of the code which had good OpenMP scalability, portion of non-OpenMP code has increased, and no performance benefit is observed anymore
    - except for CDO algorithms, where coding is more thread-friendly, and scaling is good at over 4 threads per task



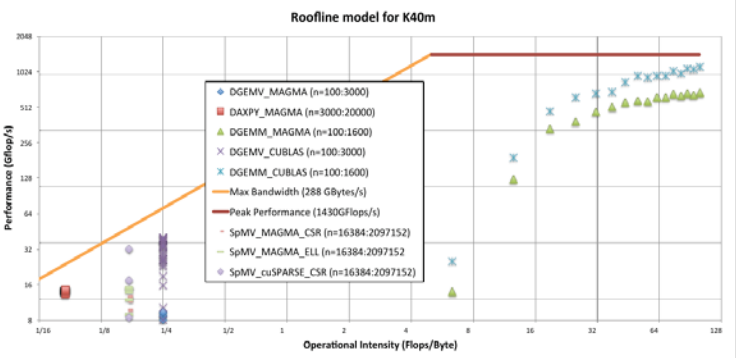
Hybrid parallelism : performance

- Bandwidth is the main limitation (mostly memory-bound performance characteristics)
- On several clusters, best results observed with 2 threads per task, MPI for all the rest
  - example on 51-million cell tube bundle test case
  - recommandation: test with both 1 and 2 OpenMP threads
    - use more threads only in case of MPI memory usage issues when memory per core is limited
  - 2 x Intel Xeon® Processor E5-2680 v4 (Haswell) /node: 14 cores (28 HT), 2 QPI links, 4 channels
    - at 24 tasks/node, 3 tasks/channel; at 28 tasks/node, 3 or 4 tasks/channel
  - Since 2016, with other performance improvements in parts of the code which had good OpenMP scalability, portion of non-OpenMP code has increased, and no performance benefit is observed anymore
    - except for CDO algorithms, where coding is more thread-friendly, and scaling is good at over 4 threads per task



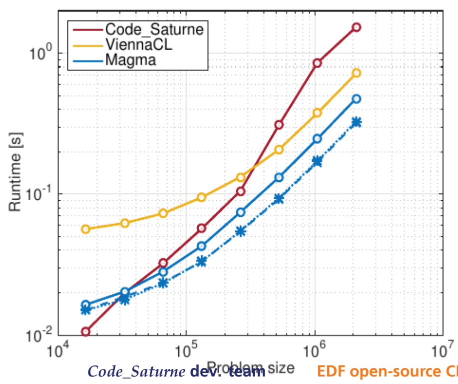
Hybrid parallelism : roofline

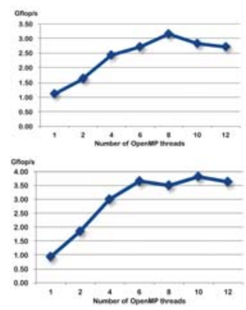
- Roofline model
  - helps estimate maximum attainable performance for a given algorithm
  - several variants (cache-aware or not)



Hybrid parallelism : first GPU tests

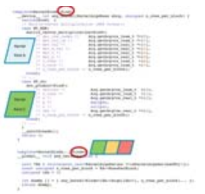
- Comparison of performance for a Jacobi (diagonal)-preconditioned CG (single MPI rank, Full OpenMP)
  - time for 100 iterations
  - *Code_Saturne* using CSR structure on Intel Xeon E5-2620 (Ivy Bridge) with hyperthreading, using 8 to 10 threads
  - ViennaCL or Magma on NVIDIA Tesla K40c GPU. The default block size is 256, which is also the size of the matrix slices in the SELLP format

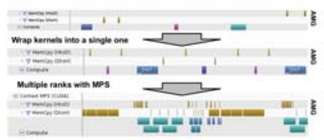




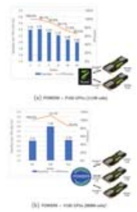
Hybrid parallelism : GPU progress

- Work in progress
- Work at IBM at Daresbury allows speedups near to 3 on ORNL's Summit (IBM POWER9 CPUs and NVIDIA Volta GPUs with NVLink)
  - Only the linear solvers are run on GPU
  - Initial work used OpenMP tasking, but compilers not mature enough
  - Current version used advanced CUDA tasking
    - https://www.slideshare.net/ganesannarayanasamy/cfd-on-power?qid=4a1b26dd-3e83-455f-8783-340f6b5559f6&v=&b=&from_search=13
  - Poster at SC18
    - https://sc18.supercomputing.org/proceedings/tech_poster/poster_files/post149s2-file3.pdf
  - Pull request on GitHub
    - https://github.com/code-saturne/code_saturne/pull/22

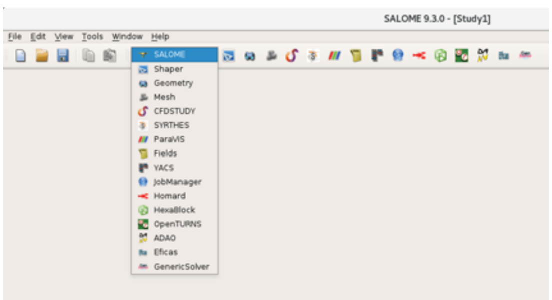




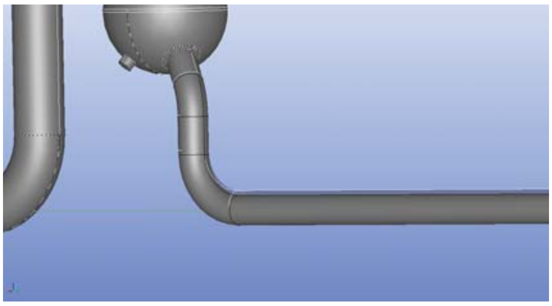
CPU+GPU speeddup over CPU-only and efficiency (strong scale)

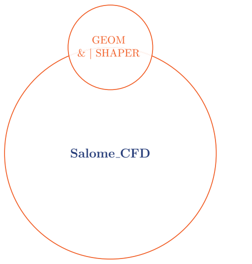


Salome_CFD in short

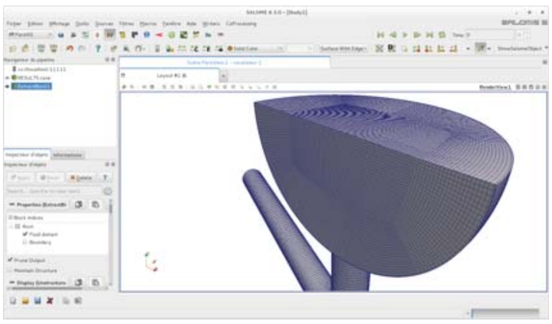


Salome_CFD in short

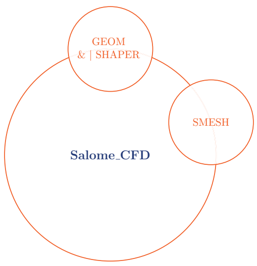




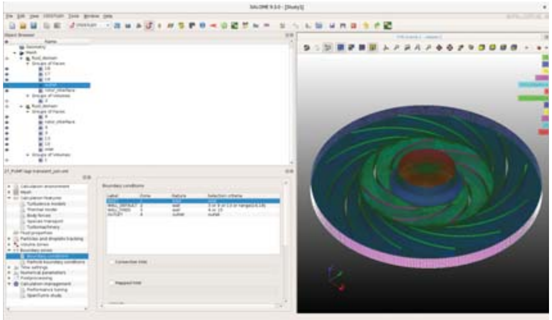
Salome_CFD in short

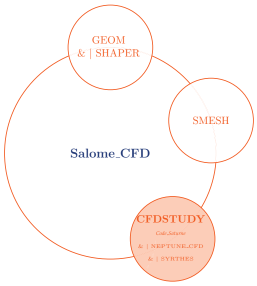


Advanced scripting capabilities



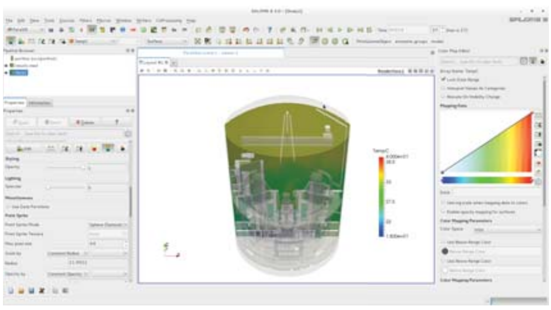
Salome_CFD in short

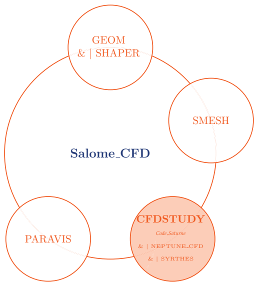




Single-phase solver *Code_Saturne*
Multi-phase solver NEPTUNE_CFD

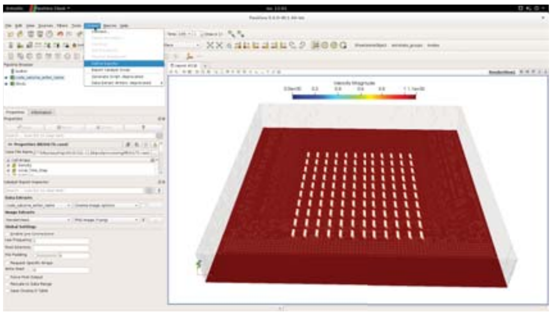
Salome_CFD in short

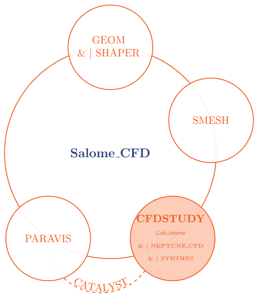




Visualisation / Remote visualisation for Big Data

Salome_CFD in short

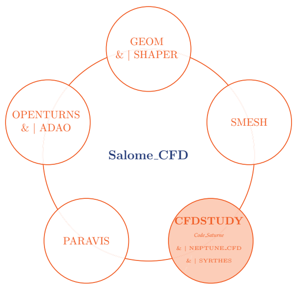




Visualisation / Remote visualisation for Big Data

In-situ and live visualization

Salome_CFD in short



UQ studies

Design

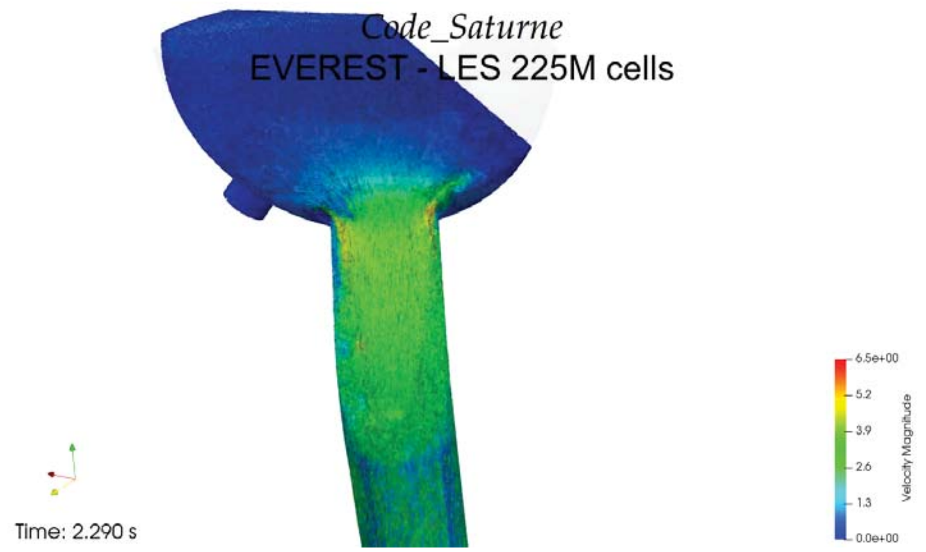
Other future directions

- → Pursue integration with other SALOME platform modules
  - Only integrate directly with components which are HPC compatible
  - Or in a manner compatible with HPC (client-server vaiants)
  - For future ensemble calculations, may benefit from OpenTURNS and Melissa integration for driving of uncertainty determination
- → Test parallel meshers / add readers if required
- → Using in memory data staging (avoiding files) with ADIOS, HDF5, or similar technologies may mitigate IO volumes
- → Optimize for future ensemble calculations
  - Pseudo code coupling (actually postprocessing coupling) may allow determine key statistics with less I/O and archival
  - This needs to be done in a fault-tolerant manner, so one run crashing does not cause the loss of the whole ensemble
  - Continue collaboration on the Melissa platform (https ://github.com/melissa-sa/melissa)

Conclusion messages

Code_Saturne can and must still evolve !
Suggestions and experimentation are welcome...
We can help (feedback/support loop) !

Thank you for your attention !



Richard Graham

Mellanox Technologies

Improving Application Performance: Mellanox's Collaboration with UK HPC

Dr. Richard Graham is Senior Director, HPC Technology at Mellanox Technologies, Inc. His primary focus is on HPC network software and hardware capabilities for current and future HPC technologies. Prior to moving to Mellanox, Rich spent thirteen years at Los Alamos National Laboratory and Oak Ridge National Laboratory, in computer science technical and administrative roles, with a technical focus on communication libraries and application analysis tools. He is cofounder of the Open MPI collaboration, and was chairman of the MPI 3.0 standardization efforts.

Abstract

Recently, Mellanox Technologies has become an active program of collaboration with end-users in the UK, focused around improving performance and scalability of the selected user applications as well as doing it's share of training up the next generation of HPC practitioners. The work has involved changes to the HPC-X, such as adding a more efficient MPI_Alltoallv algorithm to improve CASTEP performance, along with initial work to develop very efficient support for persistent MPI_Alltoallv, to further improve this code's performance. Mellanox is also partnering with the DiRAC program by sponsoring industry placements for graduate and recent graduate students. This work is application focused, using key applications to understand how these may benefit from using Mellanox's in network computing capabilities, such as SHARP and the BlueField SmartNIC. This work also aims to identify opportunities for improvements to HPC-X to better serve the applications.

Finally, following successful optimization of the Halo Exchange used by ICON, Mellanox is looking to improve the support for this general class of data exchange by relevant UK codes, and using it's InfiniBand support for hardware gather/scatter in the process. This presentation will describe Mellanox's approach to such collaborations, some of the on-going work, and current results from this work.

Karthee Sivalingam

Cray

Towards Understanding Exascale IO Needs – Insights from LASSi on ARCHER

Karthee is a Research Engineer at Cray EMEA Research Lab in Bristol UK, providing deep support for ARCHER (as part of the CoE) and also controbuting EU Research projects like SODALITE. After obtaining doctorate in Particle Physics from the University of Edinburgh, has worked previously at the University of Reading (Met Office), STFC Daresbury Lab (Hartree) and Infosys (IT). Research interest includes Analytics for Monitoring, HPC for Big data analytics and AI and HPC in the cloud.

Abstract

With a shared high-performance filesystem, application performance has become more dependent not only on peak IO capability but also on the degree of contention around shared resources. The HPC IO landscape has changed due to new Big Data and AI workloads. Understanding how this mix of application workloads from different domains with different IO requirements impacts the IO performance is very important. Today, ARCHER, the UK National Supercomputing service supports a diverse range of applications such as Climate Modelling, Biomolecular Simulation, Material Science and Computational Fluid Dynamics. LASSi is a framework developed as part of the ARCHER Centre of Excellence to analyse application IO usage and contention on the shared resource (Lustre file system).

LASSi combines the application job data from the scheduler with Lustre IO statistics to construct the IO profile of applications interacting with the filesystem. In this talk, we explain how a metric-based approach can be used to analyse application slowdowns in hours, which previously took days. This study highlights application groups that make unusual demands on the filesystem and an unexpected significant issue with applications launched within taskfarms or job arrays. We explore patterns in IO usage from application groups, for example CFD and python-launched applications. We will present an overall picture of IO usage of filesystem and application groups on ARCHER and show how this has changed over time. Such information can be further used for reengineering applications, resource allocation and filesystem sizing for future systems.

As we approach the Exascale, application IO requirements are evolving. Storage environments will change to support them. It will always prove valuable to understand how application interact with the memory/storage system to gain new insights on future performance challenges.

Towards understanding Exascale IO needs – insights from LASSi on ARCHER

Computing Insight UK 2019

Karthee Sivalingam * Harvey Richardson

CRAY EMEA Research Lab (CERL)

Adrian Tate

Numerical Algorithms Group (NAG)

Manchester

5-6 December, 2019

"A supercomputer takes a compute-bound problem and turns it into an I/O-bound problem"

Prof. Ken Batcher

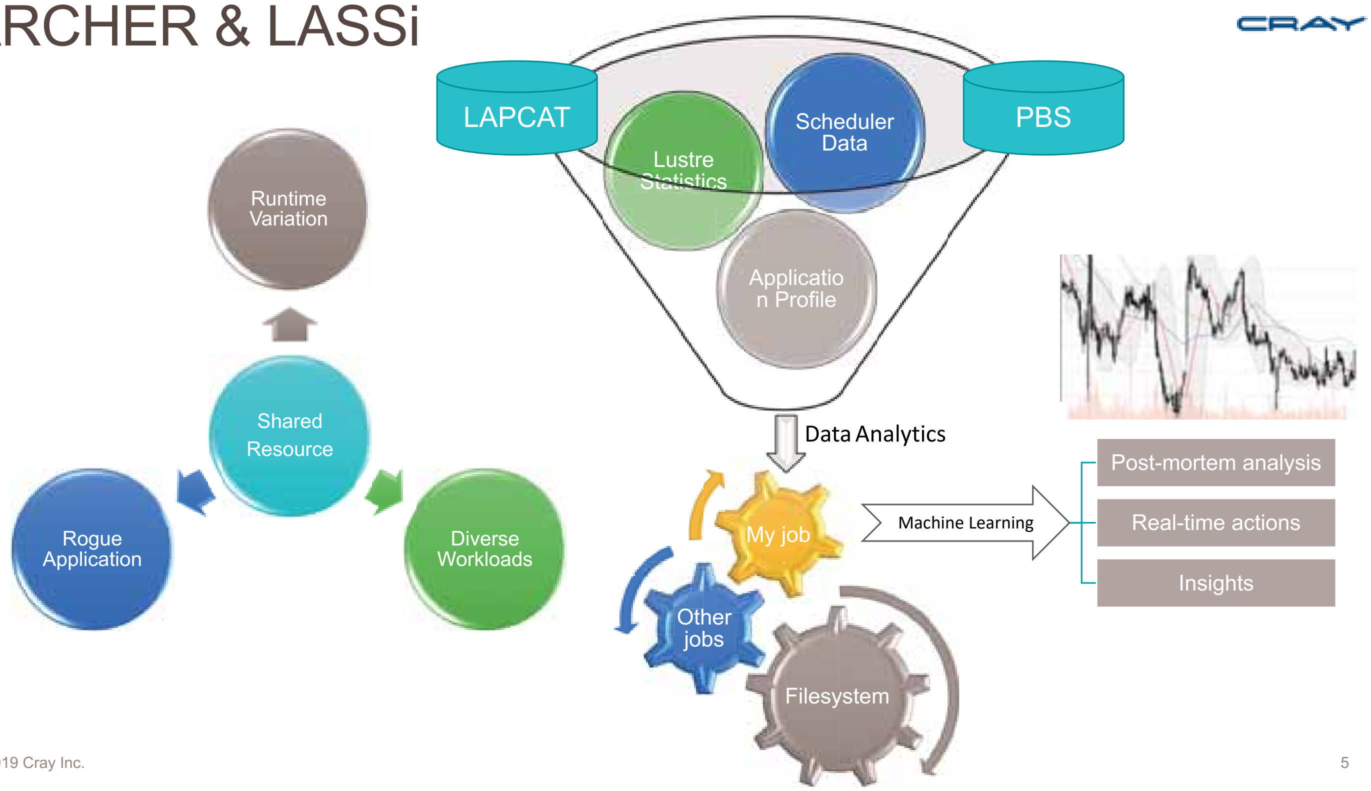
ARCHER

- Cray XC30
- 4920 nodes
- 12 core Intel Ivy bridge (64 GB)
- High-performance Lustre storage system
- Cray Aries interconnect

LASSi: The Big Picture

- Gain better understanding of performance issues in a complex workload for a shared HPC system
- Quite often some shared resource is a bottleneck
- Our focus at the moment is on I/O
- Based on statistics available from LAPCAT - collects the Lustre stats over Cerebro and stores them in a mysql DB on a management server
- Our approaches:
  - Extend work done at HLRS (looking at network contention)
  - Directly gain insights from statistics and aggregations.
- We have built a framework to triage problems and support analysis
- Eventually: Fast triage, early warning, 'health' status

ARCHER & LASSi



A different approach based on risks

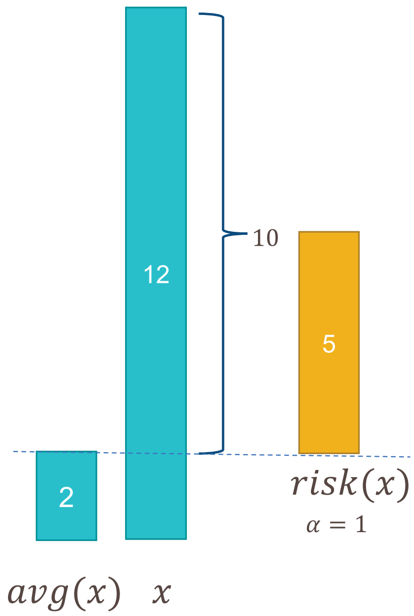
The simplest way to look at risks is perhaps:

- In isolation, slowdown will happen only when an application does more IO than expected (for example due to a configuration or code change)
- Also users will report slowdown only when they encounter more IO in a filesystem than expected
- We will use this idea as a metric for risks

Risk metrics

$$risk_{fs}(x) = \frac{x - \alpha * avg_{fs}(x)}{\alpha * avg_{fs}(x)}$$

- x is any IO operation OSS or MDS
- Risk is calculated for each application run
- We use averages for IO operation for each filesystem
- We calculate risk as
  scale of deviation from α times the avg on a filesystem
- Higher value of risk denotes a higher risk of slowdown



Metrics for IO

Quantity

$$risk_{oss} = risk_{read\_kb} + risk_{read\_ops} + risk_{write\_kb} + risk_{write\_ops} + risk_{other}$$

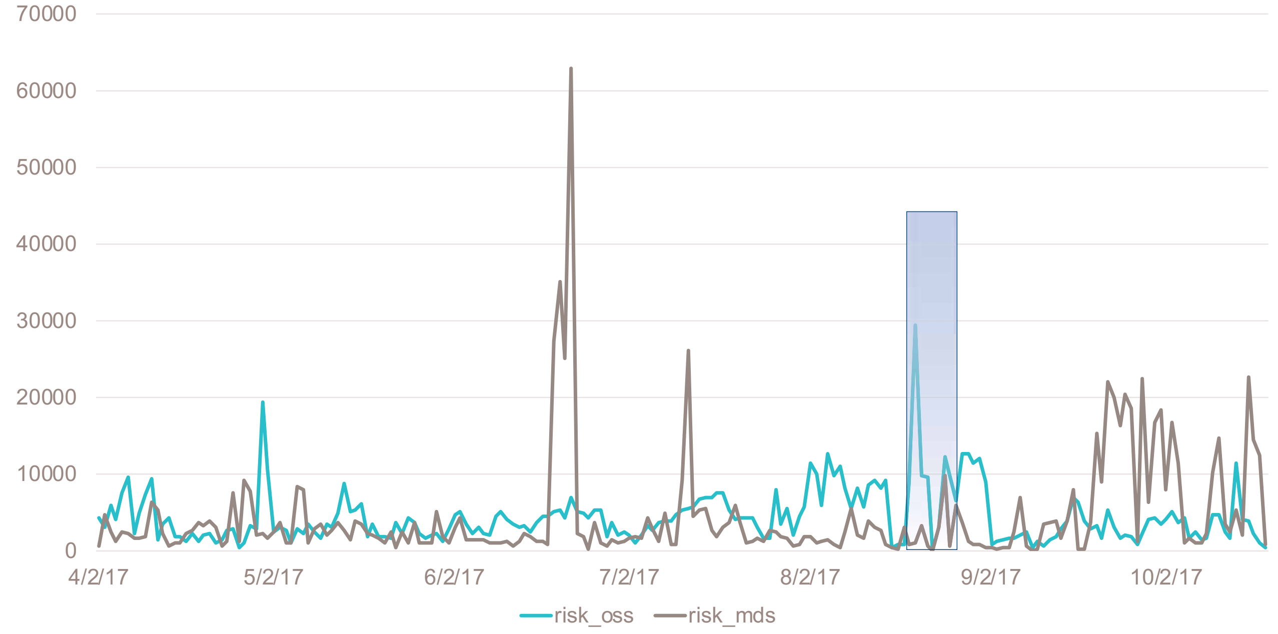
$$risk_{mds} = risk_{open} + risk_{close} + risk_{getattr} + risk_{setattr} + risk_{mkdir} + risk_{rmdir} + risk_{mknod} + risk_{link} + risk_{unlink} + risk_{ren} + risk_{getxattr} + risk_{setxattr} + risk_{statfs} + risk_{sync} + risk_{cdr} + risk_{sdr}$$

Quality

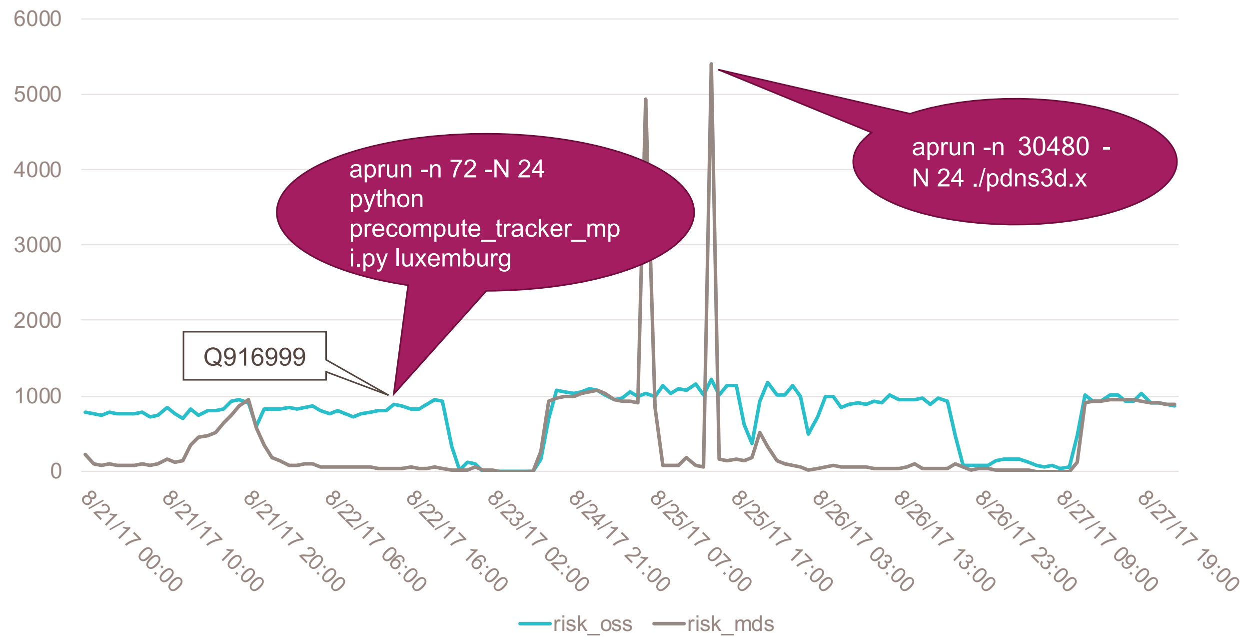
$$read\_kb\_ops = \frac{read\_ops * 1024}{read\_kb}$$

$$write\_kb\_ops = \frac{write\_ops * 1024}{write\_kb}$$

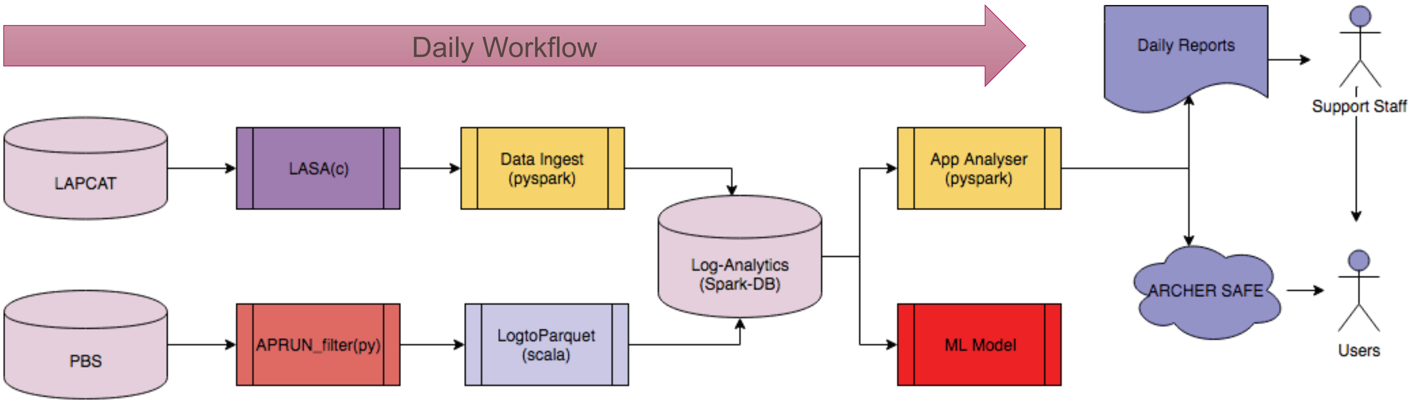
fs2 daily risk



fs2 hourly risk



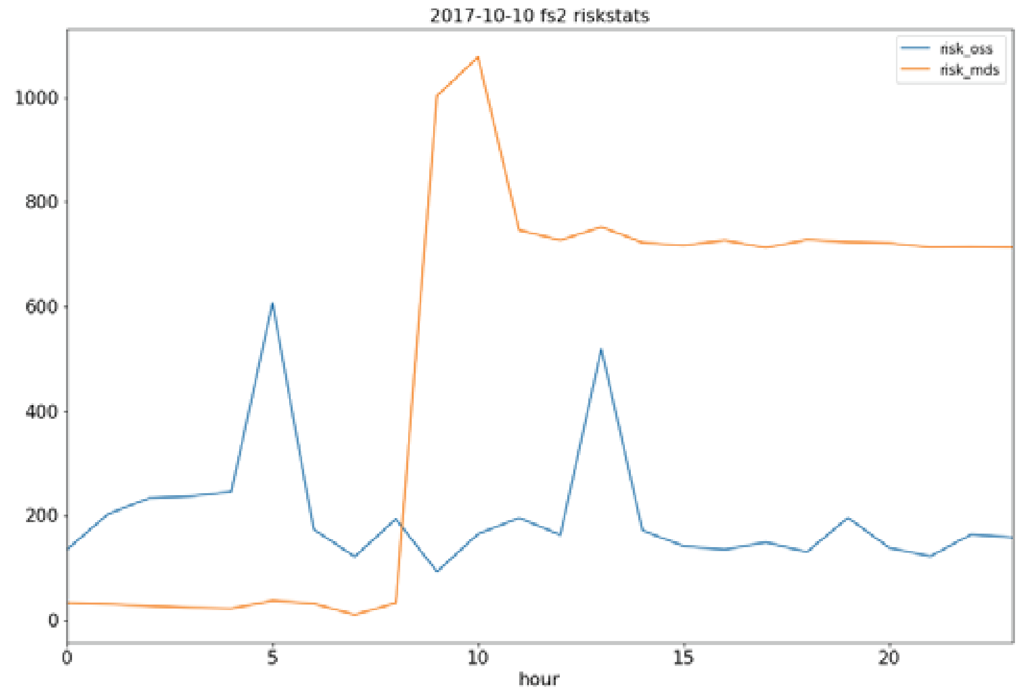
Architecture



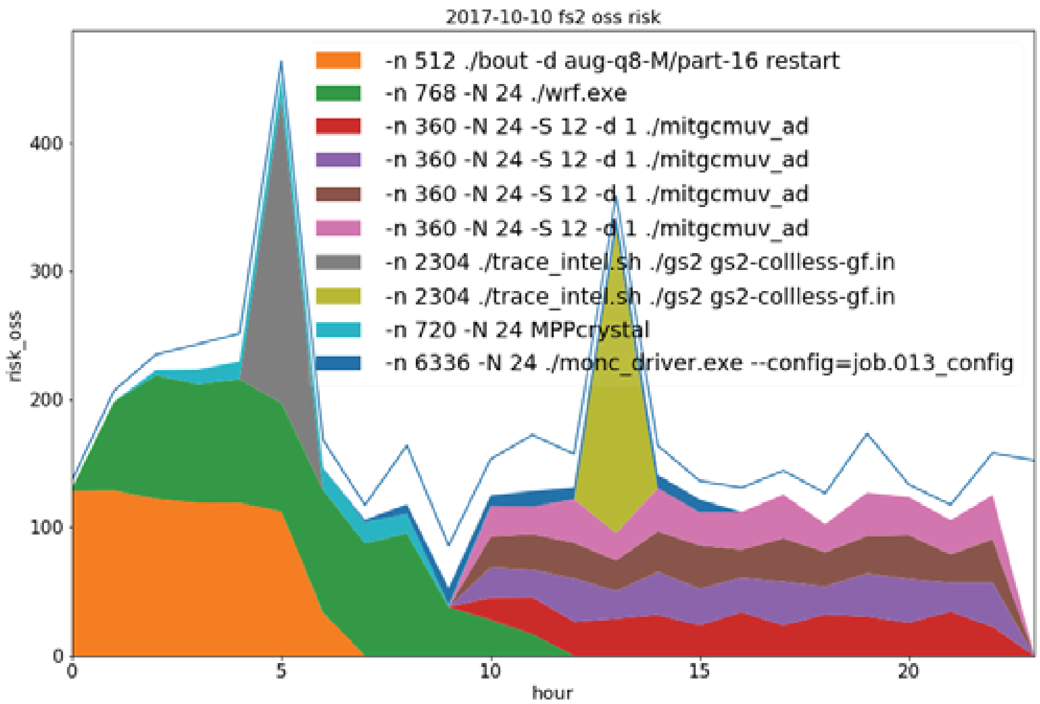
LASSi – a tool for real-time analysis.

- Provides an automated workflow
- Risk metrics can be fine tuned by assigning weights
- Risk model has been validated by comparison with actual reported slowdown incidents
- The existing automated process could be easily extended to enable real-time analysis (daily)
- Generates daily reports for ARCHER helpdesk and Cray support

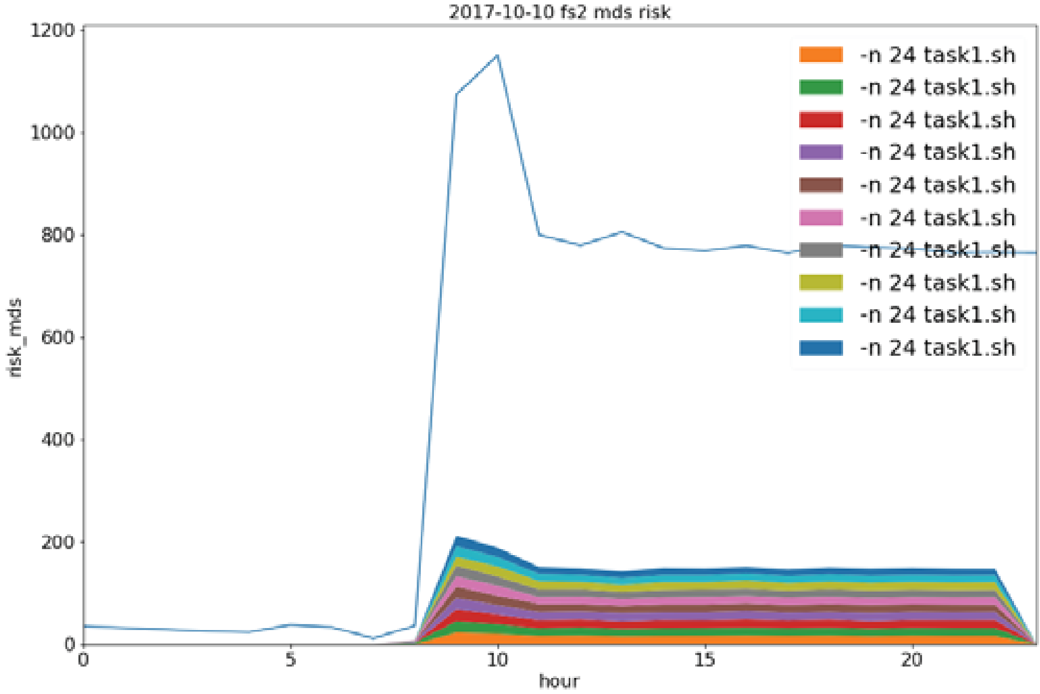
Examples of displays for helpdesk (risk)



Example of displays for helpdesk - daily risk to oss



Example of displays for helpdesk - daily risk to mds



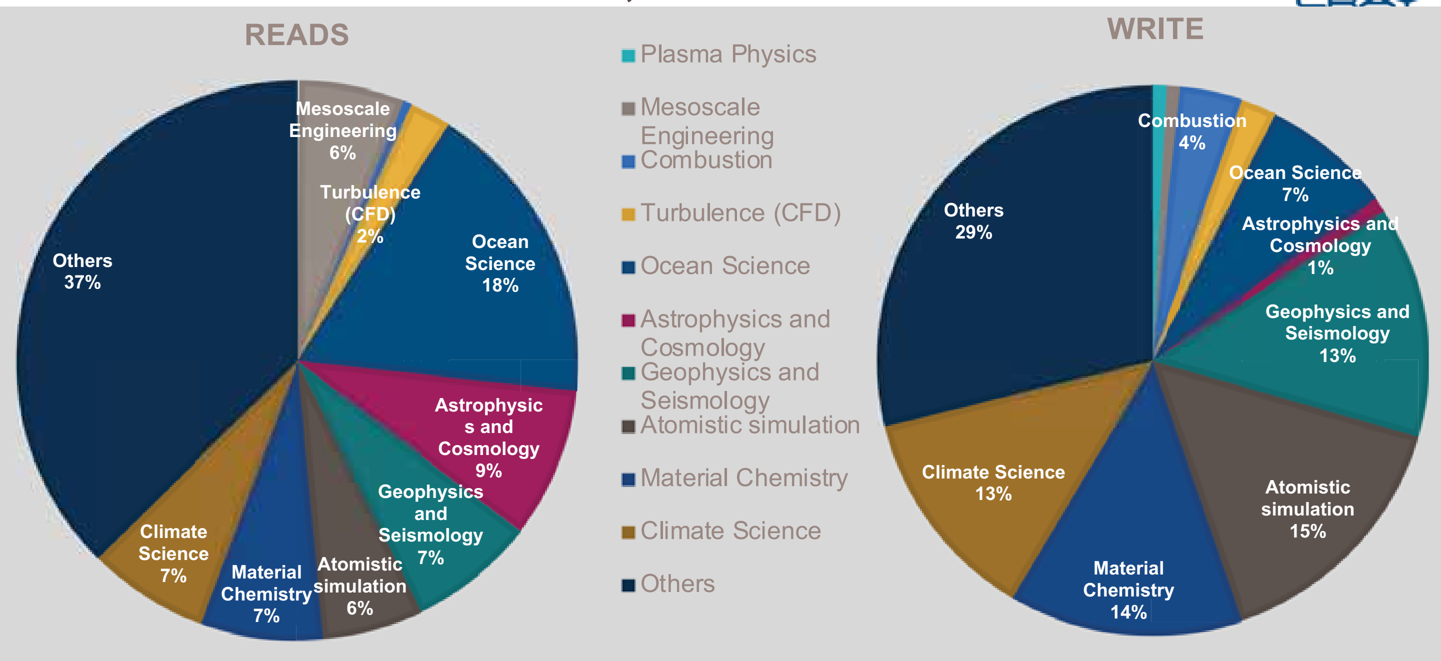
What can LASSi offer?

- A coarse IO profile of each application running
- Identification of abnormal filesystem IO usage
- Identification of abnormal application IO usage
- Identification of exact times when the filesystem is at risk of slowdown
- Identification of exact applications causing the risk of slowdown
- A prototype towards real-time analysis of risks and triggers

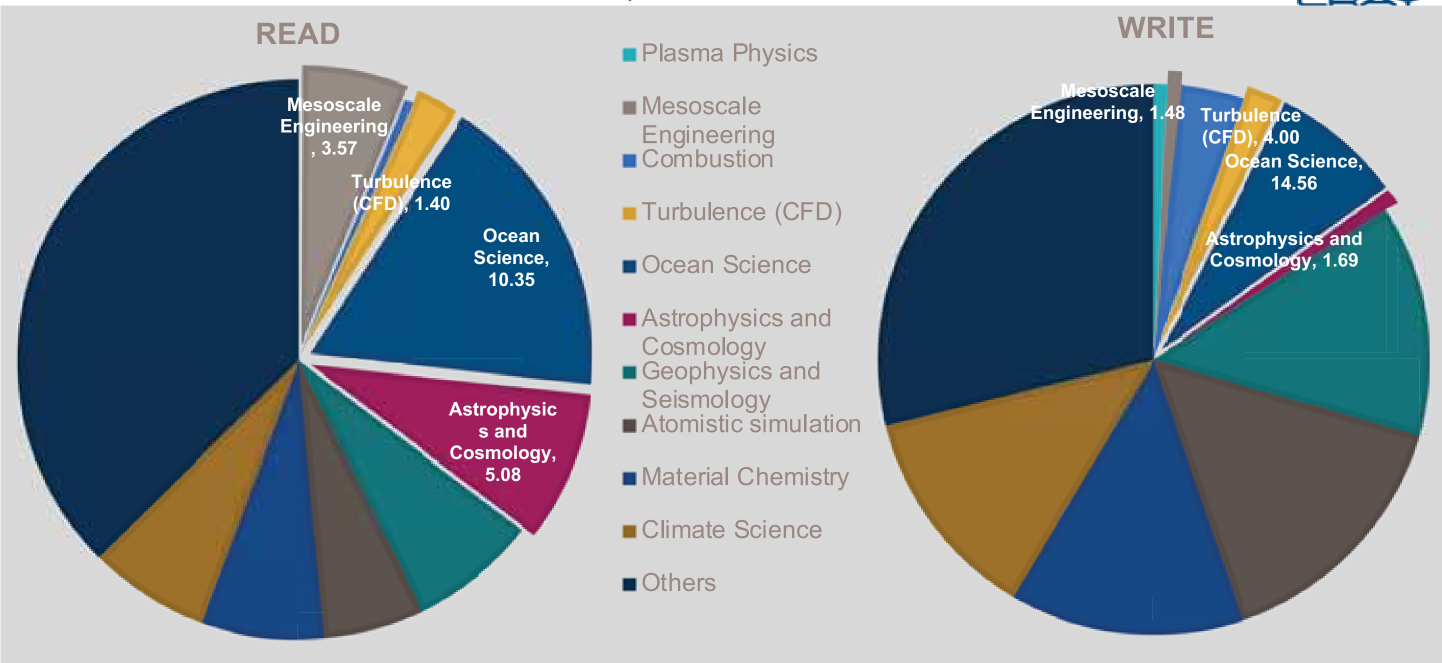
ARCHER Projects

- Mesoscale Engineering : lammps, Foam
- Turbulence : HYDRA, incompact3D, solver
- Combustion : boffin, senga, Foam
- Ocean Science : OPA, Nemo, mitgcmuv
- AstroPhysics and Cosmology : UKRmol
- GeoPhysics and Seismology : vasp, buildcell, axisem3d, wein2k
- Atomistic Simulation : castep, vasp, elk
- Material Chemistry : aims, vasp, nwchem
- Climate Science : UM_atmos, mitgcmuv, nemo, wrf

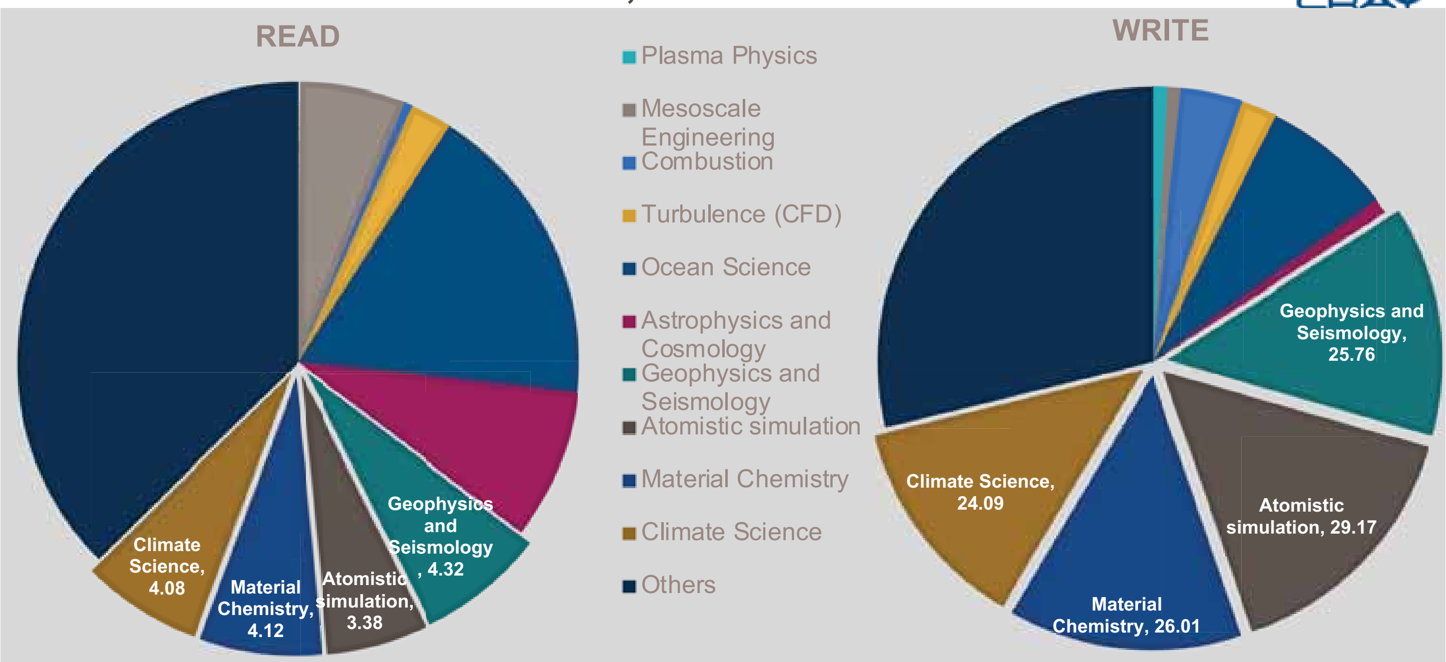
ARCHER: Read ~ 59 PB, Write ~ 192 PB



ARCHER: Read ~ 59 PB, Write ~ 192 PB



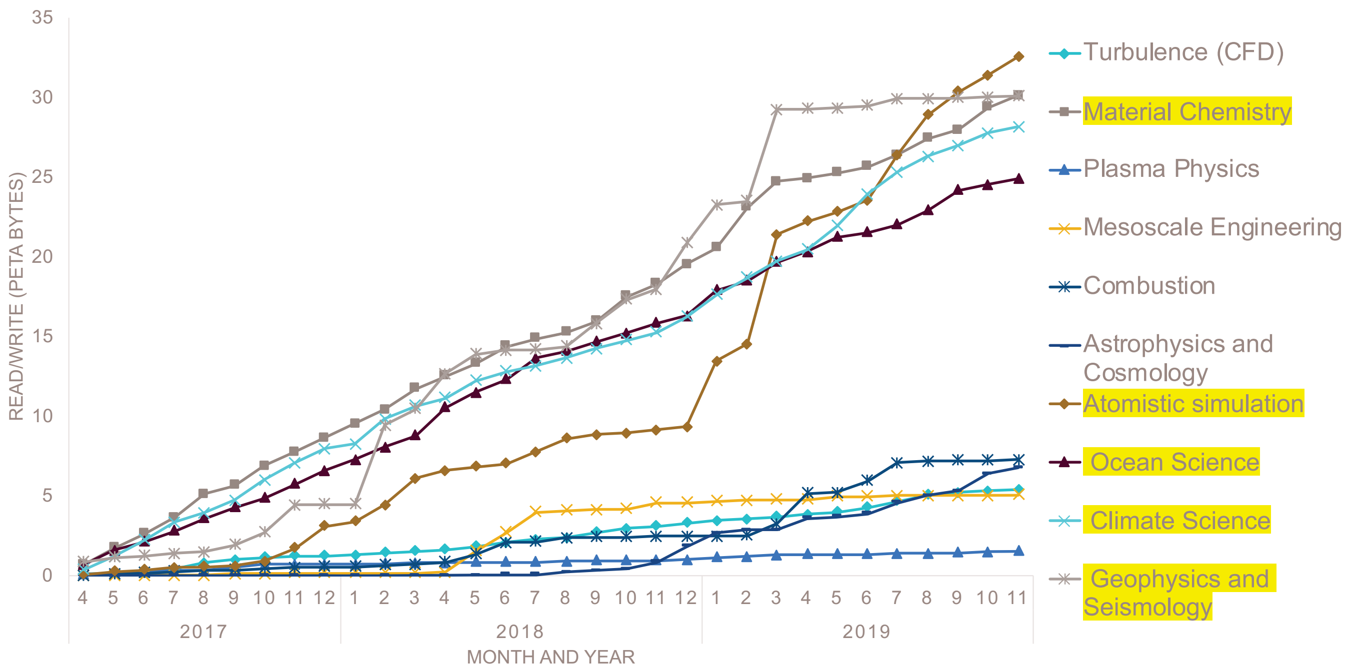
ARCHER: Read ~ 59 PB, Write ~ 192 PB



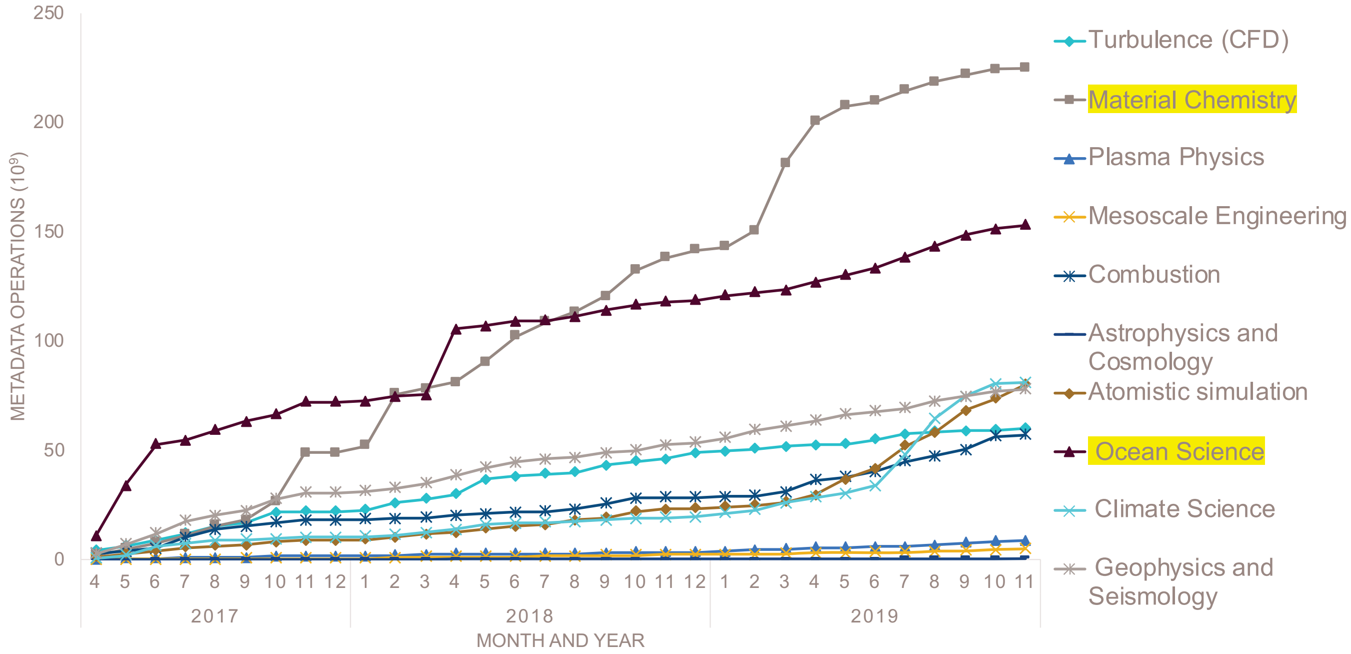
So what about Exascale?

- Technology
  - Initial Exascale systems will still use Lustre (gperformant with NVRAM)
  - We are likely to move to object store (key-value store as backend)
  - On top of this will use standard APIs like MPI-IO, NetCDF, HDF5
  - Will still want a POSIX layer (with its scaling limitations)
  - Projects like DAOS are interesting with increasing AI, Big Data applications
- Instrumentation and analysis still important
- Can we spot trends in applications/science as we move forward?
- Are we seeing changes today on ARCHER?

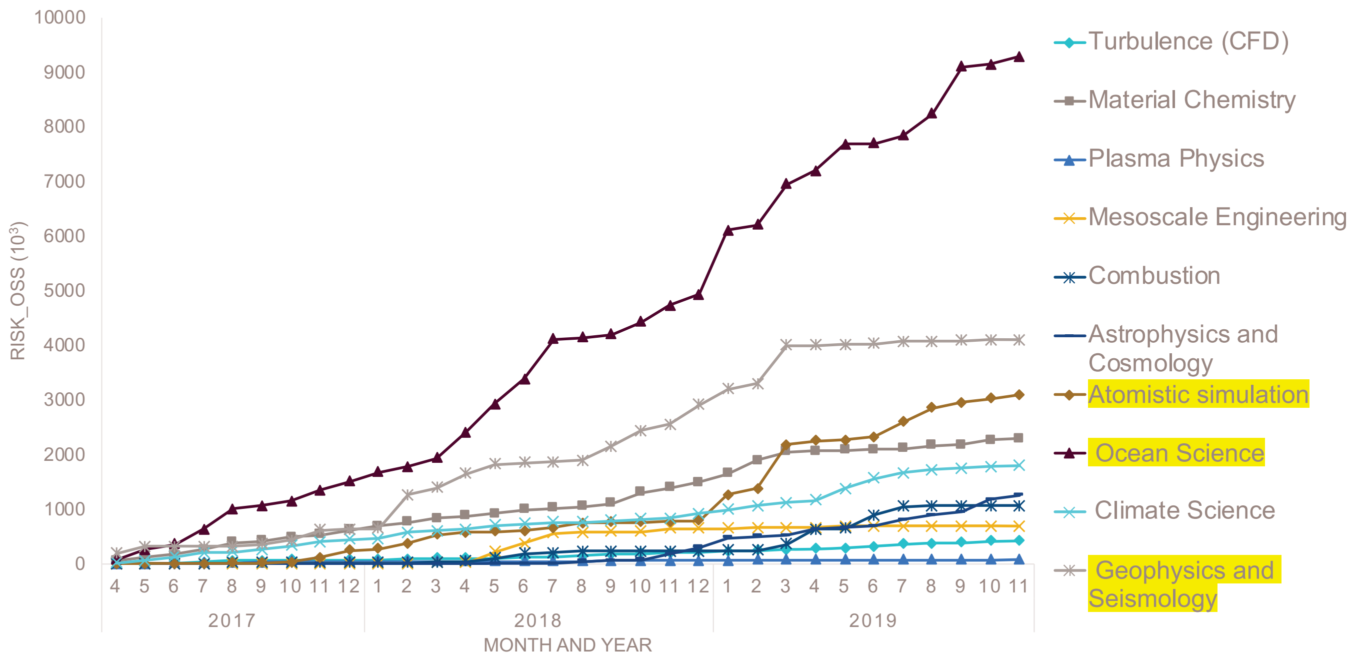
Read/Write in ARCHER



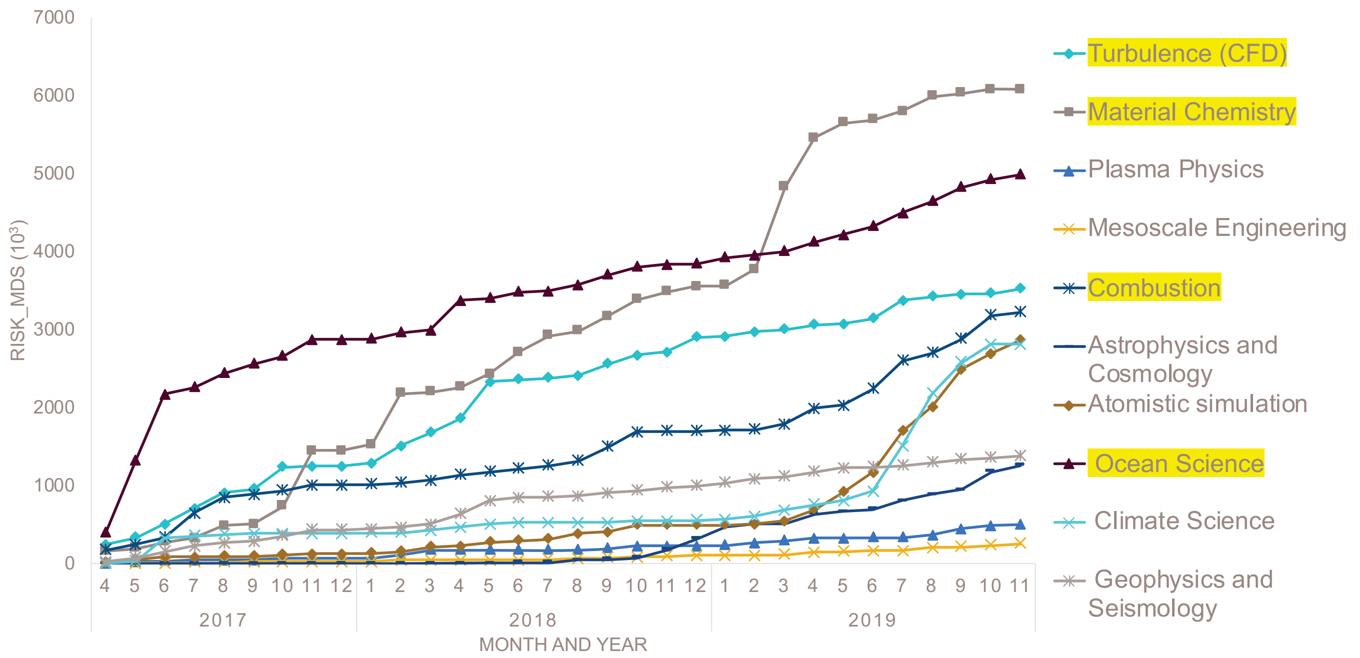
Metadata operations in ARCHER



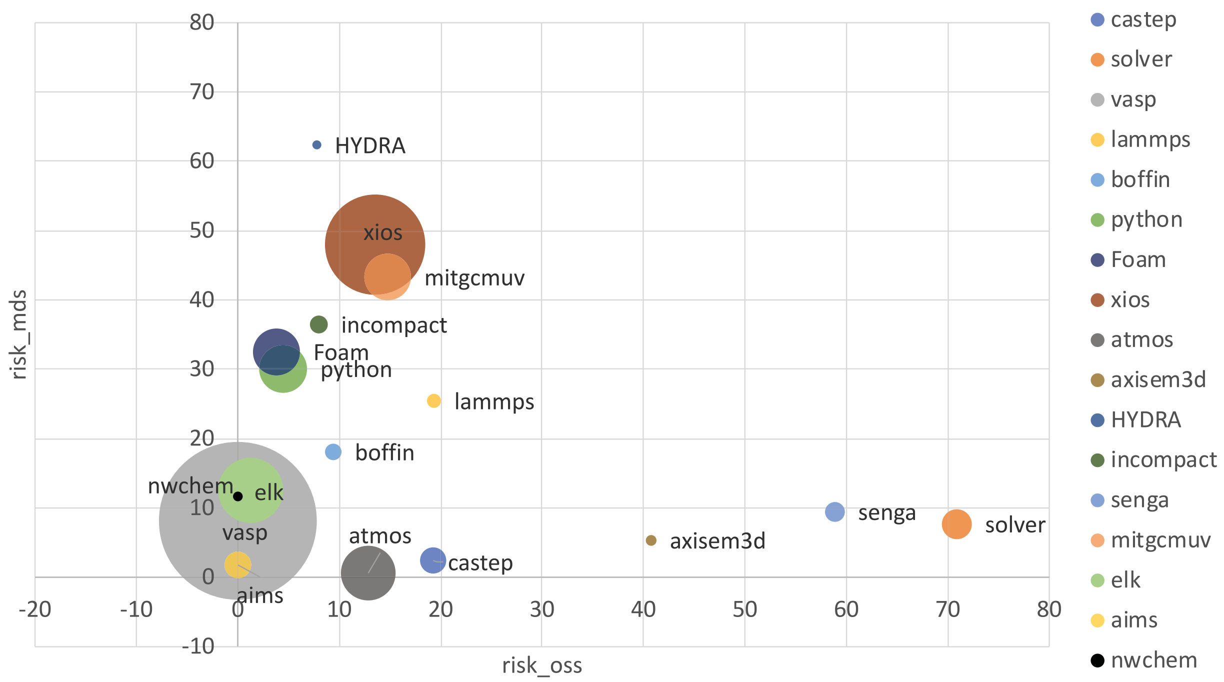
Risk to OSS in ARCHER



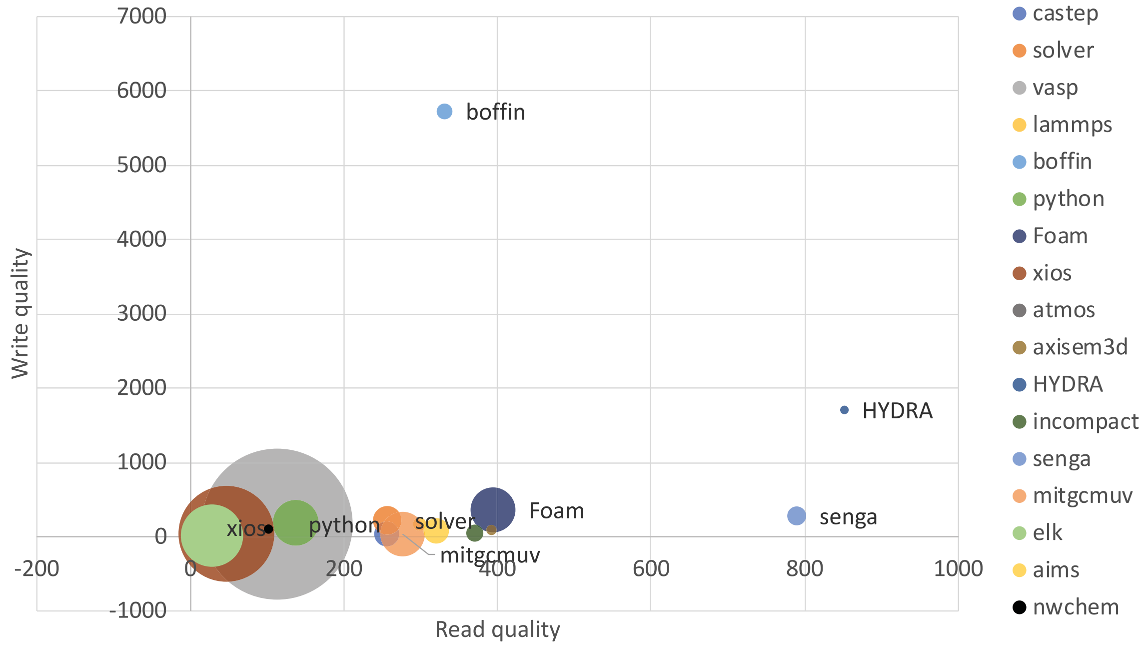
Risk to MDS in ARCHER



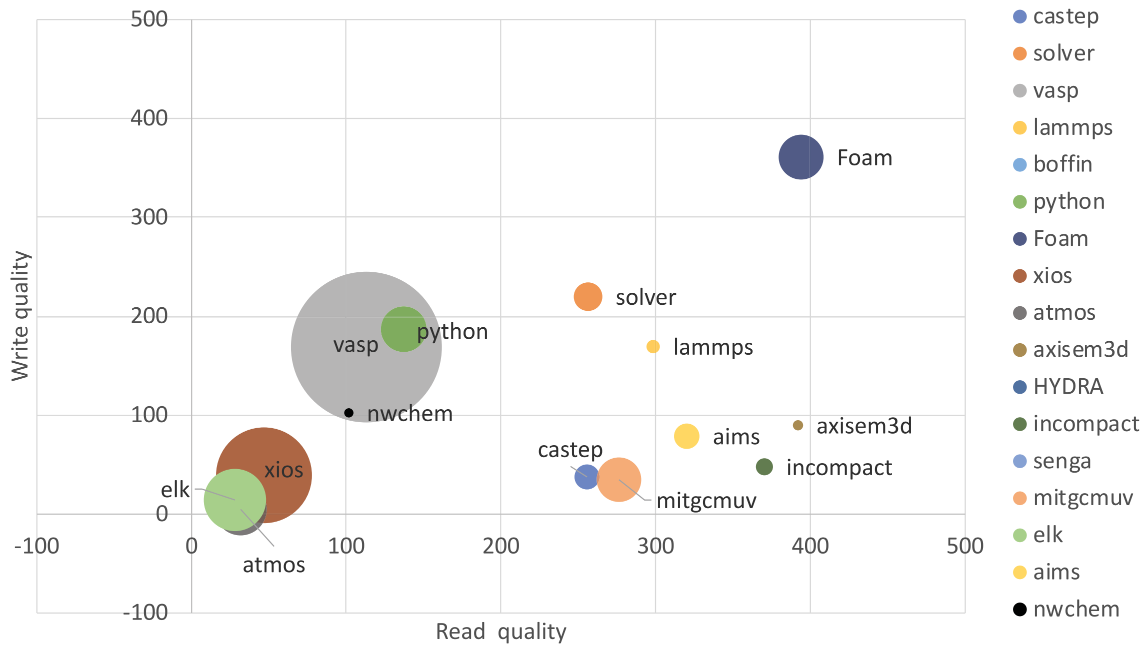
Risk to OSS vs MDS



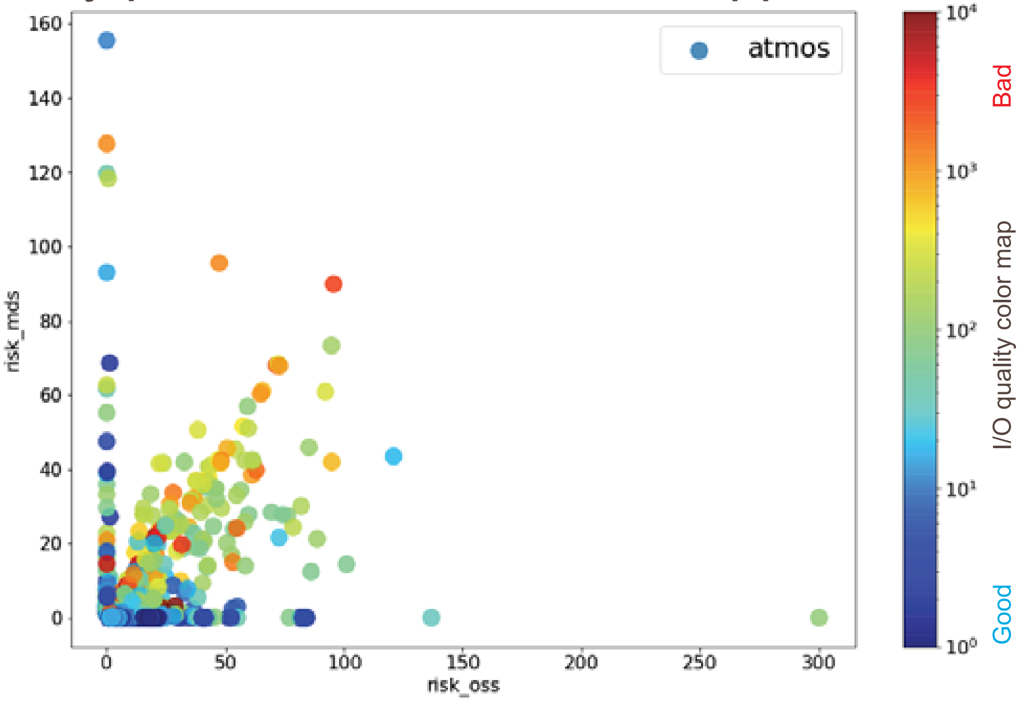
Read vs Write quality



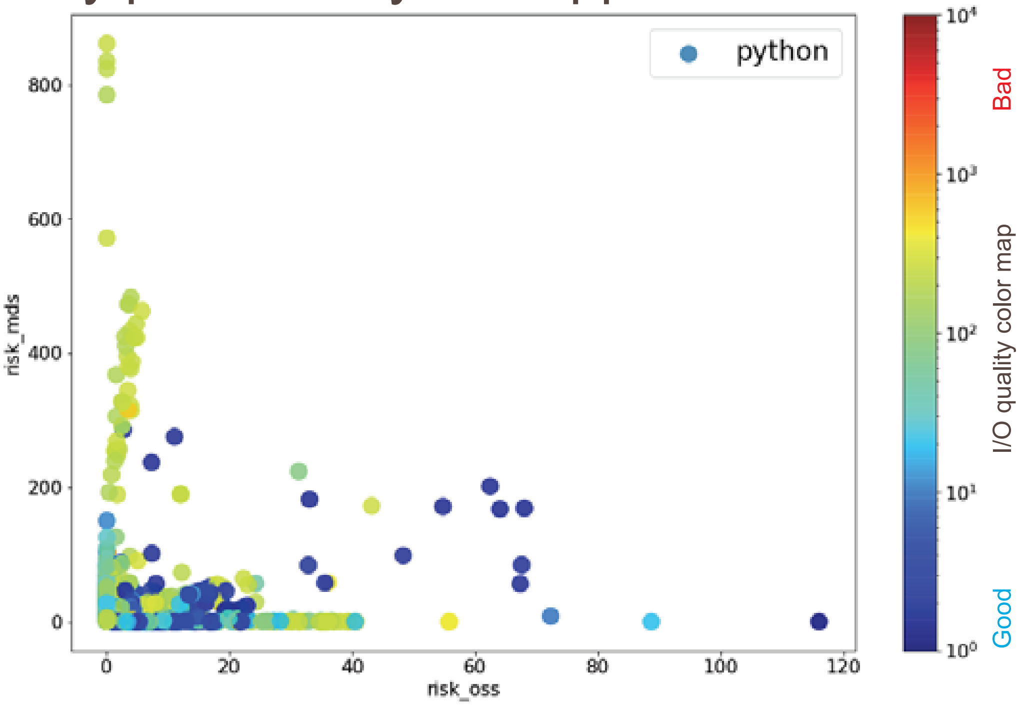
Read vs Write quality



Risk/quality profile of Climate/NWP applications

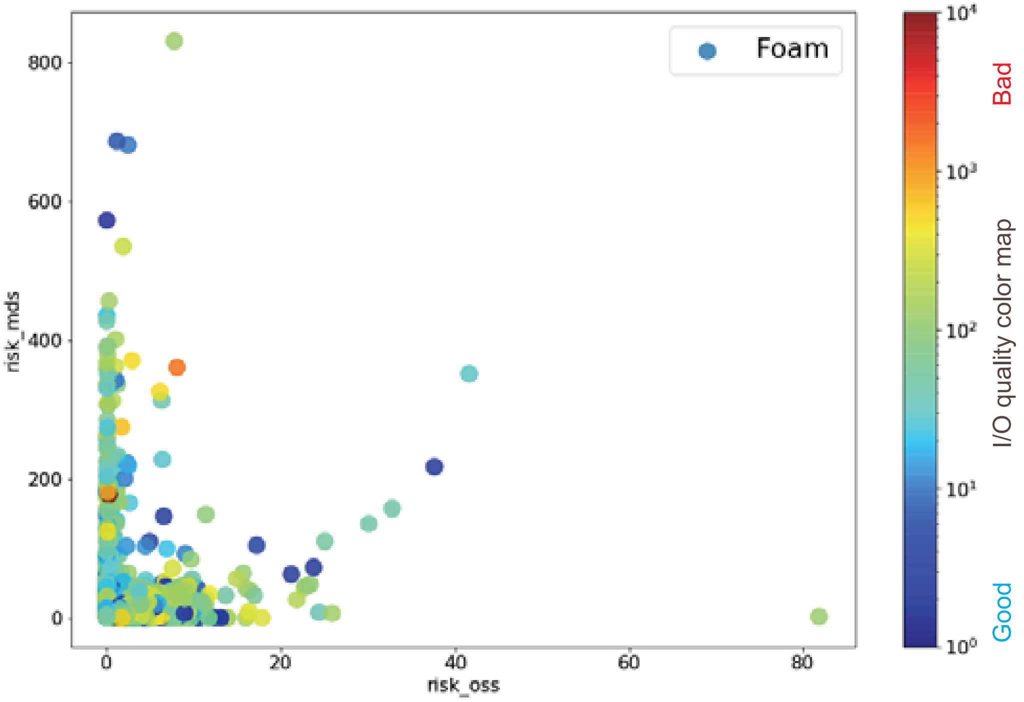


Risk/quality profile of Python applications



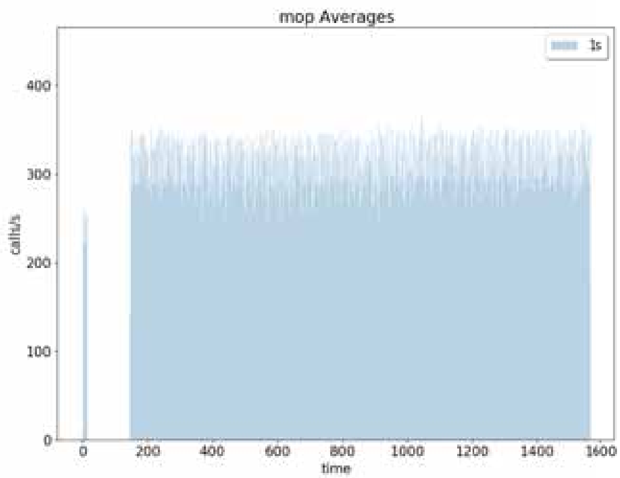
Risk/quality profile of CFD applications

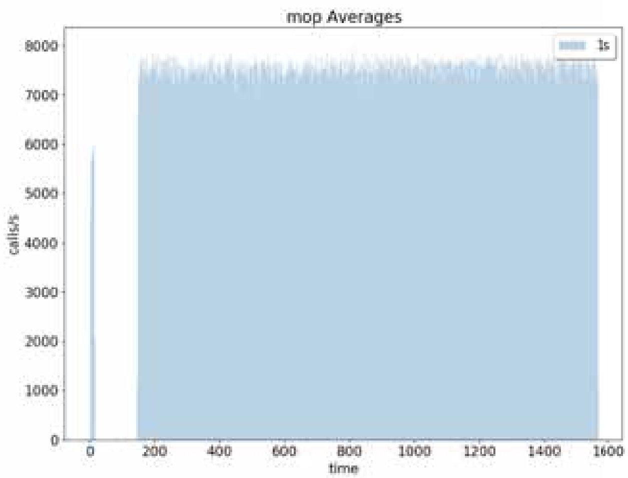
CRAY



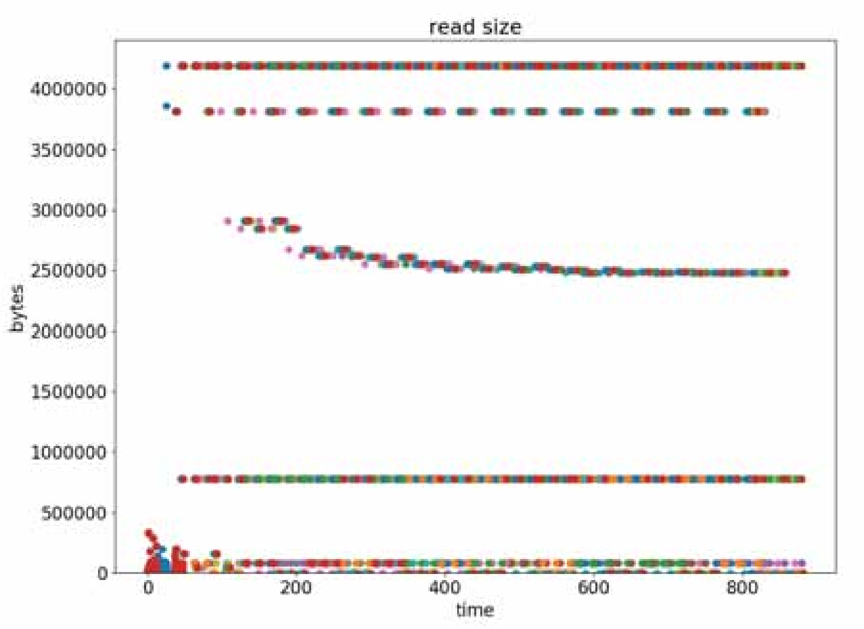
CP2K run in task farm (24 task)

CRAY

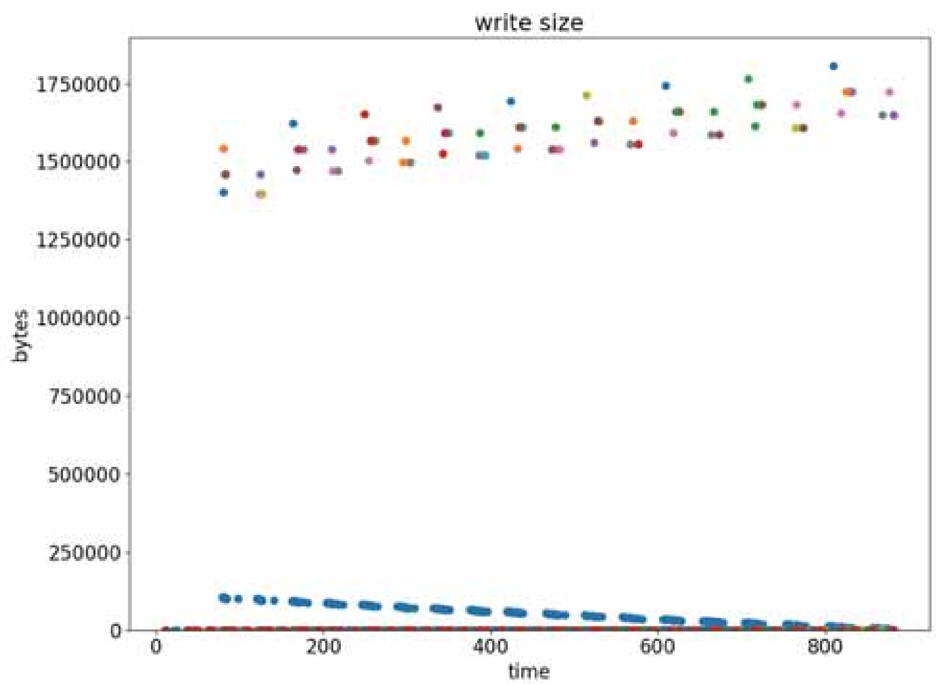




Python application RW size



Python application RW size



Summary

- LASSi provides an application-centric, non-invasive approach based on metrics to analyse slowdown due to IO
- Valuable in understanding application I/O behaviour on ARCHER
- Different communities/applications stress the filesystem in different ways
- For some communities these requirements are changing rapidly as the scale up
- Need to work with Project managers, Scientists and application developers to manage IO requirements and demands
- Continuous monitoring and analysis important in Exascale resource management.

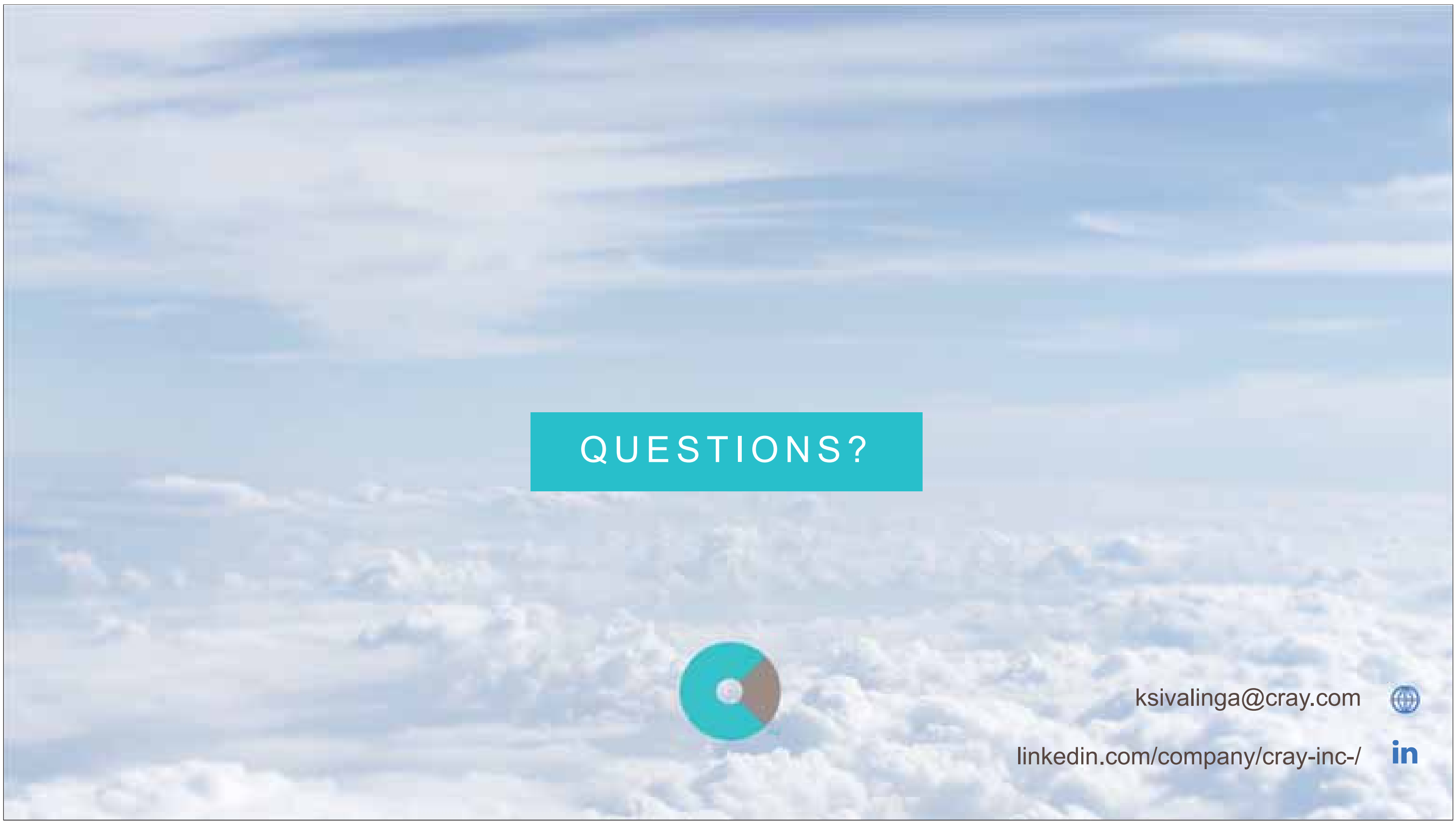
Acknowledgements

ARCHER helpdesk

CSE support (EPCC)

EPSRC

Cray EMEA Reseach Lab


QUESTIONS?
ksivalinga@cray.com
linkedin.com/company/cray-inc-/

Mark Thomson

STFC

UKRI E-Infrastructure Roadmap: Next Steps

Professor Mark Thomson is Executive Chair of the Science and Technology Facilities Council (STFC). STFC, which is one of the nine councils of UK Research and Innovation, responsible for particle physics, astrophysics, space science and nuclear physics. He is also responsible for the large-scale multidisciplinary research facilities at the UK National Laboratories. Within UKRI, Professor Thomson leads on infrastructure, including e-Infrastructure, and is currently directing the work to produce the UK's first Research and Innovation Infrastructure Roadmap, which will be released in 2019.

Professor Thomson has held national and international research leadership roles at the forefront of particle physics in both neutrino physics and collider physics. Most recently, he has been the co-leader of the Deep Underground Neutrino Experiment (DUNE), a collaboration of over 1000 scientists and engineers. Beyond his own research, Professor Thomson has held numerous research oversight roles in the UK and abroad. In 2013, he published "Modern Particle Physics", a textbook that has been widely adopted for undergraduate courses at universities around the globe.

Kate Marshall

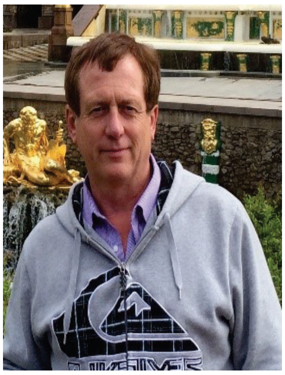
IBM

Quantum Computing for the 21st Century

Kate has recently finished her studies in Masters level Physics at University College London, specialising in Astoparticle and Neutrino Physics. She is very proud to have joined IBM in the last year, as a Technical Consultant and IBM Q Ambassador. She has had a long-standing interest in Quantum Computing and Communication, as well as where this field meets the main focus of her degree in Particle Physics, such as the use of Majorana quasiparticles as a potential basis for Quantum Computers.

Abstract

This talk will cover IBM's take in the race to build commercially useful Quantum Computers. In particular, we will look at the technology IBM is using and where we see applications arising in Quantitative Finance, Chemical Innovation, Production and Transport industries and more. We will also explore how to measure progress in this fast moving and competitive industry, as well as how anyone can get involved in using and writing software for IBM Q Quantum Computing Systems using our Qiskit SDK.

Michael Bane and Shane Rigby

University of Liverpool and Atos

Quantum Computing Ambitions from University of Liverpool

Michael lectures in high performance computing and emerging technologies at the University of Liverpool, and manages the Centre for AI Solutions. Michael's research centres on energy efficient computing.

Shane is an experienced business professional, with a visionary approach to new business development, sales, and marketing. He has more than 25 years of experience, with a significant portion of this time focused on the IT&C industry.

As the Business Development Executive for Deep Learning and Quantum Learning, Shane is responsible for helping customers take a data driven journey, using latest AI technologies and Supercomputing to discover, and unlock, unique business potential in their data. As AI/ML/DL and soon Quantum becomes mainstream, the customers engaging with the will be using proven infrastructure, analytical modelling, self-learning and creative pricing, to gain a compelling advantage in most key vertical markets. Using the latest HPC, GPU and Quantum learning technology, combined with leading edge services, Shane is helping to change the Artificial Intelligence landscape, including fully automated discovery and interpretation.

Having lived in Hong Kong for four years, covering the whole of Asia, Shane has gained significant global experience. Besides residing in Asia, Shane has lived and worked in the USA, Europe, and Russia, and held Director and VP positions for many of his clients. Early in his career, Shane worked for Redifusion Computers and pioneered a patented Computer emulation hardware/software platform.

Shane graduated with an MPhil - Master of Philosophy, Advanced Master's and Bachelor's degrees in Electronics and Electrical Engineering from Brighton University in the UK. He is a Corporate Member of the IEE and IET and holds a Chartered Engineering lifetime status.

Abstract

The University of Liverpool aims to foster quantum education and is on a journey to become a leading UK quantum applications centre. With a programme of activity taking place including several workshops, Michael will outline his strategy and aspirations for the University, whilst Shane will guide attendees through the role of the Atos Quantum Learning Machine, how it is benefiting the plans at the University, and how other organisations can prepare for quantum computing.

UNIVERSITY OF LIVERPOOL

Quantum Computing Ambitions at Liverpool

Dr. Michael K. Bane

Quantum Computing Ambitions at Liverpool
Dr. Michael K. Bane

UNIVERSITY OF LIVERPOOL

1881
Founded in 1881 as the original 'red brick'

81%
of our research 'world-leading' or 'excellent'

Top 200
Ranked in the top 200 universities worldwide

Campuses in China and London

33,000
online and on campus students

£652 million
gross value added to Liverpool City Region in 2015/16

100% online

9
Associated with 9 Nobel Laureates

UNIVERSITY OF LIVERPOOL

Dept. of Computer Science

@LivUni_CompSci

- 50+ academics, ~850 UG+PG students
  - #1 Russell Group for social mobility in student recruitment
  - Awards for Tech-Enhanced Learning
- 97% REF outputs as world-leading or internationally excellent
- Area of expertise
  - Algorithms & Theoretical Comp Sci
  - AI, Robotics, ML
- Expanding range of industrial collaborators

Quantum Computing (QC)

- Motivation
- Current interest
- Ambitions

QC: Motivation

- Potential recognised
  - ability to do some things very fast
  - searches, cryptography, probabilities (with feedback ==> AI)
- Perceived barriers
  - inability to know how to model many things
  - quantum noise

UNIVERSITY OF LIVERPOOL

QC: Motivation

- Potential yet barriers == RESEARCH OPPORTUNITIES

- @LivUni QC Network
  - inter-departmental
  - exploring potentials
  - overcoming barriers
  - funding for further research

UNIVERSITY OF LIVERPOOL

Computer Science

- Dr. Alexei Lisitsa
  - Automated verification of quantum algorithms & programs
  - QC for verification of classical (& quantum) algo & programs

  - Cryptography
  - [T] RNG

  - Applications of QC in automated reasoning, computer-assisted mathematics, …

  - 1 student: factorisation & network optimisation via Quantum Annealing

UNIVERSITY OF LIVERPOOL

Mathematical Sciences

- Dr. David Schaich
  - Quantum simulation of lattice quantum field theories
  - NISQ technology – use & development
  - IBM-Q
  - 1 post-doc to support above
    (speak to me if you think that is you!)

UNIVERSITY OF LIVERPOOL

Mathematical Sciences

- Dr. David Lewis
  - Riemann Zeta Function
  - Accelerate Riemann Hypothesis verification processes
  - via combo of classical + quantum (Grover's?) algorithms
  - from N/2 to sqrt(N) order complexity

Dudek (2014) proved that the Riemann hypothesis implies there is a prime p satisfying

$$x - \frac{4}{\pi}\sqrt{x}\log x < p \le x$$

for all $x \geq 2$. This is an explicit version of a theorem of Cramér.

https://en.wikipedia.org/wiki/Riemann_hypothesis

Electrical & Electronic Engineering

- Prof. Simon Maskell
  - Quantum MC
  - Sequential MC, embarrassingly parallel but with global update
  - +1 PhD
    (speak to me…)

Chemistry

- Dr. Max Birkett
  - Functional Materials Discovery
  - Use of QC massive parallelism to explore search spaces
  - Exploration of novel materials / construction of specific quantum circuits for specific algorithms

UNIVERSITY OF LIVERPOOL

Integrative Biology (partner: STFC)

- Dr. Daniel Ridgen
  Dr. Ronan Keegan
  - QC for structural biology
  - Processing of images
    ==> optimisation of vast combinatorial landscape

UNIVERSITY OF LIVERPOOL

Computer Science

- Dr. Michael Bane
  - Evaluation of emerging tech
    https://emit.tech
  - Reducing environmental impact of
    high end computing
  - 2 students:
    > cryptography / bitcoin hashing
    > solving large systems of linear equations

And...

- Luke Anatassiou (EEE): quantum simulation of quantum gyro

- Anthony Mtitimila: Research Partnerships & Innovation
- Dr. Manhui Wang (CSD): QC training provision

Liverpool QC Network

- QLM training
  - 20 Liverpool
  - +3 apologies
  - + interest from KT partners

- Network
  - Email / meetings
  - Peer-to-peer support
  - Grant submissions

Liverpool QC Network

- Access…
  - Atos / QLM
  - Atos / myQLM
  - IBM-R / IBM-Q
  - D-wave
  - Bristol Univ
- Working with partners
  > hubs
  > grant bids
  > development

Ambitions

- Evaluation of potential
  => simulation via QLM (on-site + researcher access)
- Active input to barrier removal (QC development)
  - What vendors could do; how end users could use; new algos; …
- Teaching
  - UG, PGT & CPD/industry
- Increasing liaison with key partners
  - Atos, IBM, D-Wave (etc)
  - working with various Hubs & ISCF programme

UNIVERSITY OF LIVERPOOL

https://cgi.csc.liv.ac.uk/~mkbane/

Atos Quantum Learning Machine

A future proof approach to quantum computing application development

Shane Rigby,
DL & Quantum BD Executive, Atos UK&I

Atos

The Quantum Computing Party Hasn't Even Started Yet

But your company may already be too late

- **Here's what you can do now to be ready:**
  - **Create a formal effort to explore quantum computing applications.**
    - It needs people and resources, but not many to start. Treat it as a research and development expense with a high probability of paying off in three to five years.
  - **Identify where quantum computers can help your company most.**
    - Likely to involve optimizing complex systems that are difficult or impossible to model today.
  - **Build relationships with quantum computer makers.**
    - As companies refine their hardware, they are eager to help potential customers develop software.
  - **Cultivate emerging talent.**
    - The biggest problem that many companies will face as quantum computers become available is the shortage of software engineers who know how to use them
  - **Build prototype quantum applications in your field.**
    - What's important is that you create new algorithms that use the distinctive mathematics of quantum computers

Atos

Neven's law vs Moore's law

- **Dec 2018 Quantum = PC**
- **Jan 2019 Quantum = best WS**
- **Feb 2019 Quantum = Google super computer**
- **Mar 2019 Quantum = 1M Google CPUs**
- **Jul 2019 Quantum > Largest HPC/GPU capability**
- **End 2019 Quantum supremacy ?**
- **Quantum computers are gaining computational power relative to classical ones at a "doubly exponential" rate**

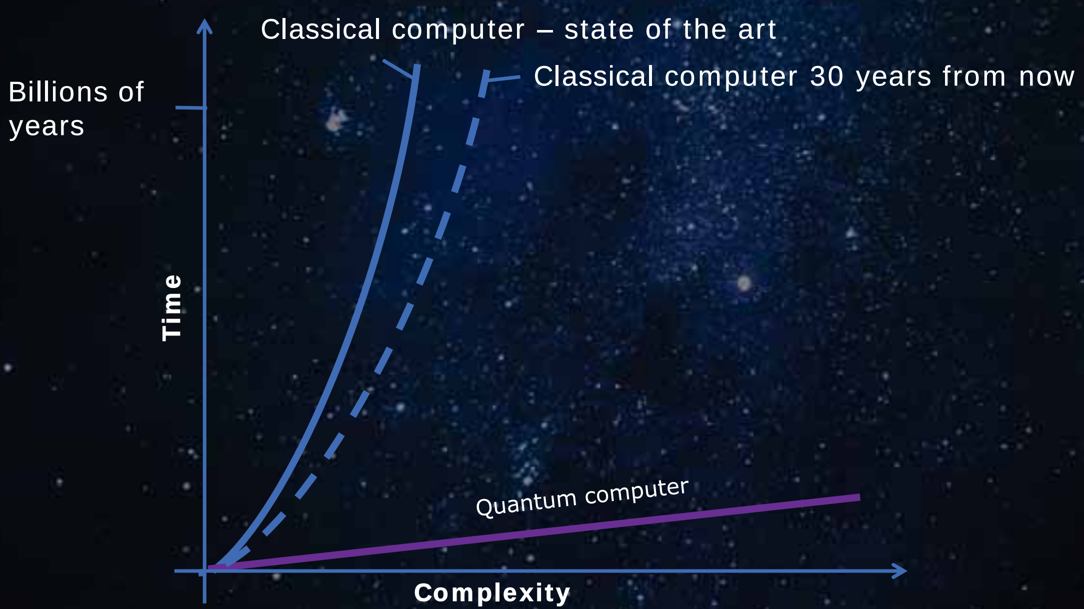
- **Moore's Law** $2: 2^1, 2^2, 2^3, 2^4$
- **Neven's Law** $2: 2^{2^1}, 2^{2^2}, 2^{2^3}, 2^{2^4}$

Source Quantamagazine Jun2019

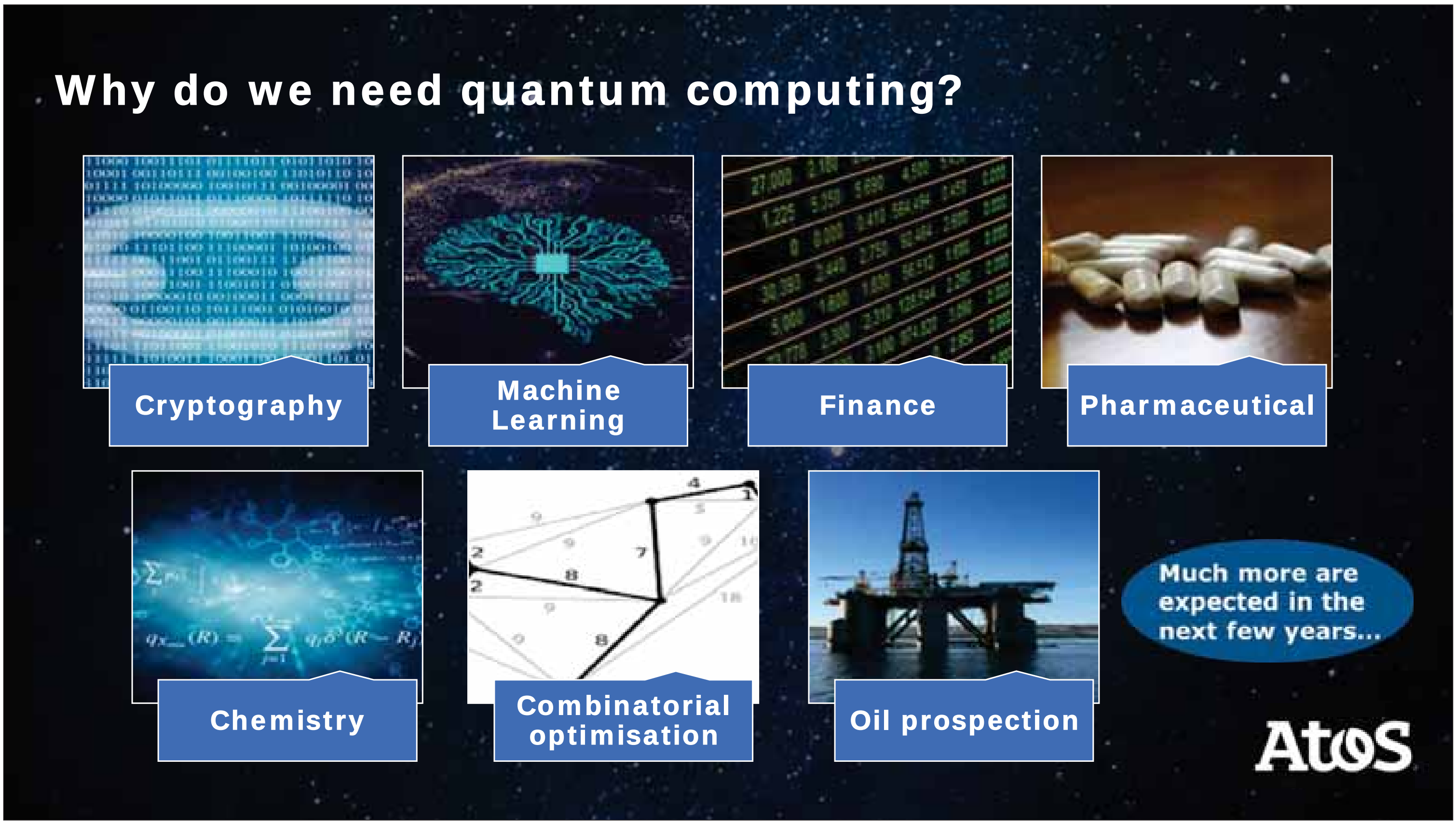
Google's "Bristlecone" 72 QuBit quantum processor.

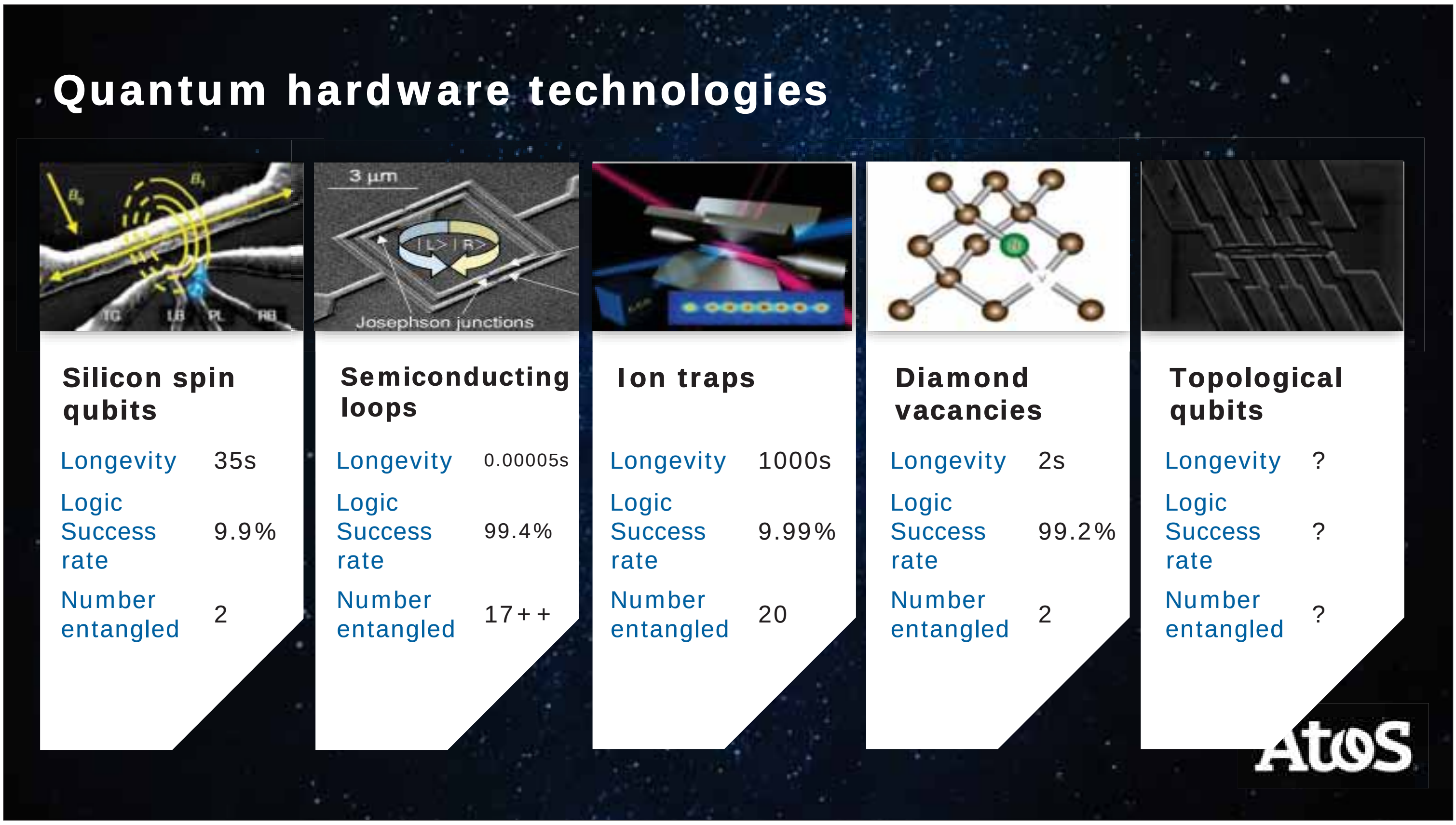
Atos

Quantum computing speedup



Atos


Why do we need quantum computing?
Cryptography
Machine Learning
Finance
Pharmaceutical
Chemistry
Combinatorial optimisation
Oil prospection
Much more are expected in the next few years...
Atos


Quantum hardware technologies
Silicon spin qubits
Longevity 35s
Logic Success rate 9.9%
Number entangled 2
Semiconducting loops
Longevity 0.00005s
Logic Success rate 99.4%
Number entangled 17++
Ion traps
Longevity 1000s
Logic Success rate 9.99%
Number entangled 20
Diamond vacancies
Longevity 2s
Logic Success rate 99.2%
Number entangled 2
Topological qubits
Longevity ?
Logic Success rate ?
Number entangled ?
Atos

The Atos Quantum learning machine

A must-have to prepare your future quantum algorithms

- A quantum simulator, it is not a quantum computer
- Allows Quantum algorithms development without quantum hardware constraints
- **Simulates** Quantum Processing Units
- Fully debugged development environment

Atos

Atos: a leader in HPC and pioneer in quantum solutions

Some of our recent successes

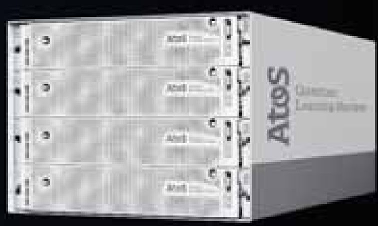
Current QLM use cases include

- Pick and Grab logistics
- Aero Engine design
- Flight Dynamics
- Life Sciences
- University driven projects
- Battery Design e-cars
- Quantum secure comms

SURF SARA

BAYER

cea



ccrt

fh OBERÖSTERREICH

Hartree Centre
Science & Technology Facilities Council

OAK RIDGE National Laboratory

JÜLICH Forschungszentrum

RWTH AACHEN UNIVERSITY

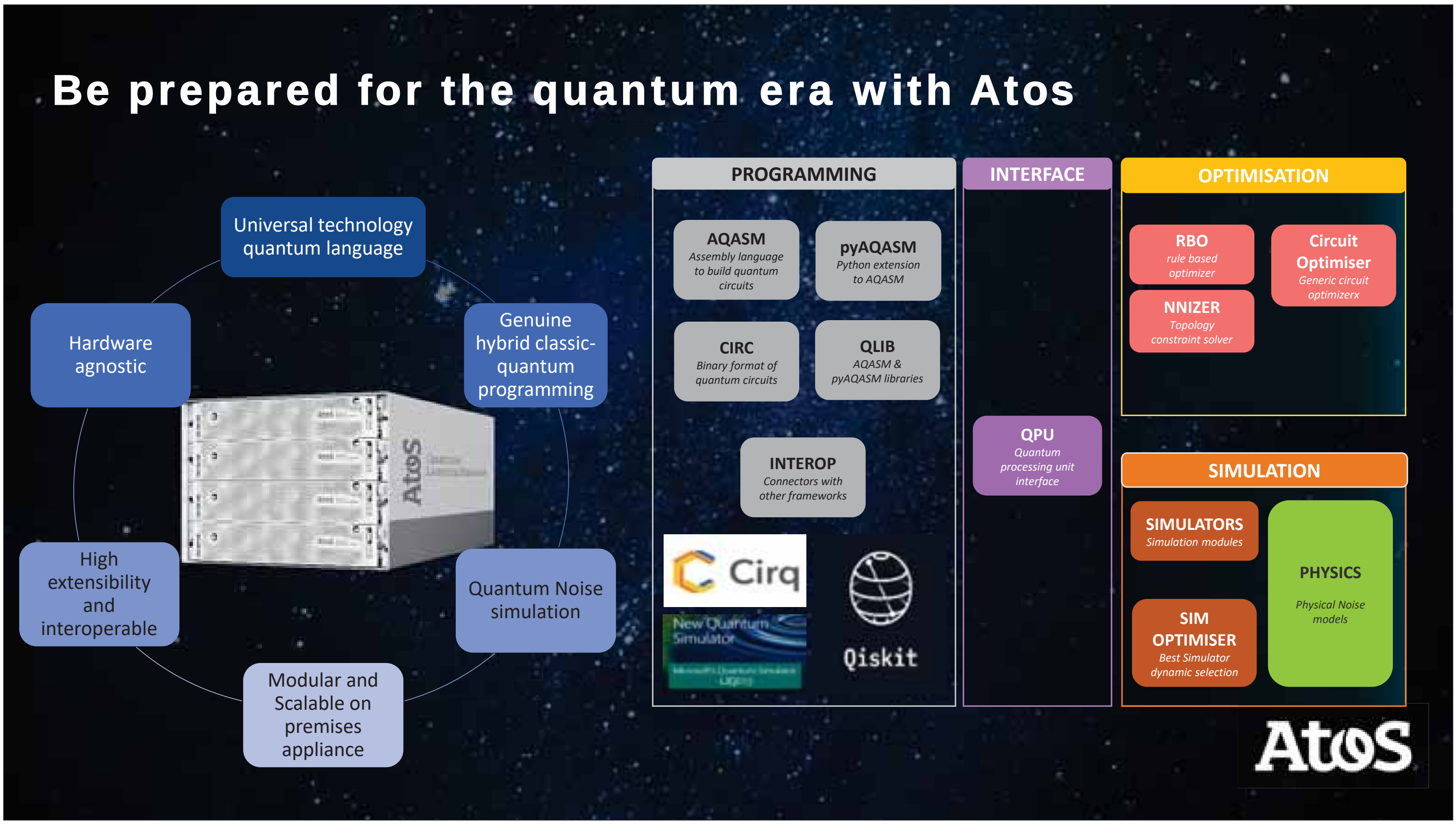
DTU

Argonne NATIONAL LABORATORY

TOTAL

UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

Atos


Be prepared for the quantum era with Atos
Universal technology quantum language
Hardware agnostic
Genuine hybrid classic-quantum programming
High extensibility and interoperable
Quantum Noise simulation
Modular and Scalable on premises appliance
PROGRAMMING
AQASM
Assembly language to build quantum circuits
pyAQASM
Python extension to AQASM
CIRC
Binary format of quantum circuits
QLIB
AQASM & pyAQASM libraries
INTEROP
Connectors with other frameworks
Cirq
New Quantum Simulator
Qiskit
INTERFACE
QPU
Quantum processing unit interface
OPTIMISATION
RBO
rule based optimizer
NNIZER
Topology constraint solver
Circuit Optimiser
Generic circuit optimizerx
SIMULATION
SIMULATORS
Simulation modules
SIM OPTIMISER
Best Simulator dynamic selection
PHYSICS
Physical Noise models
Atos


Atos
ZAPATA
Atos and Zapata Computing partner to deliver full-quantum computing solution
Atos

The leader in quantum computing algorithms

>> ~30 people, with 15 PhDs

>> over 20,000+ academic citations of our papers in quantum computing

>> 4 locations: Boston, Toronto, *Europe*, *Japan*

>> Fortune 100 customers

>> 10+ quantum hardware partners

>> $25m+ in funding

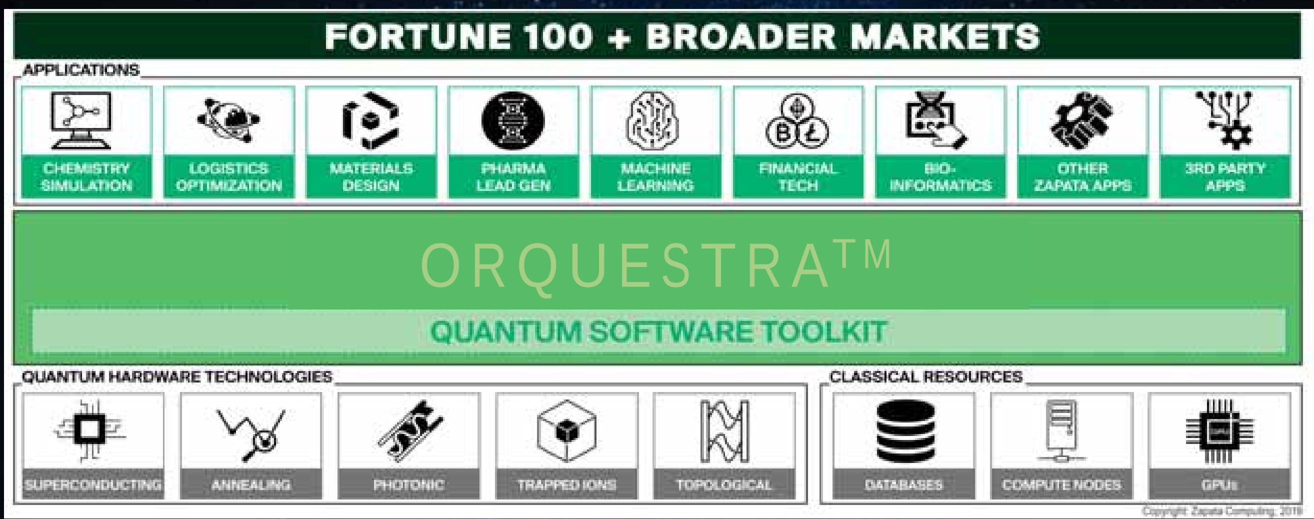
>> 30+ proprietary algorithms

Founded in 2017 and based on technology developed at Harvard University, Zapata Computing is the leading enterprise software company for quantum solutions.

Zapata's software platform Orquestra™ offers workflows and quantum algorithms for the next generation of high-performance computing in industries like oil & gas, finance, aerospace.

COMCAST VENTURES · Prelude VENTURES · pitango VENTURE CAPITAL · THE ENGINE · pillar · BASF The Chemical Company · BOSCH

The go-to quantum applications platform



Atos


myQLM
• Join the Quantum Programmer Community
Atos
Atos


myQLM
A complete toolset to build your own Quantum Algorithms
Discover AQASM and pyAQASM
WRITE
your own quantum algorithms
jupyter
Explore Jupyter Notebook tutorials
Adapt QLIB algorithms
On your laptop using pyLinalg or your own simulator
RUN & TEST
your quantum circuits
On your Atos QLM
Create myQLM user communities
SHARE
tips and codes with the community
Collaborate with other frameworks' users
Atos

What functionalities are included in MyQLM?

The programming environment of the Atos QLM with open source simulators

PROGRAMMING

AQASM
Publication of the syntax + compiler/decompiler to competing opensource platforms

pyAQASM
Source code availability under license for QLM customers

CIRC
Binary format of quantum circuits
Published specs and Interop Kit with C++, Python, js, Java...

QLIB
AQASM & pyAQASM libraries
QRAM, oracle emulator, arithmetic libraries and more!

rigetti

INTEROP
Connectors' source codes: build your own!

Open Source

ProjectQ

Cirq

SIMULATION

Open Source

SIMULATOR
pyLinalg
Source code of this simulator: build your own!

Atos

Thank you

For more information, please contact:

Shane Rigby
M+44 7970 125855
shane.rigby@atos.net

Visit Atos on stand 21
Proud to be Gold Sponsor of CIUK 2019

Atos

Josh Borrow

Institute for Computational Cosmology, Durham University

Post-Processing Tools for Next-Generation Cosmological Simulations

Cosmological simulations are used by researchers to study the formation and evolution of galaxies within the context of the cosmic web. They are some of the most computationally taxing simulations in the entire field of physics, often running on many thousands of cores for months at a time. The next generation of simulations will produce over a petabyte of on-disk data per simulation, with individual objects to be studied representing only megabytes of data. Thus, a suite of efficient and user-friendly tools to extract only the necessary data are required. Here, we present swiftsimio, a tool-kit for the open-source SWIFT cosmological simulation, which contains science analysis, visualisation, and post-processing submodules. We show how it can be used to efficiently analyse any size of simulation through its integration with SWIFT thanks to custom metadata produced on the fly while running the main code.

SWIFTsimIO

Post-processing the next-generation of cosmological simulations

Josh Borrow

Institute for Computational Cosmology, Durham University

Gas Density

Temperature

Dark Matter

Shocks

Enrichment

Movitation

We run **large-scale simulations** of the universe with up to **100 billion particles**.

Largest simulations now producing **petabytes of data**.

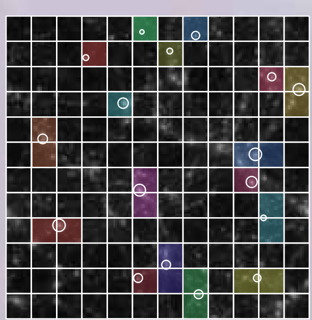
Individual snapshots are around **10 TB**, and represent a **huge dynamic range**.

However, **individual objects** (galaxies) in these simulations **only represent < 1 GB** of data.

Need to **efficiently extract** these objects!

0

Metadata



Key to solving **big data challenges**: producing enough **metadata** to efficiently slice the data at a later stage.

Physicists think spatially – package (very cheap) **spatial metadata with outputs**.

Run **on-the-fly object finders** to deal with huge **dynamic range**, along with a **top-level grid**.

Store each file **ordered by top-level cell**.

Figure 1: Left shows the top-level cell grid (projected in 2D) of a typical cosmological volume simulated with the SWIFT code. Objects identified by the on-the-fly object finder are shown as white circles, with the top-level cells identified by swiftsimio to read from the snapshot highlighted in various colours. This reduces the data size sigfnicantly; each cell contains only around a hundred thousand particles.

1

Reading Data

Metadata is **stored for every object** in the simulation, including properties such as mass, size, temperature, etc. such that **often it is not necessary to go back to the particle data**.

When it is necessary, thanks to the **spatial metadata**, the **time to read** a fixed volume of data is **completely independent** of the **size of the dataset** (see Figure 2).

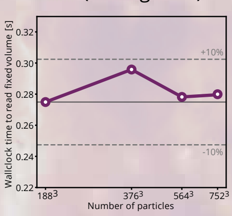


Figure 2: Left shows the (constant) time to access a fixed volume of data (here much larger than any object in a typical simulation) as a function of the size of the dataset. The rightmost dataset represents over 100 Gb of particle data. The dashed lines show a range of ±10% in read time.

2

Visualisation

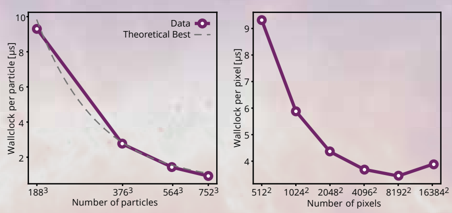


Figure 3: The left panel shows the cost of making a fixed 4096x4096 image of a dataset of different sizes. The cost per particle actually decreases as the (particle) resolution increases as each particle is smoothed over fewer pixels. `swiftsimio` shows very close agreement to theoretical best scaling here. The right panel shows how the cost per pixel scales as a function of the image resolution for a fixed particle count (188^3). This should be constant, but overheads dominate for small images, with large images generally being cheaper.

Once loaded, **very cheap to visualise** data.

Accelerated routines with numba to produce SPH-smoothed visualisations.

See background for examples!

Visualisation is **not just for** making pretty **pictures**; **projected quantities** map directly to **astronomical observables**.

3

Conclusion

`swiftsimio` allows users of the SWIFT simulation code deal with **huge snapshots trivially** through the use of **spatial metadata**.

It turns a **petascale big data** analysis problem into something **simple** to perform even on a **laptop computer**.

The code is **available** on GitHub (`swiftsim/swiftsimio`) and on PyPI.

4

5 million light-years

Matthew Carter

University of Liverpool, Big Data and High-Performance Computing MSc Student

Determination of the Usability of FPGA Technology to Accelerate Option Pricing Algorithms

This project explores the feasibility of using FPGAs to accelerate numerical options pricing algorithms. Specifically, efficient implementations of Binomial Tree and Monte Carlo options pricing algorithms were illustrated on a FPGA. Prototypes of these algorithms were developed on a Zynq-7020 FPGA and compared against a Cortex-A9 CPU. A full implementation was developed on an Alveo u200 datacentre card and compared against an Intel Xeon E5649. The speed-up of these algorithms range from 1.1x to 20x on the Zynq-7020 and 0.12x to 13x on the Alveo u200.

Determination of the Usability of FPGA Technology to Accelerate Option Pricing Algorithms

Matthew Carter, Department of Computer Science, University of Liverpool
MSc Big Data and High-Performance Computing
Primary Supervisor: Dr Michael Bane; Secondary Supervisor: Professor Jeremy Smith

Abstract

High-performance computing is an increasingly important topic in the world of finance, particularly in the field of options pricing. As CPU performance plateaus, it is imperative that financial institutions look towards alternative hardware, such as Field Programmable Gate Arrays (FPGAs), to satisfy their demands. This Master's project explores the feasibility of using FPGAs to accelerate numerical options pricing algorithms. Specifically, implementations of Binomial Tree and Monte Carlo options pricing algorithms were illustrated on a FPGA. We demonstrate how Taylor approximations can be used to further accelerate algorithms whilst minimising resource usage.

Prototypes of these algorithms were developed on a Zynq-7020 FPGA and compared against a Cortex-A9 CPU. A full implementation was developed on an Alveo u200 datacentre card and compared against an Intel Xeon E5649. The speed-up of these algorithms range from 1.1x to 20x on the Zynq-7020 and 0.12x to 13x on the Alveo u200. Figure 1 shows the runtime of the Binomial Tree algorithm to price European options on the Zynq-7020 and Alveo u200 against their respective CPUs. The FPGA implementation of the binomial tree algorithm is 1.79% whereas the FPGA implementation of the Monte Carlo algorithm is 1.78%.

Field Programmable Gate Arrays

Field Programmable Gate Arrays (FPGAs) are a form of reprogrammable hardware that can be configured by a user to perform a desired function. This makes them much more versatile when compared to custom hardware such as ASICs that are designed to perform specific functions. Due to their reprogrammable and parallel nature, FPGAs have become a topic of interest in several domains ranging from consumer electronics to high-performance computing.

Pynq-Z2 (left) Alveo u200 (right)

- The Zynq-7020 FPGA used throughout this project was embedded into a Pynq-Z2 development board and the Alveo u200 was installed in a server at the University of Liverpool.
- Both the development board and datacentre card were donated by Xilinx as part of the Xilinx University Programme.

Hypotheses

Primary Aims

1. Determine the speed-up, accuracy and energy efficiency of FPGA implementations of Monte Carlo and binomial tree algorithms to price European options in comparison to CPU implementations.
2. Compare the FPGA algorithms to determine which yields the greatest speed-up and accuracy whilst minimising energy use.

Secondary Aims

1. Implement both Monte Carlo and binomial trees algorithms to price American options and perform the analysis described above.

Experiments Performed

- Experiments were performed to determine the run time, accuracy and energy usage of the previously mentioned algorithms.
- The run time of the algorithms was measured using the timing functions in Python, C++ and OpenCL.
- The accuracy of the Zynq-7020 implementation was determined by comparing its output against the output of an Arm Cortex-A9 processor.
- The accuracy of the Alveo u200 implementation was determined by comparing it against an Intel Xeon E5649 processor.
- The quality of the algorithms solution was assessed by comparing their output against the closed form solution determined through the Black-Scholes formula.
- The experiments performed to measure energy usage were largely inconclusive due to the age of the CPU architectures used.

Accuracy of Algorithms

	Platform		
Algorithm	CPU	Zynq-7020	Alveo u200
EU Binomial Tree	2.78541565	2.78521299	2.78521300
US Binomial Tree	8.01245689	8.01226807	8.01227000
EU Monte Carlo	2.76810241	2.76814928	2.76810290

Table 1: Estimated Stock Price ($) on CPU, Zynq-7020 and Alveo u200

- Stock screeners such as Yahoo! Finance display information to 4 decimal places.
- Both binomial trees algorithms are accurate to three decimal places whereas the Monte Carlo algorithm is accurate to four.

Time Performance

The following steps were taken to improve runtime performance:

- Loop pipelining and unrolling to exploit parallelism.
- Storing data in BRAM on the FPGA device.
- Partitioning arrays to increase local memory bandwidth.

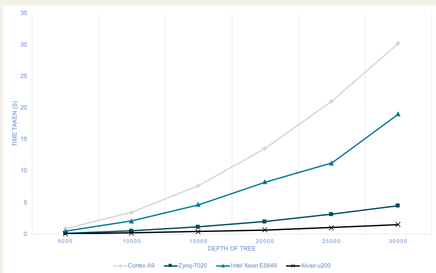


Figure 1: Runtime of Binomial Trees Algorithm with Varying Depth

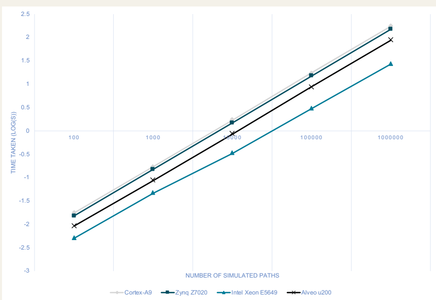


Figure 2: Runtime of Monte Carlo Algorithm with Varying Number of Paths

- Good speedups were achieved for the binomial trees algorithm.
- Further work is needed to achieve similar results for the Monte Carlo algorithm.

Conclusions and Future Work

The algorithms built throughout this project demonstrate how FPGAs can be used to accelerate options pricing libraries, whilst maintaining a similar accuracy to the CPU implementations. A summary of the speedup achieved by each algorithm is shown in Table 2.

	Speedup	
Algorithm	Zynq-7020	Alveo u200
European Binomial Tree	2x	3x
US Binomial Tree	5x	1.6x
European Monte Carlo	-	0.09x

Table 2: Summary of Timing Results

Suggestions for future work

- Gain a deeper knowledge of how optimise algorithm designs for FPGAs
- Repeat the analysis using more modern CPU architectures

MSc Big Data and High-Performance Computing

The Big Data and High-Performance Computing Master's programme is run by the Department of Computer Science at the University of Liverpool and has the option of spending a year in a relevant industrial setting. The modules were designed with input from the Hartree Centre and intend to equip students with knowledge and skills that are increasing in demand world wide.

Topics include research methods, applied algorithmics, data mining and visualisation, machine learning, multi-core and multi-processor programming, and optimisation. Students learn through a combination of taught lectures, assessments and group projects. They have the opportunity to demonstrate their knowledge and research skills in an independent project such as the one described throughout this poster.

The University of Liverpool works with a number of industrial partners such as automobile manufacturers, high-performance computing service providers, chip manufacturers and data scientists. The University is actively seeking opportunities to improve our teaching, ensuring that it is continually relevant to data science, artificial intelligence and high-performance computing and welcome discussions with other industrial partners.

To discuss further please contact Dr Michael Bane (m.k.bane@liverpool.ac.uk), alternatively you can follow the following QR code.

Affiliations

UNIVERSITY OF LIVERPOOL

XILINX

Contact Details

Matthew Carter: m.j.carter2@liverpool.ac.uk

The final thesis for this Master's project can be found on Dr Michael Banes university website, which can be found by searching *mkbane* on Google. All code produced throughout this project are on Github, *mjcarter95*.

Julita Inca Chiroque

University of Edinburgh

Benchmarking the performance of HPC architectures using CP2K

CP2K is a chemistry application that performs atomistic simulations that can vary from solid and liquid states to biological systems. This work present basically two stages: the compilation of CP2K, and benchmark the performance of CP2K on different HPC architectures. For this purpose, three samples of water were used (64, 128 and 256 molecules) as well as the parallel approaches MPI, OpenMP and hybrid.

Jake Foster

University of Birmingham,
Advanced Research Computing

Advanced Research Computing - from a Year in Industry Student

I am a Computer Science undergraduate student currently on my year in industry. My placement is with the Advanced Research Computing (ARC) team at the University of Birmingham. My poster describes the year in industry placement programme run by ARC, along with its other student work opportunities. It also covers the first major project I've been involved with. An automatically-generated website documenting all the applications installed on our HPC systems BlueBEAR, BEAR Cloud and CaStLeS.

UNIVERSITY OF BIRMINGHAM

ARC
Advanced Research Computing

BEAR
BIRMINGHAM ENVIRONMENT FOR ACADEMIC RESEARCH

Created by Jake Foster

I am a Computer Science undergraduate student currently in my year in industry. My placement is with the Advanced Research Computing (ARC) team at the University of Birmingham.
This is an overview of the different opportunitites for students at ARC and the skills you will gain from this experience, as well as an insight into the first project I worked on over the summer and the tools that were required to achieve the end result.
ARC is committed to trying to develop up and coming HPC sysadmins and RSEs to work in the sector.

Students In ARC

The University of Birmingham has actively been trying to encourage more students into the field.

They offer both part time work and 12 month industrial placements for students.

Part Time Work

At ARC, you have the opportunity to work flexible hours around your studies. You are given a project to work on over an extended period of time. It is a great opportunity to gain real life experiences and make some money during your studies.

Industrial Placement

Over the course of the year, you will be a part of the engagement team, the research software team and the infrastructure team. This provides a wider range of applicable skills than a typical job.

Over my first few months I was a member of the Research Software Group. I was given the task of redesigning the Application Documentation. The web pages were outdated and the information inaccurate. For more information, see 'My first project'.

I have also assisted with multiple inductions, informing the new staff about the services Advanced Research Computing can offer them.

My first project – Apps Documentation

The website documenting all the applications installed on our HPC systems BlueBEAR, BEAR Cloud and CaStLeS on bluebear was sporadically maintained and therefore constantly out of date and inaccurate.
The goal was to create a new, *fully automated* website that would display the new data automatically on application installs.
The project was split into multiple stages:

1. Scraping the data

All data from the old website was extracting using *BeautifulSoup* – A Python3 web scraping library, and stored in JSON files. At this point in the project, we had not decided how the final project would be constructed so we wanted the data to be in an easily accessed data structure.

2. Creating the new webiste

The new website was built using *Django 2*. All the previously scraped data was stored in a postgres database.

3. Maintaining the information

When we have installed new software on BEAR we run a script that checks the permissions etc. for any software that the script hasn't seen before.
It also automatically adds any newly installed software to our bear-apps Database.

SCAN ME
https://bear-apps.bham.ac.uk

Guido Giuntoli

Barcelona Supercomputing Centre

Hybrid CPU/GPU FE2 Multi-Scale Approach applied to Aircraft Wing Panels

This poster presents the results of a new implementation of the Finite-Element Squared (FE2) multi-scale algorithm that is achieved by coupling macro-scale and micro-scale to simulate the behaviour of composites used in aircraft wing panels. The multi-physics code Alya is used for the macroscale in its MPI version only and is coupled to the micro-scale code, Micropp, which runs on CPU/GPU. The computational performance has been derived from results obtained on the CTE-POWER cluster of the Barcelona Super Computing Center (IBM POWER9 + V100 Nvidia GPUs). They show that this new method o er good scalability for real size industrial problems and also that the execution time is dramatically reduced using GPU-based clusters.

Authors: Guido Giuntoli, Judicael Grasset, Charles Moulinec, Stephen Longshaw, Mariano Vazquez, Guillaume Houzeaux & Sergio Oller

Hybrid CPU/GPU FE2 Multi-Scale Approach applied to Aircraft Wing Panels

Guido Giuntoli [1], Judicaël Grasset [2], Alejandro Figueroa [3], Charles Moulinec [2], Stephen Longshaw [2], Guillaume Houzeaux [1], Mariano Vázquez [1] & Sergio Oller [4]

[1]Barcelona Supercomputing Center (BSC), Spain, [2]STFC, Daresbury Laboratory, UK, [3]George Mason University, USA & [4]Universitat Polytècnica de Catalunya, Spain

Objective

- Solve the largest problem possible with the FE2 multi-scale method using massive parallel computing. Meshes of around 1 M elements for the macro-scale and 100^3 elements for the micro-scale are considered.

Numerical Method

Macro-Scale
(Only CPUs and MPI)

$$\begin{cases} \nabla \cdot \overline{\sigma} = 0 \\ \overline{u} = \overline{u}_d \\ \overline{\sigma} \cdot \hat{n} = \overline{\sigma}_n \cdot \hat{n} \\ \overline{\epsilon} = \nabla_s \overline{u} \\ \overline{\sigma} = \ldots \end{cases}$$

$$\begin{cases} \nabla \cdot \sigma = 0 \\ \sigma = f(\epsilon, q) \\ u_d = \overline{\epsilon} \cdot x \\ \overline{\sigma} = \frac{1}{V} \int_\Omega \sigma \, dV \end{cases}$$

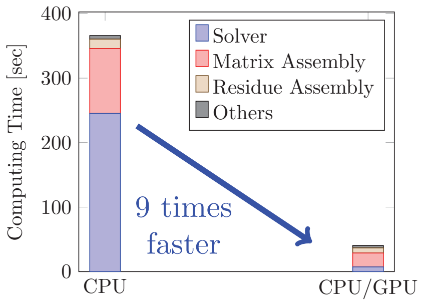
Micro-Scale
(CPU + GPU)

- The main reason of why HPC is needed is because one micro-scale FE problem should be solved for each Gauss point in the macro-scale at each time step (millions of FE problems).

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

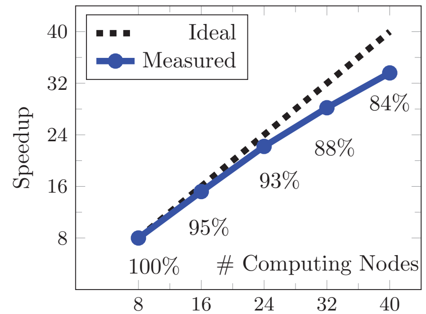
Improvements with CPU/GPU

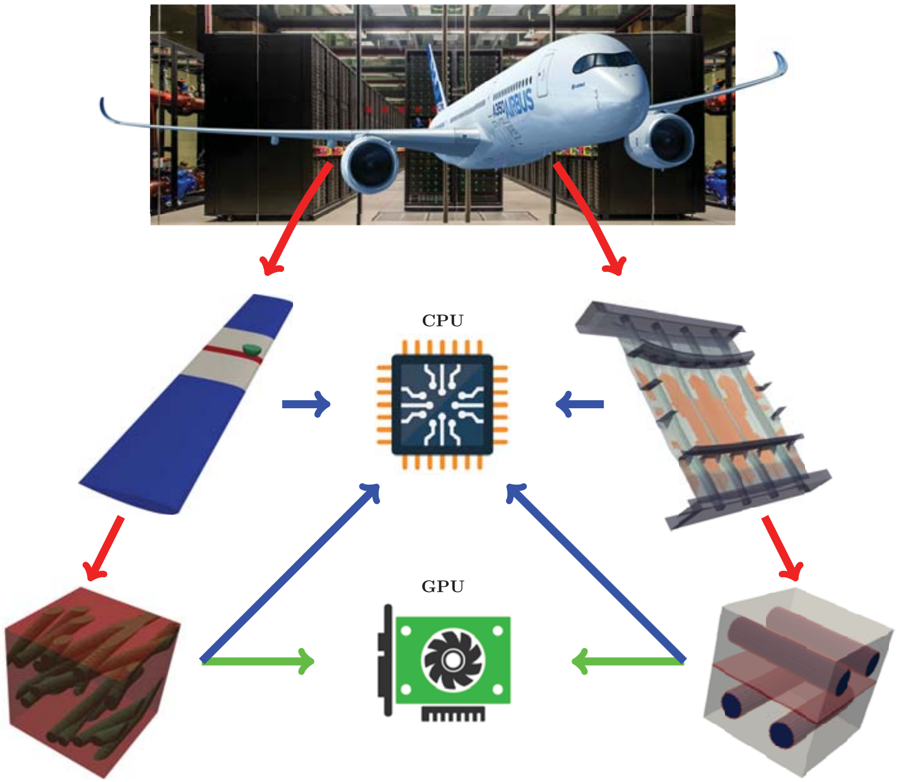
The micro-scale code Micropp, the most computational intensive part, was ported to CPU/GPU for accelerating the entire execution:



Parallel Performance

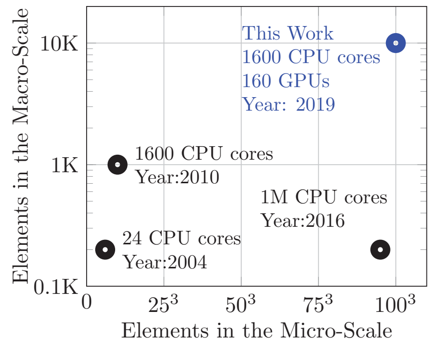
Micropp was coupled with Alya for solving the multi-scale. The application scales well up to 40 computing nodes (160 V100 Nvidia GPUs):





Conclusion

- Our implementation has solved the largest case ever achieved, 10 K elements in the macro-scale and 100^3 elements in the micro-scale:



- The application (Alya + Micropp) is ready to solve problems of 1 M elements in the macro-scale and 100^3 elements in the micro-scale (a total of 10^{13} degrees of freedom). A simulation in Summit machine using 24576 GPUs is planned to be performed.

Acknowledgements

To the Barcelona Supercomputing Center, to the STFC, Daresbury Laboratory, UK and to the 9th HPC-Europa3 Transnational Access programme.

Contact Information

- Web: bit.ly/2OoLeDA
- Email: gagiuntoli@gmail.com
- LinkedIn: linkedin.com/in/gagiuntoli/

Science & Technology Facilities Council

Judicael Grasset

Science and Technology Facilities Council

GPU offloading experiments for TELEMAC-MASCARET

TELEMAC-MASCARET is an open-source suite of hydraulics solvers for free-surface flow modelling originally developed by EDF and now by the TELEMAC-MASCARET Consortium. It can be used to simulate 2-D or 3-D flows, sediment transport, water quality, wave propagation in coastal areas and river. More details on the possible applications can be found on the website:[1]

Currently TELEMAC-MASCARET is parallelised with MPI and is only able to use CPUs. Current computer trends favour an increase in the number of cores in CPU, which will be handled by MPI, but also an increase in the use of accelerators, such as GPUs, which TELEMAC-MASCARET currently has no ability to use. If TELEMAC-MASCARET was run on the current top super computer in the world, Summit (Oak Ridge National Laboratory, USA), it would not be able to use most of the computing power of the machine since most of that power come from the 6 GPUs available on each node. While computers like Summit are not typical environments for TELEMAC-MASCARET users, they give us a strong hint of the future architectures on which it should be prepared to run on. In this poster we will present experimental work done on the TOMAWAC module. In particular the offloading of a computationally expensive subroutine used in a demonstrational test-case of the suite will be shown. The ooading is done with pragma-based programming method, using OpenACC on Paragon [2] an IBM OpenPOWER cluster (2 Power8 CPUs and 4 Nvidia P100 GPUs per node). We will present how these modifications can be used to either improve the precision of results or reduce the execution time. Finally, a work in progress of the offloading of a real test case provided by EDF R&D will be shown.

[1] www.opentelemac.org
[2] www.hartree.stfc.ac.uk/Pages/Our-systems-and-platforms.aspx

GPU offloading experiments for TELEMAC-MASCARET

Judicaël Grasset, Stephen Longshaw, Charles Moulinec
Scientific Computing Department, UKRI-STFC Daresbury Laboratory, Warrington, UK

Introduction

TELEMAC-MASCARET is an integrated suite of computational fluid dynamics (CFD) solvers for use in the field of free-surface flow. It currently utilises MPI parallelism.

This work demonstrates that some parts of the suite can benefit from being offloading to GPUs. Specifically this is aimed towards computationally intensive but repetitive loops within the code and is achieved using pragma-based OpenACC programming directives.

Software & machine

-TELEMAC-MASCARET V8P1 (from 2019-11-25)
-PGI compiler 19.10
-2 IBM Power8 (16 cores) & 4 NVIDIA P100, per node.

Implementation and Results

Original Code (taken from qnlin3 subroutine)

Computing the variable *k* is computationally complex. Arrays involved are accessed in a non-contiguous, random, fashion. This results in significant cache misses and CPU stalls.

Some users of the suite have reported this loop as the bottleneck of their simulation.

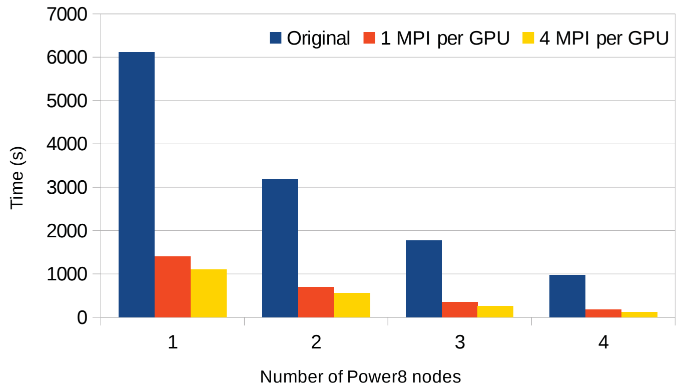
```
do loop
    do loop
        do loop
            do loop
            array(x,y,z) = array(x,y,z) + k
```

Modified GPU code (using OpenACC)

The array is first copied to the GPU, the four imbricated loops are then flattened into one and executed in parallel on the GPU. An atomic directive surrounds the array update to avoid thread race conditions.

Results from this acceleration can be seen in the graph below, showing speed-up achieved with both 1 and 4 MPI processes per GPU.

```
!$acc data copy(array)
!$acc parallel loop collapse(4)
do loop
    do loop
        do loop
            do loop
            !$acc atomic
               array(x,y,z) = array(x,y,z) + k
```



Conclusion

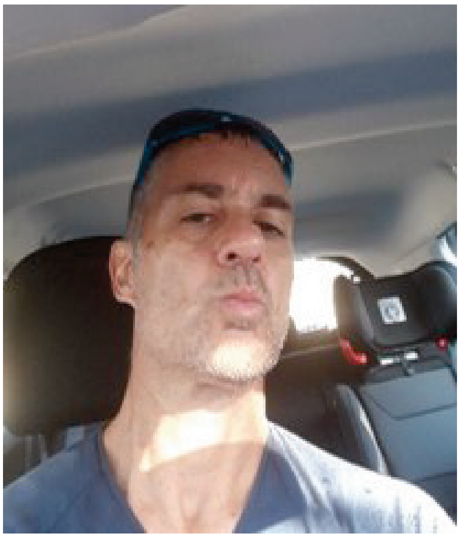
Offloading portions of the TELEMAC-MASCARET suite to GPUs is clearly beneficial. Problems that use the qnlin3 subroutine see performance increases of between 5.5x and 8x.

Importantly, this work can be easily incorporated into the existing and widely used open-source codebase due to its pragma-based approach as it involved no re-write of the original code and directives are only activated when a compiler flag is supplied.

UKRI Science and Technology Facilities Council

Acknowledgments
This work is supported by the Hartree Centre through theInnovation Return on Research (IROR) programme.

Further information



Piero Lanucara

CINECA

ChEESE, A Center of Excellence for Exascale in Solid Earth

This proposal aims at establishing a Center of Excellence (CoE) to prepare state-of-the-art codes and develop related services for upcoming Exascale supercomputing in the area of Solid Earth (SE). ChEESE is addressing extreme computing scientific and societal challenges by harnessing European institutions in charge of operational monitoring networks, tier-0 supercomputing centers, academia, hardware developers and third-parties from SMEs, Industry and public-governance. The scientific challenging ambition is to prepare 10 open-source flagship codes to solve Exascale problems on computational seismology, magnetohydrodynamics, physical volcanology, tsunamis, and data analysis and predictive techniques, including machine learning and predictive techniques from monitoring earthquake and volcanic activity. The selected codes are audit and optimized at both intranode level (including heterogeneous computing nodes) and internode level on heterogeneous hardware prototypes for the upcoming Exascale architectures, thereby ensuring commitment with a co-design approach. Preparation to Exascale is considering also code inter-kernel aspects of simulation workflows like data management and sharing, I/O, post-process and visualization. In parallel with these transversal activities, ChEESE is sustenting on three vertical pillars. First, it develop Pilot Demonstrators for scientific challenging problems requiring of Exascale computing in alignment with the vision of European Exascale roadmaps. This includes near real-time seismic simulations and full-wave inversion, ensemble-based volcanic ash dispersal forecasts, faster than real-time tsunami simulations and physics-based hazard assessments for seismics, volcanoes and tsunamis. Second, Pilots are also intended for enabling of operational services requiring of extreme HPC on urgent computing, early warning forecast of geohazards, hazard assessment and data analytics. Selected

Pilots are tested in an operational environment to make them available to a broader user community. Additionally, and in collaboration with the European Plate Observing System (EPOS), ChEESE is promoting and facilitating the integration of HPC services to widen the access to codes and fostering transfer of know-how to Solid Earth user communities. Finally, the third pillar of ChEESE aims at acting as a hub to foster HPC across the Solid Earth Community and related stakeholders and to provide specialized training on services and capacity building measures.

A Center of Excellence for Exascale in Solid Earth

Computing Insight UK 2019 (CIUK 2019) poster competition

ChEESE
Centre of Excellence for Exascale in Solid Earth

10 new High Performance Computing Centers of Excellence (CoEs) have been created under H2020 e-Infrastructures. Among these, ChEESE is a 3-year project targeting at Solid Earth (SE) for the upcoming pre-Exascale (2020) and Exascale (2022) supercomputers.

15 Exascale Computational Challenges

ChEESE will address 15 scientific, technical and socio-economic Exascale Computational Challenges (ECC) in the domain of SE.

10 Flagship codes

10 different SE open source European codes have been selected in ChEESE:

- 4 in computational seismology: EXAHYPE, SALVUS, SEISSOL, SPECFEM3D
- 2 in magneto hydrodynamics: PARODY_PDAF, XSHELLS
- 2 in physical volcanology: ASHEE, FALL3D
- 2 in tsunami modelling: T_HYSEA, L_HYSEA

12 Pilot Demonstrators

ChEESE will develop 12 Pilot Demonstrators (PDs) and enable services oriented to society on critical aspects of geohazards like hazard assessment, urgent computing, and early warning forecast.

- Urgent Seismic Simulations
- Faster Than Real-Time Tsunami Simulations
- High-Resolution Volcanic Plume Simulation
- Physics-Based Tsunami-Earthquake Interaction
- Physics-Based Probabilistic Seismic Hazard Assessment (PSHA)
- Probabilistic Volcanic Hazard Assessment (PVHA)
- Probabilistic Tsunami Hazard Assessment (PTHA)
- Probabilistic Tsunami Forecast (PTF) for Early Warning and Rapid post Event Assessment
- Seismic Tomography
- Array-Based Statistical Source Detection and Restoration and Machine Learning from Earthquake/Volcano Slow-Earthquakes Monitoring
- Geomagnetic Forecasts
- High-Resolution Volcanic Ash Dispersal Forecast

Integrate

ChEESE will integrate around HPC and HDA European institutions in charge of operational geophysical monitoring networks, tier-0 supercomputing centers, academia, hardware developers, and third-parties from SMEs, Industry and public governance bodies (civil protection), and pan-European infrastructures, such as the European Plate Observing System (EPOS) and EUDAT.

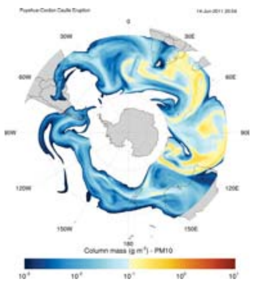


Image from: FALL3D

Image from: Salvus (https://salvus.io)

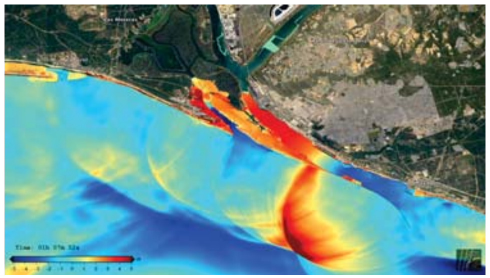


Image from: T-Hysea (https://edanya.uma.es/hysea)

Partners:

Barcelona Supercomputing Center – Centro Nacional de Supercomputación; Bull atos technologies; CINECA; Icelandic Meteorological Office; IPGP; Istituto Nazionale di Geofisica e Vulcanologia; LMU Ludwig-Maximilians-Universität München; CNRS; NGI; ETH zürich; Technical University of Munich TUM; Universidad de Málaga; HLRS

The ChEESE project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement Nº 823844.

www.cheese-coe.eu
cheese-coe@bsc.es
ChEESE CoE
@Cheese CoE

Alexander Lloyd

University of Birmingham

Building the next generation JupyterHub platform for research

At the University of Birmingham, there is a need for access to computing resource without requiring Linux command line skills. Jupyter notebooks provide an intuitive user interface for users to develop and run their code from a web browser. Our next-generation deployment utilises Kubernetes to orchestrate these Notebooks with pre-configured environments. It also provides the functionality to submit these jobs to the universities HPC cluster.

Building the next generation JupyterHub platform for research.

BEAR
BIRMINGHAM ENVIRONMENT FOR ACADEMIC RESEARCH

Alexander Lloyd, University of Birmingham

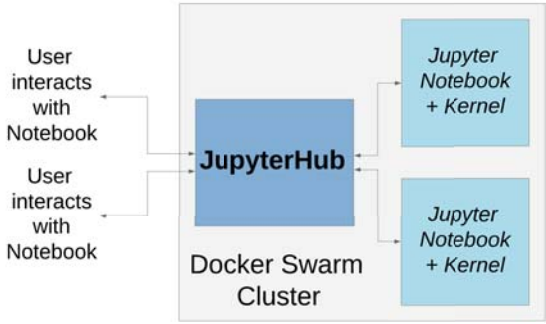
Background

Jupyter Notebooks have become more popular than ever providing a friendly programming environment with access to vast computing power at the click of a button. We wanted to provide a federated environment where users can use Jupyter notebooks and submit their projects to BlueBEAR – The University of Birmingham's Super Computer. This project is to support the ever increase data science workloads and make access to supercomputing resources easier for non-traditional users.

Result

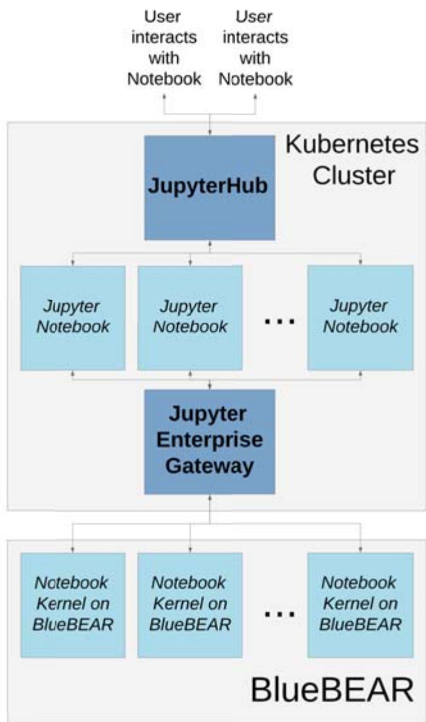
First Iteration:

- Docker Swarm cluster to orchestrate containers.
- JupyterHub authenticates using hosts PAM daemon.
- Mounted Users home directories inside notebook container.
- Each notebook container used a large amount of disk space – did not scale well.



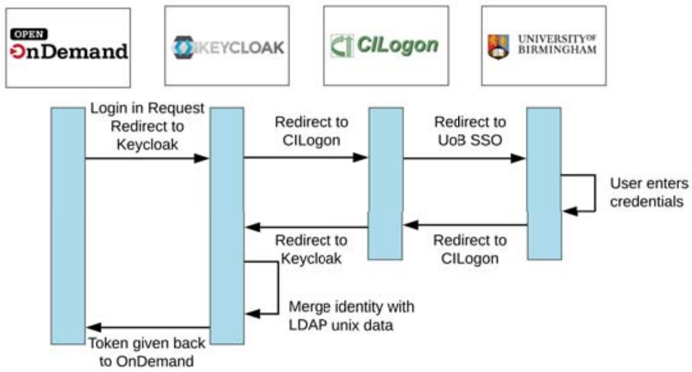
Second Iteration:

- Using Jupyter Enterprise Gateway to separate the kernel and notebook.
- Use Kubernetes to orchestrate JupyterHub and Jupyter Enterprise Gateway.
- Users can easily change the kernel running from inside the notebook.
- We can run the kernels anywhere! E.g. on BlueBEAR!
- Use Keycloak with CILogon. Using the University's SSO and LDAP to authenticate users.
- Using Spectrum Scale Storage driver will enable users home directory to be mounted inside containers. Using native Spectrum Scale driver rather than mounting NFS folder.



Current Implementation:

- Notebooks deployed with Open-OnDemand.
- Launch Jupyter Notebook, MATLAB and R Studio straight from the browser.
- OnDemand launches the Nginx process as the user → completely isolated user permissions.



Next Steps

- Cluster Keycloak (HA).
- Cluster Open-OnDemand (HA).
- Use Kubernetes for other workloads in BEAR, e.g. CI from Gitlab.

Acknowledgement
The work presented in this poster was funded by Advanced Research Computing as part of their undergraduate employment programme to build skills in the community.

UNIVERSITY OF BIRMINGHAM

Pawel Markiewicz

University College London

High performance image reconstruction and analysis platform for PET imaging in large clinical trials

The aim of the project is the development of fast and high-throughput image reconstruction and analysis software solution which, for the first time, enables performing customised and task-oriented imaging on large scale clinical trials, such as the Dementias Platform UK (DPUK). The DPUK network includes 7 centres equipped with simultaneous positron emission tomography and magnetic resonance imaging (PET/MR) scanners used for in-vivo imaging of the brain in neuro-degeneration.

The key features of this open-source software include state-of-the-art physical models of the tomographic data acquisition implemented on graphical processing units (GPU). This enables fast processing of large datasets (approximately 10 GBs per scan) multiple times with varying model parameters for high accuracy and precision quantitative image reconstruction and analysis. Such processing has already revolutionised uncertainty estimation of imaging endpoints used for high precision dementia imaging in feasible times. Furthermore, since the DPUK network consists of different scanning technologies, the software implementation inherently accommodates for the varying technologies using unified photon detection models, allowing thus for more harmonised multi-centre clinical trials.

Since all the fast GPU routines are available in Python, this enables researcher to fast prototype novel imaging methods (see below the link to the software documentation for more details). This software is currently used at UCL for the first DPUK dementia large cohort study as well nationally for the DPUK network. Also, it has been fully adapted by the large PET centre at the Washington University, USA and the Technical University of Munich, Germany.

The software documentation can be found at https://niftypet.readthedocs.io.

Ahmed Ammar Naseer

The School of Engineering,
The University of Manchester

Virtual Prototyping in the Cloud

Despite the widespread usage of cloud computing in various disciplines, one area in which cloud computing is not harnessed to the full extent is in engineering simulations. It is imperative the engineering community exploit this agile, scalable and reliable paradigm. In this project an online pay-as-you-go service for finite element analysis is built and its potential use demonstrated through a proof of concept case study. A Linux Virtual Machine is configured on Microsoft Azure with the open source finite element analysis software – ParaFEM. A simple elastic problem is evaluated on up to eight virtual CPUs to test the parallel nature of the cloud. This is followed by the proof of concept case study – simulating a single realisation of a stochastic Monte Carlo Simulation of a graphite brick used in a nuclear reactor. It is observed that using eight virtual CPUs compared to one virtual CPU results in a compute time that is nearly four times faster. Furthermore, it is also demonstrated that a probabilistic simulation can be completed in the same time frame as a deterministic simulation on the cloud. The results suggest usage of cloud computing for engineering simulations could significantly reduce the time spent in the design phase, thereby allowing engineers to focus more on other areas of importance.

Ahmed Ammar Naseer
Dr Lee Margetts
The University of Manchester

VIRTUAL PROTOTYPING IN THE CLOUD

INTRODUCTION

Despite the widespsread usage of cloud computing in our daily life and in disciplines such as computer science, cloud computing is not harnessed to the full extent in the engineering sector.

This project aims to bridge this gap by building a pay-as-you-go service for finite element analysis using Microsoft Azure and demonstrate its potential use through a proof of concept case study.

WHAT IS CLOUD COMPUTING?

It is a computing paradigm by which computing resources could be remotely accessed on demand (Mell and Grance, 2011). There are a number of cloud computing service providers such as Amazon, Google, and Microsoft. For this project Microsoft's cloud platform Azure was chosen.

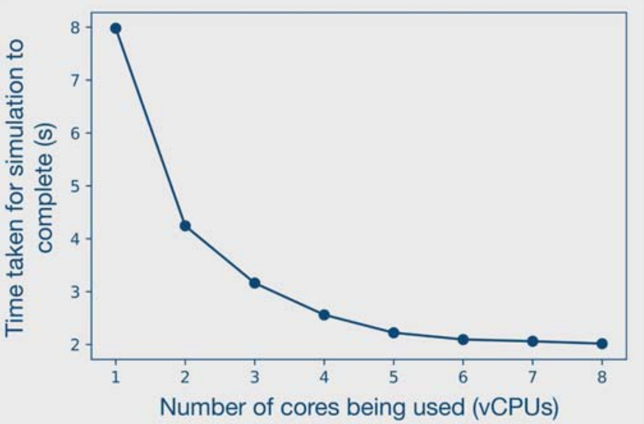


Fig 3: Results from the simple elastic problem showing the parallel performance of ParaFEM on Azure.

METHODOLOGY

1. An acount was made and setup on Microsoft Azure.
2. A Secure Shell (SSH) key pair was generated.
3. An instance of a Linux Virtual Machine (VM) was setup in Azure.
4. Open source FEA software ParaFEM was ported onto the Linux VM, compiled and thoroughly tested.
5. A simple elastic problem was run on one core (vCPU), followed by running it on two cores and subsequently on increasing number of cores upto 8 cores.
6. A single simulation of a Monte Carlo simulation was run for a graphite brick (resembling a thermo-mechanical problem) on one core.
7. The Virtual Machine was then stopped.

RESULTS & ANALYSIS

As seen from figure 3, when the simple elastic problem was analysed, with increasing number of cores time taken for simulation reduced. Indicating the program is behaving as expected.

The simulation of graphite brick took approximately 30 minutes on one core. As this is part of a Monte Carlo simulation, in practice 1000 such simulations maybe carried out in a single analysis. This equates to 21 days of work on a normal workstation. However, with Azure, 1000 cores can be provisioned and each simulation run simultaneously reducing total computing time to 30 minutes. Using 125 standard H8 machines on Azure, this would cost a company a maximum of approximately £90 for a similar analysis.

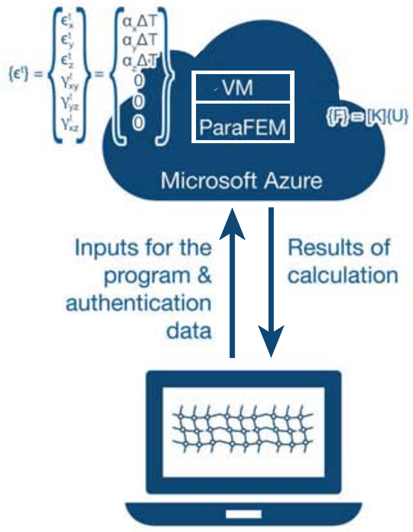


Fig 4: System architecture with two important equations. Equation on left is used to account for thermal strain in thermo-mechanical problem. Equation on the right is global load vector equation.

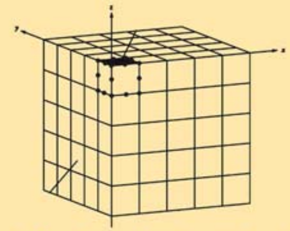


Fig 1: Mesh used in the simple elastic problem to test the parallel performance (Source: Smith et al., 2013).

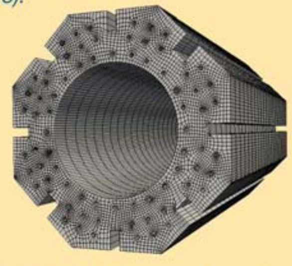


Fig 2: Mesh of the graphite brick used as the thermo-mechanical problem (Source: Arregui-Mena et al., 2015).

CONCLUSIONS

There are benefits of using more than one core for FEA.

With cloud computing, a probabilistic simulation can be completed in same time as a deterministic simulation.

FUTURE WORK

A web interface for users to interact with the software needs to be built.

REFERENCES

Arregui-Mena, J. D., Margetts, L., Griffiths, D., Lever, L., Hall, G. & Mummery, P. M. (2015). Spatial variability in the coefficient of thermal expansion induces pre-service stresses in computer models of virgin Gilsocarbon bricks. Journal of Nuclear Materials, 465, 793-804.

Mell, P. & Grance, T. (2011). The NIST definition of cloud computing.

Smith, I. M., Griffiths, D. V. & Margetts, L. (2013). Programming the Finite Element Method: Wiley.

CIUK 2019 POSTER COMPETITION WINNER

Muhammad Omer

The University of Manchester

Inspecting Bridges using Imaging, Virtual Reality and AI

Rapid urbanization, poor structural maintenance techniques and recent bridge collapse incidents have persuaded policy makers to pay greater attention to structural rehabilitation. To prevent any negative socio-economic impact, timely inspection of structures becomes of prime importance. A novel technique which automatizes the inspection procedure and addresses all the limitations of current methods is proposed by the authors. The work investigates whether the use of artificial intelligence (AI), data analytics and extreme scale computing can help engineers automatically detect structural defects in the built environment. As a first step, a typical reinforced concrete bridge is selected as a case study. Bridge inspection is performed by two different approaches; conventional inspection (on site) and virtual reality (VR) inspection (in the office). In the newly proposed technique, VR inspection, a laser scanner is used to capture a 3D digital copy of the bridge, incorporating all its defects. This image is post-processed and imported into a virtual reality application written using Unity, a software development kit for authoring computer games. The resulting VR app is evaluated by conducting a critical comparison between conventional inspection and VR inspection. The results achieved so far demonstrate promising improvements over the conventional inspection technique. The project is currently investigating the use of machine learning to identify structural defects, so that inspection of the built environment can be performed automatically. This research will benefit civil engineers responsible for inspecting structures and policy makers who may wish to update codes of visual inspection. Furthermore, the research aims to lessen the negative impacts on society, the economy and human life that often arise when infrastructure is not properly maintained.

Computing Insight UK 2019
5 – 6 December 2019

Inspecting Bridges using Imaging, Virtual Reality and AI

“Instead of Office Going to the Bridge, Bridges are Coming to the Office.”

MANCHESTER 1824
The University of Manchester

1. Aims and Objectives

Develop digital twins of infrastructure and automate bridge inspection using the concepts of **“Smart Cities.”**

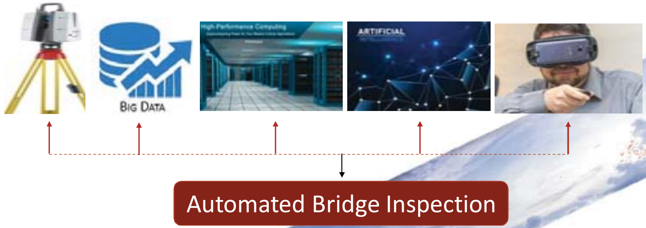


Fig 1. Cutting edge technology used for automated inspection

Use AI and extreme scale computing to detect structural features.
Develop data sub sampling algorithms to reduce the memory foot print.

2. Introduction

Due to rapid urbanization and need of optimised infrastructural maintenance, there is a growing trend of what is called **“Smart Cities.”**

Fig 2. Reasons for the need of infrastructural assessment

Smart City in described in three key words known as 3I’s **“Instrumented, Interconnected and Intelligent City.”**

3. Motivation of Study

Principle Inspection/Visual Inspection is the primary technique for assessing the serviceability and performance of bridge structure. Following are the difficulties faced by engineers in inspecting bridges

Interpretation of results — Accessibility to critical areas — Poor Lightening — Safety of inspectors

Fig 3. Limitations of conventional inspection technique

Fig 4. Digital twins of a bridge using 3D scanner

4. Methodology of Proposed Workflow

The Mancunian Way is chosen the case study. The flowchart below summarizes all the major step required. The bridge and field experimental setup is shown in *figure 6* and *figure 7*. Blue annotates the stations and yellow annotates the targets.

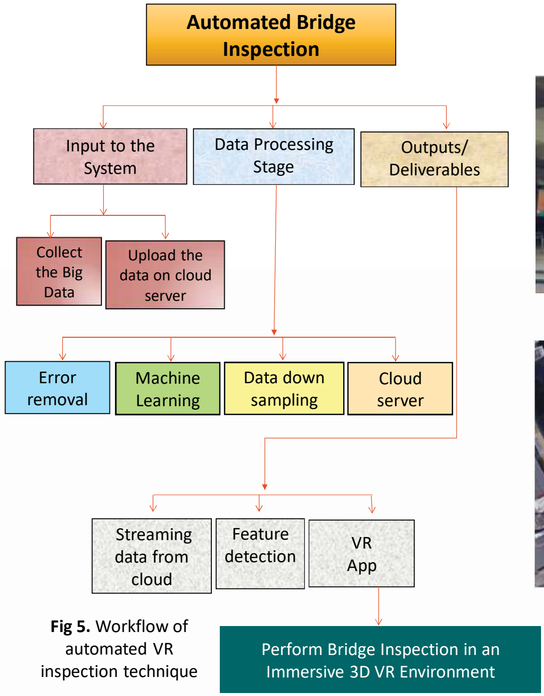


Fig 5. Workflow of automated VR inspection technique

Fig 6. Photograph of the Mancunian Way

Fig 7. Experimental setup

5. Results and Discussion

Results from bridge inspection in VR are demonstrated in *figure 8.*

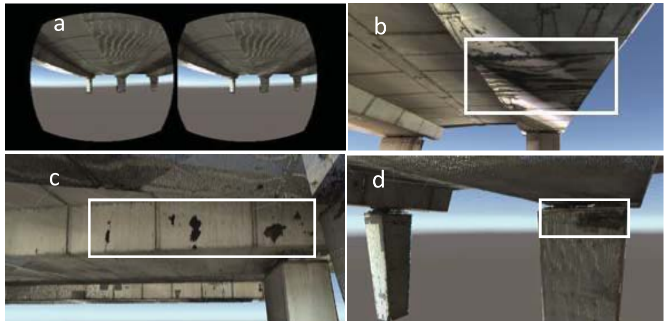


Fig 8. Bridge inspection in the VR. Features shown are: a) stereoscopic view; b) spalling on girders; c) thermal cracks on the beams; d) spalling in piers

Critical comparison between Principle Inspection in VR and conventional method of Inspection is shown in *table 1.*

Comparison Criteria	Conventional Inspection	VR Inspection
Accessibility to critical areas	Difficult	All areas are accessible
Ease of Data collection	Subjective	Effective
Consistency in findings	Low	Consistent
Interpretation of results	Based on experience	Repeatable
Safety of inspector	Needs improvement	Excellent
Time (Disruption)	Depends on the scale of inspection	Less
Documentation	Needs improvement	Excellent
Cost per inspection	Costly	Cheaper in long term

Table 1. Critical comparison of the conventional inspection technique and the VR inspection technique

6. Conclusion and Future Work

The new approach promises to be highly effective in terms of interpretation of results, accessibility to critical areas and safety of inspectors and time consumption.
In the next stage, AI, extreme scale computing and cloud computing will be used to detect and quantify cracks, reduced material strength and other irregularities.
Validation test on different bridges will be performed to verify the accuracy and efficiency of the framework.

Authors of the poster:
Omer. M, Margetts. L, Mosleh. M, Cunningham. L.

For further details, visit the article or scan the QR Code
Omer *et al* (2019), *Use of Gaming Technology to Bring Bridge Inspection to the Office*, Journal of Structure and Infrastructural Engineering

Presented by
Muhammad Omer (PhD Civil Engineering)
eng.muhammad.omer@gmail.com

William Saunders

University of Bath

Fast electrostatic solvers for kinetic Monte Carlo simulations

Kinetic Monte Carlo (KMC) is an important computational tool in physics and chemistry. The method permits the description of time dependent dynamical processes which are not in equilibrium. Recently KMC has been applied successfully to model energy materials such as Lithium-ion batteries and organic solar cells. We consider KMC where particles are localised to specific sites in a material and interact via electrostatics. The frequent calculation of these electrostatic interactions is usually the bottleneck of the simulation. To address this issue, we recently developed a variant of the Fast Multipole Method which dramatically reduces the computational cost of the electrostatic energy computations. Our algorithm scales linearly in the number of charges for each KMC step, something which had not been deemed to be possible before. In this poster we provide an overview of the KMC algorithm and our contribution for computing the electrostatic interactions. Furthermore we present initial performance results, with up to 128M charges, on the ARM Tier 2 facility Isambard and traditional X86 hardware.

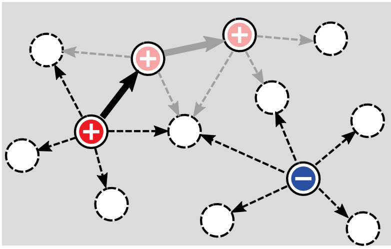
Authors: Will Saunders, Eike Mueller, James Grant and Alison Walker

Fast electrostatic solvers for Kinetic Monte Carlo simulations

Will Saunders (W.R.Saunders@bath.ac.uk)
Eike Müller, James Grant, Alison Walker
University of Bath

Kinetic Monte Carlo (KMC)

- Simulates long timescale behaviour by integrating out short timescale motion [1][2]
- Used in the study of energy materials, e.g. Photovoltaic cells [3][4], Batteries [5] and OLEDs [6][7]
- Long timescales are extremely expensive with other approaches
- The system evolves by "hopping" N charges between sites



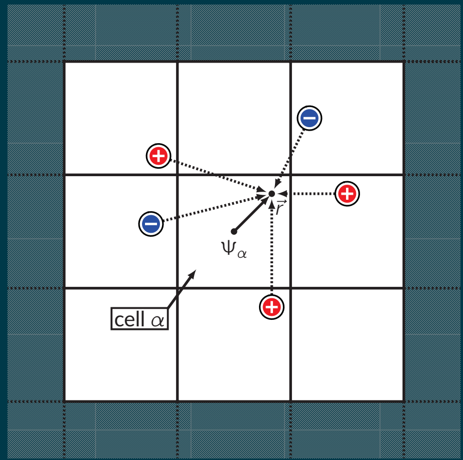
Algorithm

1. Find all possible hops as charges block each other
2. **(Propose)** Compute the relative probability of occurrence for each hop: **This probability is a function of the change in electrostatic energy for the hop. Electrostatic interactions are the computational bottleneck.**
3. **(Accept)** Choose, and perform, **one** hop based on probabilities

Our Fast Multipole Method (FMM) adaptation

- Cost to compute the change in electrostatic energy **per hop** is $\mathcal{O}(1)$
- Cost to accept a hop is $\mathcal{O}(N)$
- Hence cost per KMC step is $\mathcal{O}(N)$ (**algorithmically optimal**)
- Existing methods **do not** achieve this complexity [8]

We present a **computationally optimal** algorithm to compute electrostatic interactions in Kinetic Monte Carlo based on the Fast Multipole Method



Potential field evaluation:

- Potential $\Phi(\vec{r})$ is the sum of a "Direct" part ($\text{Direct}(\vec{r})$) and an "Indirect" part ($\text{Indirect}(\vec{r})$)
- $\text{Direct}(\vec{r})$ Computes the direct interactions with charges in the cell α and its nearest neighbours (dotted lines)
- $\text{Indirect}(\vec{r})$ Evaluates the local expansion Ψ_α
- Ψ_α describes the field from charges outside the cell α and its nearest neighbours

The change in energy ΔU of a charge (magnitude q) hop from $\vec{r}$ to $\hat{\vec{r}}$:

$$\Delta U = q\underbrace{\left[\text{Direct}(\hat{\vec{r}}) + \text{Indirect}(\hat{\vec{r}})\right]}_{\text{potential at new site}} - q\underbrace{\left[\text{Direct}(\vec{r}) + \text{Indirect}(\vec{r})\right]}_{\text{potential at old site}} - \underbrace{\frac{q^2}{|\hat{\vec{r}} - \vec{r}|}}_{\text{self interaction}}$$

The **self interaction** term:

- Accounts for the fact that the data structures contain the charge at the old site $\vec{r}$
- Is extended to **periodic boundary conditions** in our approach

When a move is **accepted**:

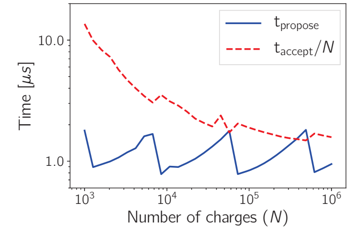
1. Update the local expansions Ψ_α in all cells α
2. Update the charge position for the direct component

arXiv:1905.04065

[1] W. Young, E. Elcock, I, Proceedings of the Physical Society 89 (3) 735, 1966
[2] D. Gillespie, J. Comput. Phys. 22 (4) 403–434, 1976
[3] R. G. E. Kimber, E. N. Wright, S. E. J. O'Kane, A. B. Walker, and J. C. Blakesley, Phys. Rev. B 86, 235206, Dec 2012
[4] T. Albes, P. Lugli and A. Gagliardi, IEEE 39th PVSC, 2013
[5] B. J. Morgan, R Soc Open Sci. 4(11), 170824, Nov 2017
[6] I. R. Thompson, M. K. Coe, A. B. Walker, M. Ricci, O. M. Roscioni, and C. Zannoni, Phys. Rev. Materials 2, 064601, Jun 2018
[7] R. Coehoorn, H. van Eersel, P. Bobbert, R. Janssen, Adv. Funct. Mater 25(13) 2024-2037, 2015
[8] H. Li, J-L. Bredas, Adv. Funct. Mater 28(29) 1801460, 2018

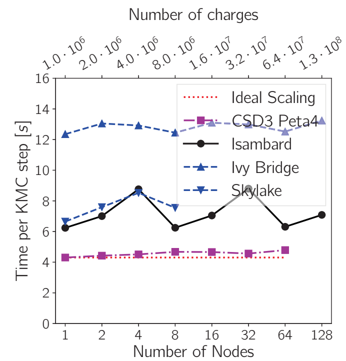
- Funding is acknowledged from the EU Horizon 2020 grants numbers 646176 (EXTMOS) and 824158 (EoCoE-II)
- This research made use of the Balena High Performance Computing (HPC) Service at the University of Bath
- This work used the Isambard UK National Tier-2 HPC Service (http://gw4.ac.uk/isambard/) operated by GW4 and the UK Met Office, and funded by EPSRC (EP/P020224/1)
- This work was performed using resources provided by CSD3 operated by the University of Cambridge Research Computing Service. Provided by Tier-2 funding from EPSRC grant EP/P020259/1, and DiRAC funding from STFC.

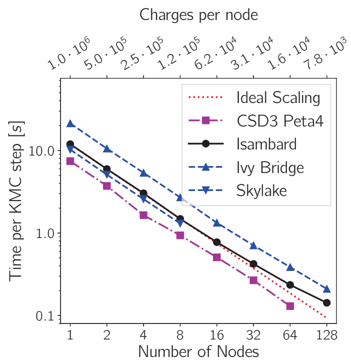
Computational Complexity



Time per proposed move t_{propose} and time for accepting a proposal per particle t_{accept}/N as a function of the number of charges.

Parallel Scaling





CSD3 Peta4:	Intel Xeon Gold 6142 (32 cores/node)
Isambard:	Marvell ThunderX2 (64 cores/node)
Ivy Bridge:	Intel Xeon E5-2650v2 (16 cores/node)
Skylake:	Intel Xeon Gold 6126 (24 cores/node)

Conclusions

- We provide an optimal algorithm for KMC
- Our implementation:
  1. Demonstrates $\mathcal{O}(1)$ & $\mathcal{O}(N)$ complexity
  2. Scales to at least 128 nodes in a strong & weak setting

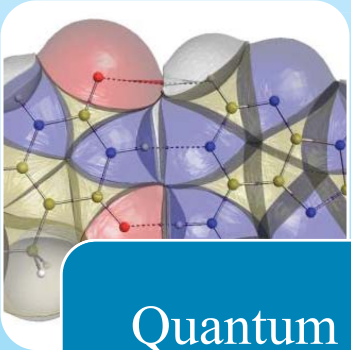
Benjamin C.B Symons

University of Manchester

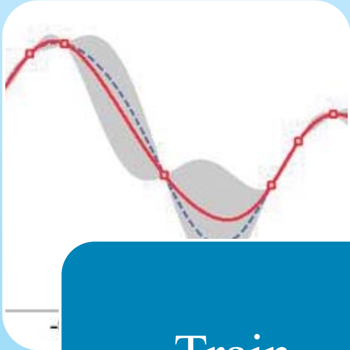
Making quantum chemistry code FFLUX HPC ready

This poster details the journey from an unoptimised, serial version of our machine learning based quantum chemistry code FFLUX to a highly efficient, parallel code suitable for HPC. Currently the code is parallelised with OpenMP and we have achieved speed-ups of 50-60x (running on 16 skylake threads) vs the unoptimised, serial code. This has allowed us to run simulations that were previously unfeasible in reasonable time frames. Implementation of MPI is also in progress which we hope will allow us to attain even greater speed-up.

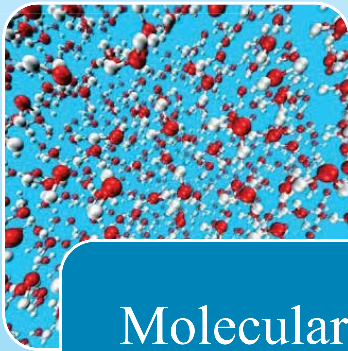
Towards a HPC ready FFLUX



Quantum Chemical Topology calculations.



Train Kriging models.



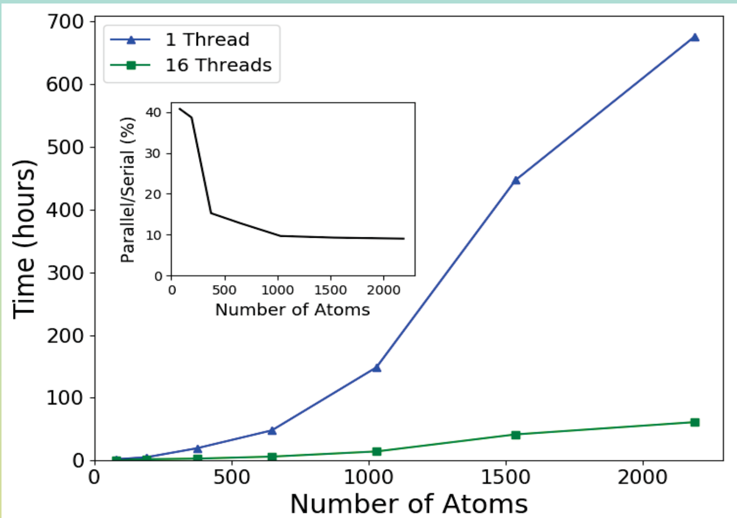
Molecular dynamics with FFLUX forcefield.

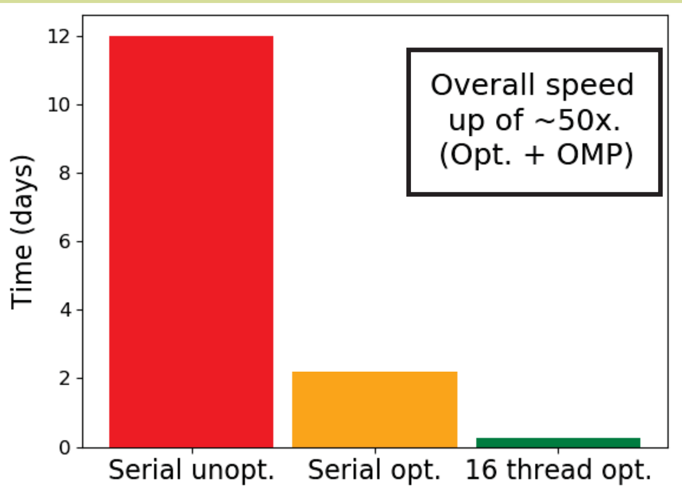
Multipolar → **Flexible** → **Polarisable**

FFLUX forcefield

Water box simulation scaling with system size for serial and parallel codes. 16 OpenMP threads shows a marked improvement over serial.

Electrostatics at L=3.





Simulation of 216 water molecules in a 19Å box for 1ns with a 2fs time step. Point charges only. Note all calculations done on Skylake Intel Xeon gold 6130.

Benjamin Symons


COMPUTING
INSIGHT UK 2020
Science & Technology Facilities Council
UK Research and Innovation
SAVE THE DATE
3-4 December 2020
MANCHESTER CENTRAL, UK
SEE YOU IN 2020!


COMPUTING
INSIGHT UK 2020
Science & Technology Facilities Council
UK Research and Innovation
3-4 December 2020
Manchester Central, UK
The CIUK 2020 Cluster Challenge
More information at
www.stfc.ac.uk/ciuk