Science and Technology Facilities Council

Conference Proceedings
STFC-CONF-2023-001

Computing Insight UK 2022

Manchester Central Convention Centre, UK
1st-2nd December 2022

G Lomas, D Jones (editors)

March 2023

Enquiries concerning this report should be addressed to:

RAL Library
STFC Rutherford Appleton Laboratory
Harwell Oxford
Didcot
OX11 0QX

Tel: +44(0)1235 445384
email: libraryral@stfc.ac.uk

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

Accessibility: a Microsoft Word version of this document (for use with assistive technology) may be available on request.

DOI: 10.5286/stfcconf.2023001

ISSN 2753-5800

Computing Insight UK (CIUK) 2022 took place on 1st and 2nd of December 2022 at the Manchester Central Convention Centre. 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.

GOLD Sponsors	SILVER Sponsors	BRONZE Sponsors
VAST	AMD	OCF
DATACORE	BIOS IT	aws
BOSTON Servers I Storage I Solutions	VANTAGE DATA CENTERS	IBM
SCAN Business		KAO DATA
		PURESTORAGE Uncomplicate Data Storage, Forever
		ddn
		CORNELIS NETWORKS
		Atos
		DELL Technologies
		intel
		WEKA
		UCL CENTRE FOR ADVANCED RESEARCH COMPUTING
		ATEMPO

Computing Insight UK 2022 Introduction

Computing Insight UK (CIUK) 2022 was the 33rd edition of an annual conference organised by the Science and Technology Facilities Council's (STFC) Scientific Computing Department (SCD). The event was held on the 1-2 December at Manchester Central and attracted a record crowd of over five hundred attendees.

The theme for the conference this year was "Sustainable HPC" with sub-themes including "Sustainable Computer and Data Centres", "Sustainability and Systems Administration", "Software Engineering to Improve Code Performance" and "Industry Perspective on Sustainability".

CIUK 2022 included an exhibition of the latest hardware and software releases plus a full, two day programme of presentations and a series of parallel breakout sessions. There was a poster competition, won by Jakub Adamski (University of Edinburgh) for his poster on "Energy Efficient Quantum Computing Simulations", and we also presented our annual Jacky Pallas Memorial Award, which this year was awarded to Dr Djenifer B. Kappel (Centre for Neuropsychiatric Genetics and Genomics, Cardiff University) for her work on "The genomic basis for precision medicine in treatment-resistant schizophrenia". Djenifer presented her work as part of the main programme during the conference.

CIUK 2022 also saw the third edition of the CIUK Cluster Challenge competition with six teams entering from UCL, Birmingham, Bristol (two teams), Durham and York. The teams completed three online challenges leading up to the conference, followed by four challenges during the conference in Manchester. Team ClusDur from Durham University took the title after a closely fought competition and earned their place at the ISC'23 Cluster Challenge competition where they will represent CIUK against the best student teams from around the word.

Computing Insight UK 2022 *"Sustainable HPC"*

Thursday 1 and Friday 2 December 2022

GOLD Sponsors	SILVER Sponsors	BRONZE Sponsors
VAST	AMD	OCF aws IBM
DATACORE	BIOS IT	KAO DATA
BOSTON Servers I Storage I Solutions	VANTAGE DATA CENTERS	PURESTORAGE Uncomplicate Data Storage, Forever ddn
SCAN Business		CORNELIS NETWORKS Atos
		DELL Technologies
		intel WEKA
		UCL CENTRE FOR ADVANCED RESEARCH COMPUTING ATEMPO

Main Programme Session Themes

	Session 1: Sustainable Computer and Data Centres
	Session 2: Sustainability and Systems Administration
	Session 3: ExCALIBUR in Practice

DAY 1 - Thursday 1 December 2022

TIME	MAIN PROGRAMME	BREAKOUT SESSIONS
From 08:30	**REGISTRATION OPEN (Charter Foyer) EXHIBITION OPEN (Gallery)**	
09:15 - 09:30	*Welcome and Introduction* **Tom Griffin (Director, Scientific Computing, STFC)**	
09:30 - 10:00	**Dr Peter Oliver (Scientific Computing, STFC)** *Design considerations for an environmentally sustainable datacentre for STFC*	
10:00 - 10:30	**Pekka Lehtovuori (CSC - IT Center for Science Ltd)** *Reaching zero carbon footprint in HPC operations*	CoSeC Computational Science Centre for Research Communities **Annual Conference 2022** Thursday 1 December @ CIUK 2022 10:00-16:00
10:30 - 11:00	**Thomas Eickermann (Jülich Supercomputing Centre)** *Towards more sustainable HPC at the Jülich Supercomputing Centre*	
11:00 - 11:30	**REFRESHMENTS**	
11:30 - 12:00	**Jacob Newman (University of East Anglia)** *Optimising HPC Workflows: Three Case Studies from a Research Software Engineer's Perspective*	
12:00 - 12:30	**Gabryel Mason-Williams (Rosalind Franklin Institute)** *DisTRaC: Accelerating High-Performance Data Processing*	
12:30 - 13:00	**Simon Atack (University of Bristol)** *Creating A Cluster - Going it Alone*	
13:00 - 14:00	**LUNCH**	
14:00 - 14:30	**Ed Threlfall (UKAEA)** *Project NEPTUNE - sustainable software for sustainable fusion energy*	
14:30 - 15:00	**Joseph Hickson, Lewis Sampson and Victoria Smart (Met Office)** *Preparing the Met Office for the next generation of supercomputers*	
15:00 - 15:30	**Ben Rogers (University of Manchester) and Phil Hasnip (University of York)** *PAX-HPC - Modelling particles at exascale: from atoms to galaxies*	
15:30 - 16:15	**REFRESHMENTS**	
16:15 - 17:00	**Martyn Guest (ARCCA, Cardiff University)** *Performance of Community Codes on Multi-core Processors.* *An Analysis of Computational Chemistry and Ocean Modelling Applications*	
17:00 - 18:00	**Keynote Presentation - Professor Michèle Weiland (EPCC, The University of Edinburgh)** **Net Zero HPC - noble dream or inevitable goal?**	
18:30 - 23:00	**CIUK 2022 Networking Event** **Revolucion de Cuba, 11 Peter St, Manchester M2 5QR (*CIUK lanyard and badge required for entry*).**	

Computing Insight UK 2022 *"Sustainable HPC"*

Thursday 1 and Friday 2 December 2022

Main Programme Session Themes

	Session 4: UKRI Net Zero Digital Research Infrastructure Project
	Session 5: Software Engineering to Improve Code Performance
	Session 6: An Industry Perspective on Sustainability

GOLD Sponsors	SILVER Sponsors	BRONZE Sponsors
VAST	AMD	OCF, aws, IBM
DATACORE	BIOS IT	KAO DATA, ddn
BOSTON Servers I Storage I Solutions	VANTAGE DATA CENTERS	PURESTORAGE, CORNELIS NETWORKS, Atos
SCAN Business		DELL Technologies, intel, WEKA
		UCL CENTRE FOR ADVANCED RESEARCH COMPUTING, ATEMPO

DAY 2 - Friday 2 December 2022

TIME	MAIN PROGRAMME	BREAKOUT SESSIONS
From 08:30	**REGISTRATION OPEN (Charter Foyer) EXHIBITION OPEN (Gallery)**	WHPC WOMEN IN HIGH PERFORMANCE COMPUTING 08:30-10:30
09:30 - 10:00	**UKRI Net Zero Digital Research Infrastructure Project** 09:30 - 09:50 Project Overview **(Martin Juckes)** 09:50 - 10:00 HPC-JEEP (**Alastair Basden and Andy Turner**) 10:00 - 10:10 IRISCAST (**Jonathan Hays**) 10:10 - 10:20 ENERGETIC (**Deepan Bhowmik and Teymoor Ali**) 10:20 - 10:30 CARBON-QUANDRI **(Daniel Schien)** 10:30 - 11:00 Panel Discussion **(Wim Vanderbauwhede, Justin O'Byrne, Martin Juckes)**	
10:00 - 10:30		
10:30 - 11:00		Spectrum Scale User Group 10:30-12:30
11:00 - 11:30	**REFRESHMENTS**	
11:30 - 12:00	**Ilektra Christidi (Senior Research Software Developer, UCL Advanced Research Computing Centre)** *Coupling the Time-Warp algorithm with a Kinetic Monte Carlo framework for exact distributed simulations of heterogeneous catalysts*	
12:00 - 12:30	**Elizabetta Boella (Lancaster University & Cockcroft Institute)** *ECsim: a massively parallel Particle-In-Cell code for plasma physics with OpenACC support*	
12:30 - 13:00	**The Jacky Pallas Memorial Presentation** **Dr Djenifer Kappel (Centre for Neuropsychiatric Genetics and Genomics - Cardiff University)** *The genomic basis for precision medicine in treatment-resistant schizophrenia*	
13:00 - 14:15	**LUNCH**	
14:15 - 14:30	**Award Presentation - The CIUK 2022 Student Cluster Challenge and Poster Competition**	UK Research and Innovation **ENERGETIC: A workshop regarding Energy Benchmarking on Heterogeneous Systems** 14:00-16:00
14:30 - 15:00	**Dr Rosemary Francis (Chief Scientist HPC, Altair)** *Ten Ways in Which Altair is Saving the Planet with HPC*	
15:00 - 15:30	**Dr Crispin Keable (Senior HPC Architect, Global HPC Strategic Sales, Atos)** *Sustainability issues as we move towards exascale class HPC architectures*	
15:30 - 16:00	**Laura Foster (techUK)** *Why is HPC integral to becoming a "science and technology superpower?"*	
16:00	**CIUK 2022 CLOSES**	

Welcome to CIUK 2022

"Sustainable HPC"

Science and Technology Facilities Council

Scientific Computing

- Two Days of Presentations
- Exhibition of the Latest Technology
- Parallel Breakout Sessions
- Poster Competition
- CIUK Cluster Challenge
- Jacky Pallas Memorial Award
- Research Zone
- CIUK 2022 Evening Networking Event

CIUK 2022

Welcome to CIUK 2022

Keynote Presentation

Thursday 1 December
17:00 – 18:00

Michèle Weiland
EPCC
The University of Edinburgh

Net Zero HPC – noble dream or inevitable goal?

UKRI Science and Technology Facilities Council
Scientific Computing

CIUK 2022

Welcome to CIUK 2022

Jacky Pallas Memorial Award

Science and Technology Facilities Council

Scientific Computing

Friday 2 December

12:30 – 13:00

Dr Djenifer Kappel

Centre for Neuropsychiatric Genetics and Genomics - Cardiff University

The genomic basis for precision medicine in treatment-resistant schizophrenia

CIUK 2022

Welcome to CIUK 2022

Science and Technology Facilities Council
Scientific Computing

Want to ask a question at the end of a presentation? We are using slido…

CIUK 2022

Welcome to CIUK 2022

Science and Technology Facilities Council
Scientific Computing

CoSeC Annual Conference

CoSeC
Computational Science Centre for Research Communities
Annual Conference 2022
Thursday 1 December @ CIUK 2022

Thursday 1 December
10:00 – 16:00

Women in HPC Breakfast

Friday 2 December
08:30 – 10:30

WHPC
WOMEN in HPC

Spectrum Scale User Group

Friday 2 December
10:30 – 12:30

Spectrum Scale User Group

ENERGETIC: A workshop regarding Energy Benchmarking on Heterogeneous Systems

Friday 2 December
14:00 – 16:00

UK Research and Innovation

CIUK 2022

Welcome to CIUK 2022

UKRI Science and Technology Facilities Council

Scientific Computing

archer2 · MMM Materials and Molecular Modelling Hub · DiRAC · cirrus Powered by epcc · JADE Tier 2 HPC

Bristol · ExCALIBUR · N8 CIR Computationally Intensive Research BEDE · GW4 Isambard · CSD3 Cambridge Service for Data Driven Discovery · hpc MIDLANDS+

CIUK 2022

Welcome to CIUK 2022

Science and Technology Facilities Council
Scientific Computing

Join the conversation on Twitter…

@CompInsightUK
#CIUK2022

CIUK 2022

CIUK 2022 Presentations

Dr Peter Oliver (Scientific Computing, STFC)

Design considerations for an environmentally sustainable datacentre for STFC

Abstract: The process of turning data into knowledge lies at the heart of research and innovation. Today, the use of digital technologies is as fundamental to modern research as theory, observation and experiment. A new environmentally sustainable datacentre is required to house computing to support the UKRI Digital Research Infrastructure (DRI) requirements of national facilities, science programmes and instruments. The presentation will explore the implications of an environmentally sustainable datacentre in both build and operation.

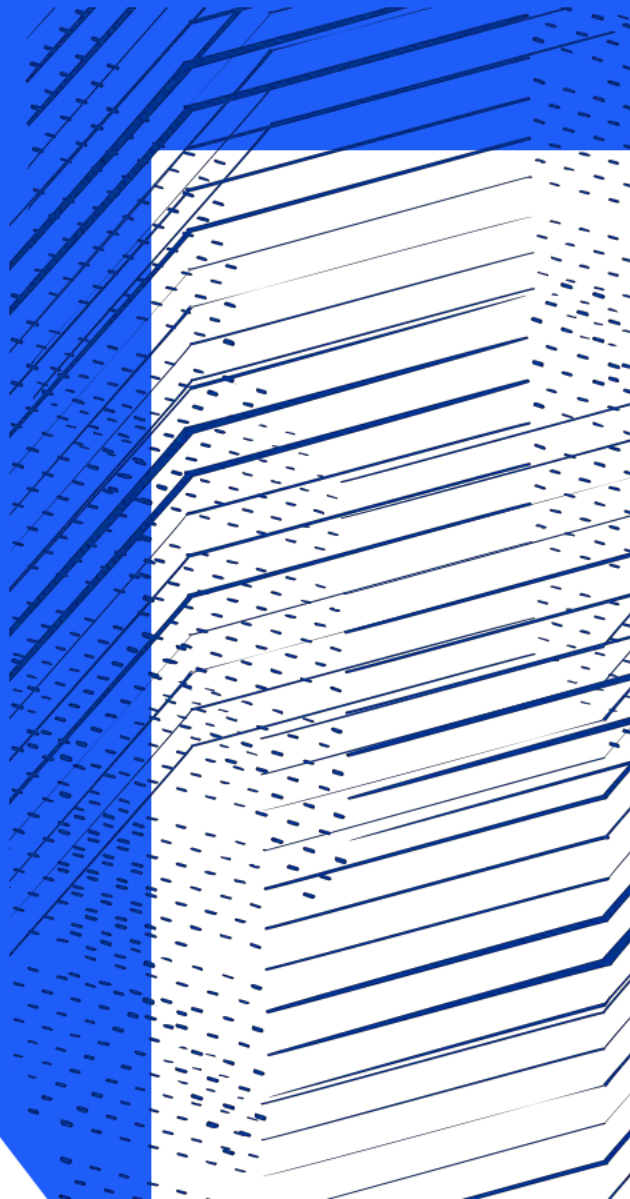
Bio: Dr Peter Oliver is Head of Operations for Scientific Computing (SC) at the Scientific and Technology Facilities Council (STFC) and is leading a new initiative to build an environmentally sustainable datacentre with a very high target for power efficiency. Peter gained his PhD from the University of Bath in 1994 with his thesis entitled “Computer Simulation of the Effects of Temperature on Oxide Surfaces.” Inspired by high performance computing (HPC), Peter joined the Rutherford Appleton laboratory in 1997, to provide user support on the national HPC service Cray J90. Since then, Peter has developed his expertise and was responsible for the specification, procurement and management of High Performance Computing Systems such as STFC’s Facilities (SCARF), national e-infrastructures (NGS & NeS) and climate and earth-system science (JASMIN). In 2013, Peter changed roles to lead the Scientific Computing Technology Division with expertise in visualisation, code optimisation, software engineering and computational mathematics and in 2016 became Head of Operations for Scientific Computing.

Science and Technology Facilities Council

Scientific Computing

Design considerations for an environmentally sustainable Datacentre

Dr Peter Oliver, Scientific Computing, STFC
peter.oliver@stfc.ac.uk



Agenda

1 Strategic Case

2 STFC Environmental Strategy

3 Build and Operation

4 Summary

Image © STFC Alan Ford

Science and Technology Facilities Council

Scientific Computing

Agenda

1 Strategic Case

2 STFC Environmental Strategy

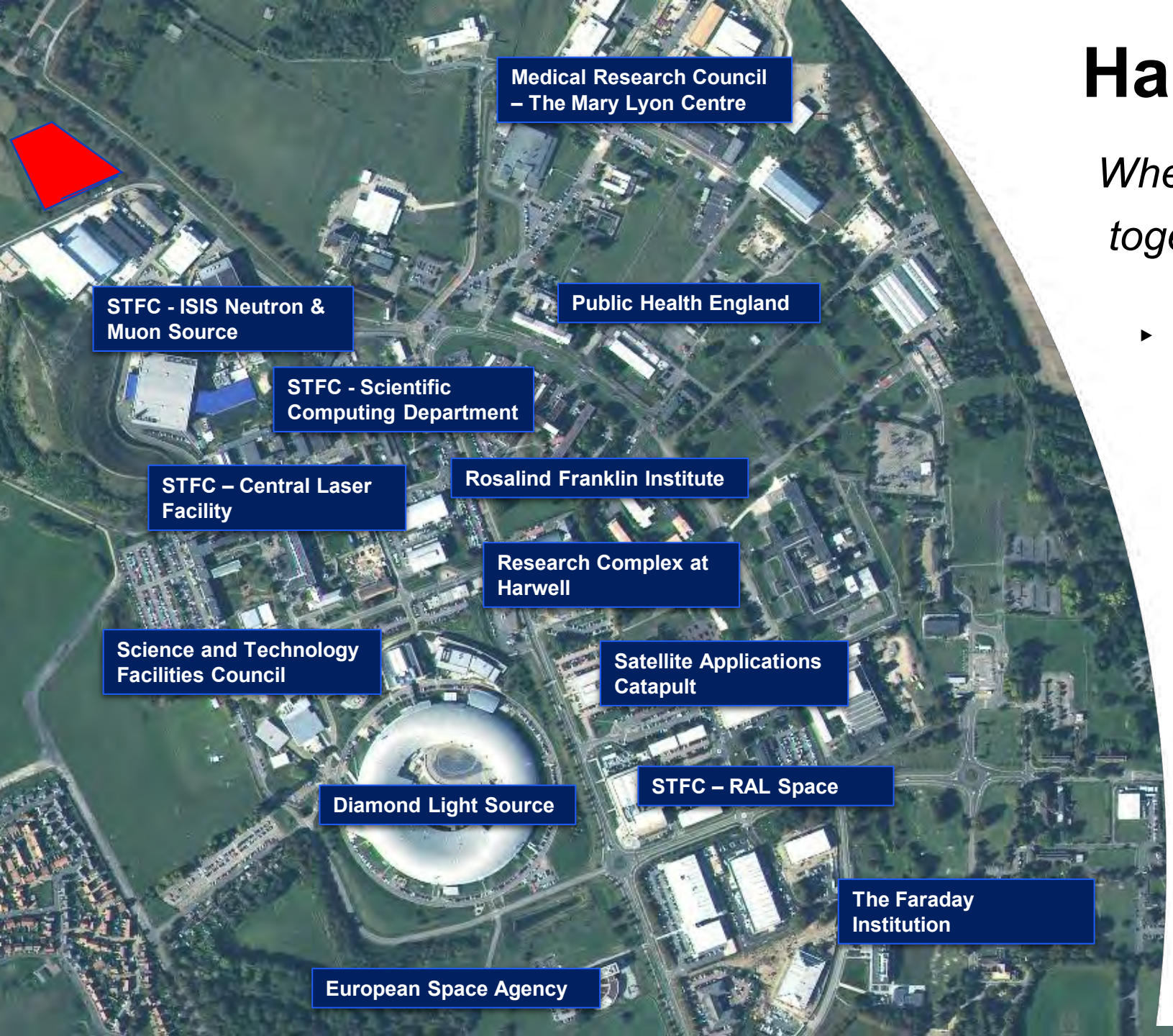
3 Build and Operation

4 Summary

Image © STFC Alan Ford

Science and Technology Facilities Council

Scientific Computing



Harwell Campus

Where science and industry comes together to tackle global problems

- £2+bn national infrastructure that can study viruses to galaxies
- 6,000 people designing and operating facilities
- Supporting research at 70+ UK universities
- Working with scientists in 60+ countries

UKRI HARWELL

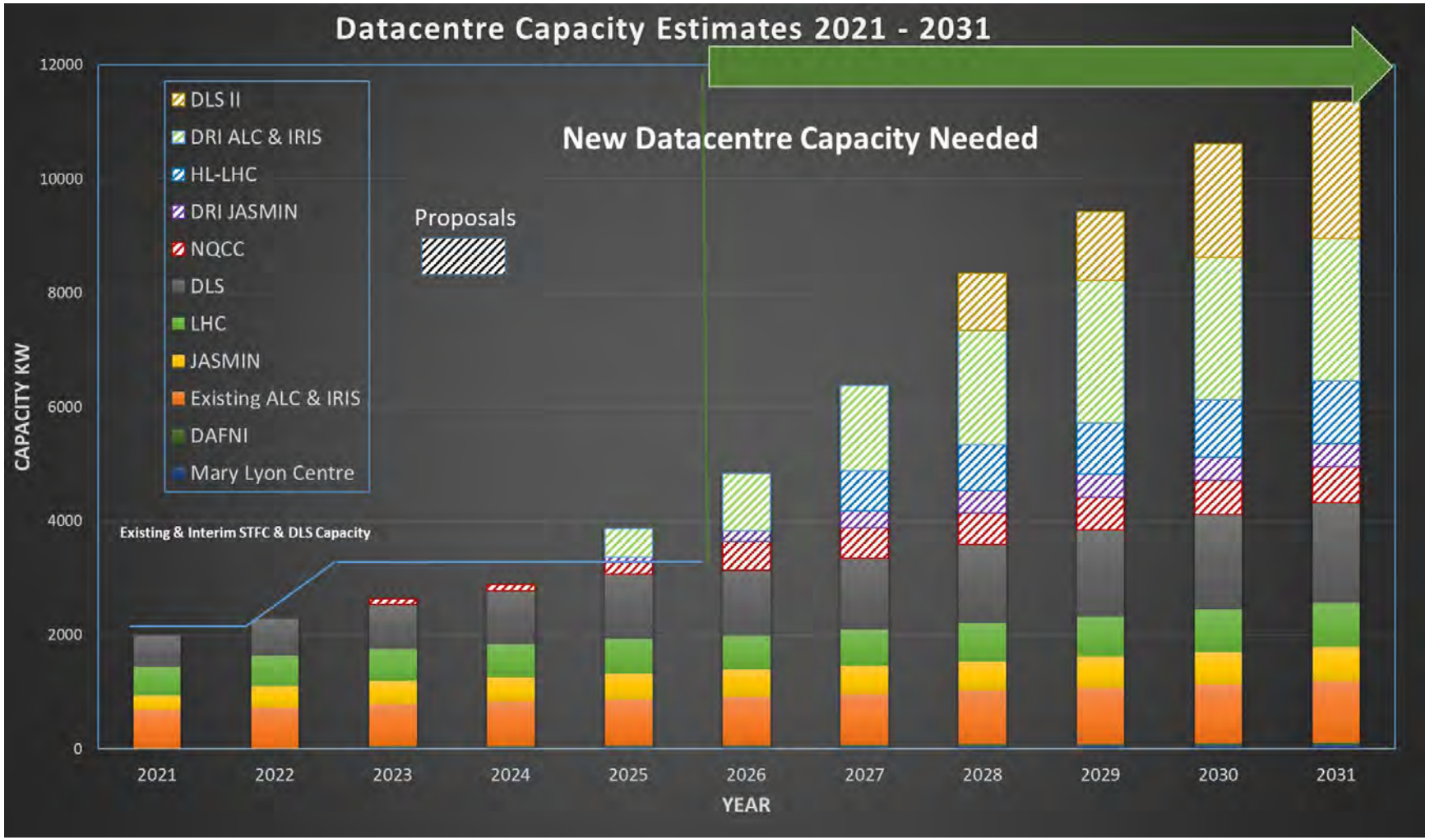
Research Computing Centre (RCC): Requirements Capture

- Major cross-council computational and storage requirements in future years
- Data rates doubling every 6-12 months.

Activity	Purpose	Interested RC
Ada Lovelace Centre (ALC)	Provides software, data services and skills to exploit data from large scale national facilities including Diamond, ISIS and CLF/EPAC	STFC/EPSRC/ BBSRC/MRC
Extreme Photonics Applications Centre (EPAC)	It will bring together world-leading interdisciplinary expertise to develop and apply novel, laser based, non-conventional accelerators and particle sources which have unique properties	STFC
IRIS	A co-operative consortium of STFC science users and compute providers, to deliver High Throughput Computing (CPU and GPU), data management and storage capability to achieve the science goals of the National Facilities (including ISIS, Diamond, CLF), of PPAN projects and instruments (including SKA, LSST, LIGO,LHC, DUNE), and CCFE	STFC
National Quantum Computing Centre (NQCC)	100+ qubit NISQ demonstrator hardware platform(s), Quantum software, algorithm & applications development High performance, scalable qubit technology development, Roadmap and architecture towards fault-tolerant general purpose quantum computing	EPSRC
Rosalind Franklin Institute (RFI)	A new national research centre at the interface between the engineering, physical, and life sciences	EPSRC
Diamond Light Source (DLS)	Diamond Light Source is the UK's national synchrotron science facility	STFC Wellcome
Diamond Light Source II (DLS II)	Diamond-II upgrade, a co-ordinated programme of development that combines a major machine upgrade with new instruments and complementary improvements to optics, detectors, sample environment and delivery capabilities, and computing, as well as integrated and correlative methods	STFC Wellcome
Electron Bio-Imaging Centre (eBIC)	eBIC provides scientists with state-of-the-art experimental equipment and expertise in the field of cryo-electron microscopy	STFC/BBSRC MRC/Wellcome
JASMIN	JASMIN is a globally-unique data analysis facility. It provides storage and compute facilities to enable data-intensive environmental science.	NERC
CCP4	Collaborative Computational Project for Macromolecular X-Ray Crystallography	BBSRC
CCP-EM	The Collaborative Computational Project for electron cryo-microscopy (CCP-EM) supports users and developers in biological EM.	MRC
Mary Lyon Centre	A national facility providing world-class expertise, tools and resources to generate and characterise genetically altered mouse models for use	MRC

Science Drivers

Major new cross council Digital Research Infrastructure (DRI) required to fully exploit science infrastructure



RCC: Cross Council Scientific Research

- RCC will need to host a variety of equipment such as
  - Large scale data storage
  - High Throughput Computing (HTC)
  - High Performance Data Analytics (HPDA)
  - High Performance Computing (HPC) and
  - Specialist computing hardware, for example Machine Learning (ML)

- This encompasses a broad range of power requirements from <10kW/rack to 100kW+/rack.
- RCC is the datacentre only and NOT not the compute & data resources
  - Computing/Data equipment purchases are driven by major scientific programmes/projects/facilities
- Analysis of requirements indicates the need for 11MW of computing load by 2030 and 20MW by 2035.
  - Phased delivery, with 6.6MW capability by 2026, option to install more capacity
  - Capacity in excess of 6.6MW would follow an additional review of requirements and funding request
  - Meet critical requirement for DLS, STFC, NERC by 2025/6 and DLS II from 2028
- Part of a UKRI eco-system of datacentres
  - Complementing Hartree Centre and national capabilities such as ARCHER 2 and Exascale.

Agenda

1 Strategic Case

2 STFC Environmental Strategy

3 Build and Operation

4 Summary

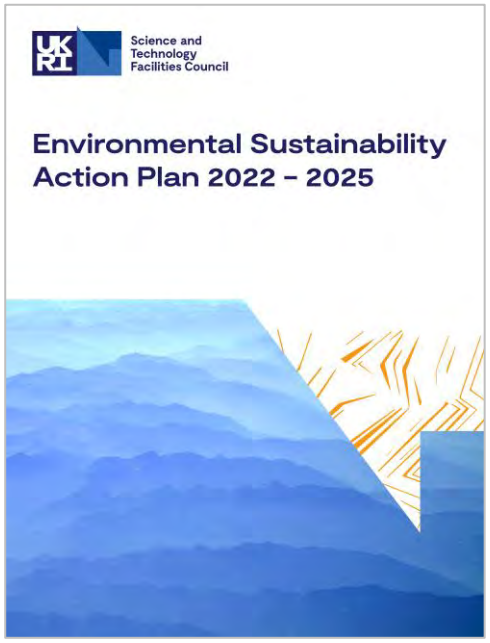
Image © STFC Alan Ford

Science and Technology Facilities Council

Scientific Computing

Environmental Sustainability

- Environmental Sustainability is a critical Strategic aim across UKRI and STFC



Science and Technology Facilities Council

Scientific Computing

Agenda

1 Strategic Case

2 STFC Environmental Strategy

3 Build and Operation

4 Summary

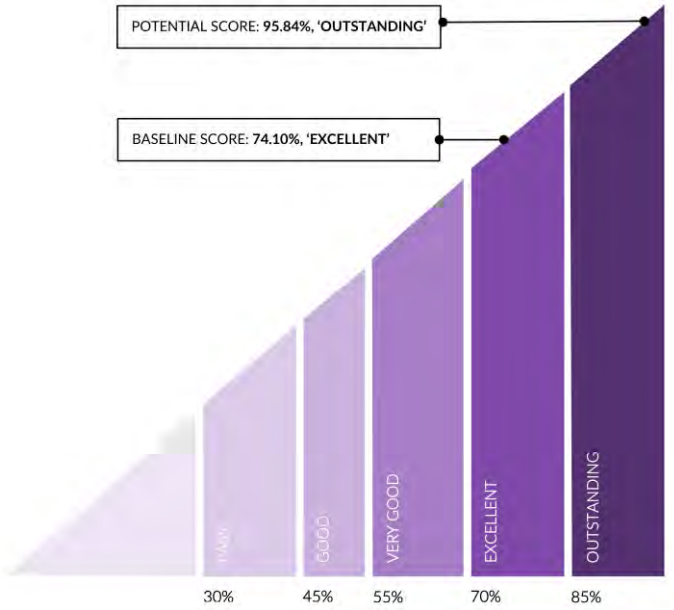
Image © STFC Alan Ford

Science and Technology Facilities Council

Scientific Computing

What does environmental sustainability mean for a datacentre build?

- The RCC is a new datacentre and can adopt international standards for the build.
- BREEAM
  - The world's leading science-based suite of validation and certification systems for a sustainable built environment.
  - External validation of the buildings design
- The current BREEAM assessment is broken down into 10 categories and 5 standards.
  - Energy
  - Health & Wellbeing
  - Innovation
  - Land Use
  - Materials
  - Management
  - Pollution
  - Transport
  - Waste
  - Water



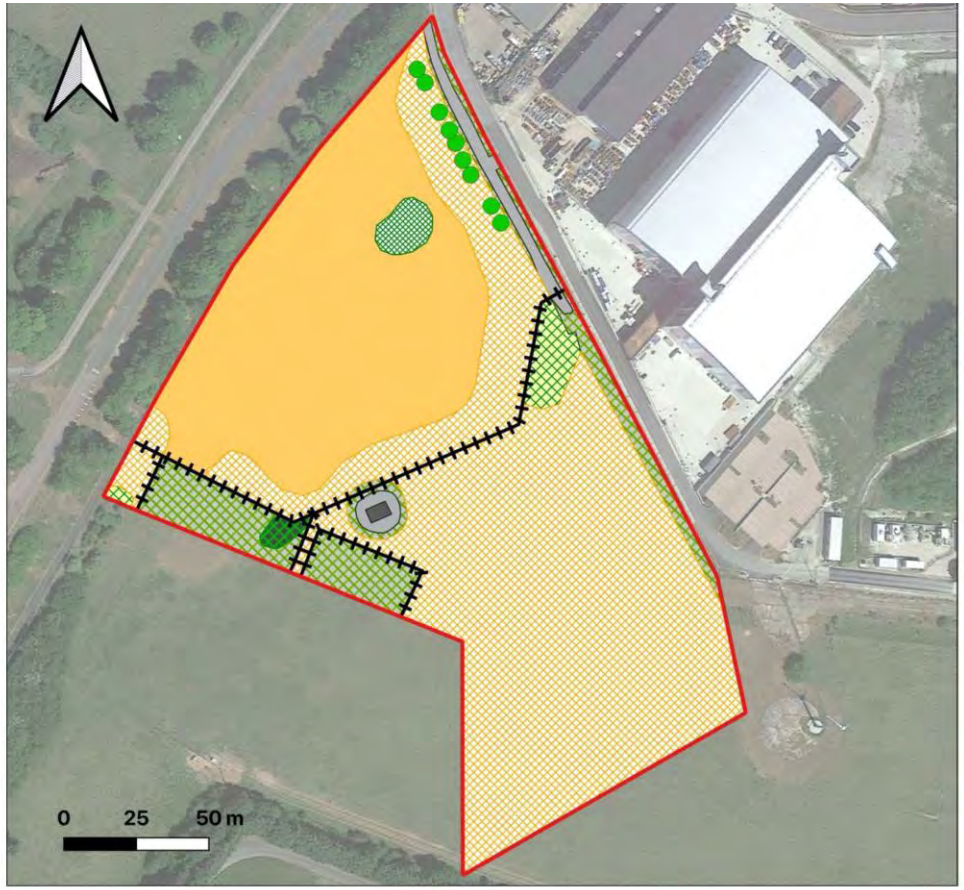
- The RCC is aiming for BREEAM Excellent but has the potential for outstanding

Science and Technology Facilities Council
Scientific Computing

BREEAM® EXCELLENT

BREEAM - Ecology

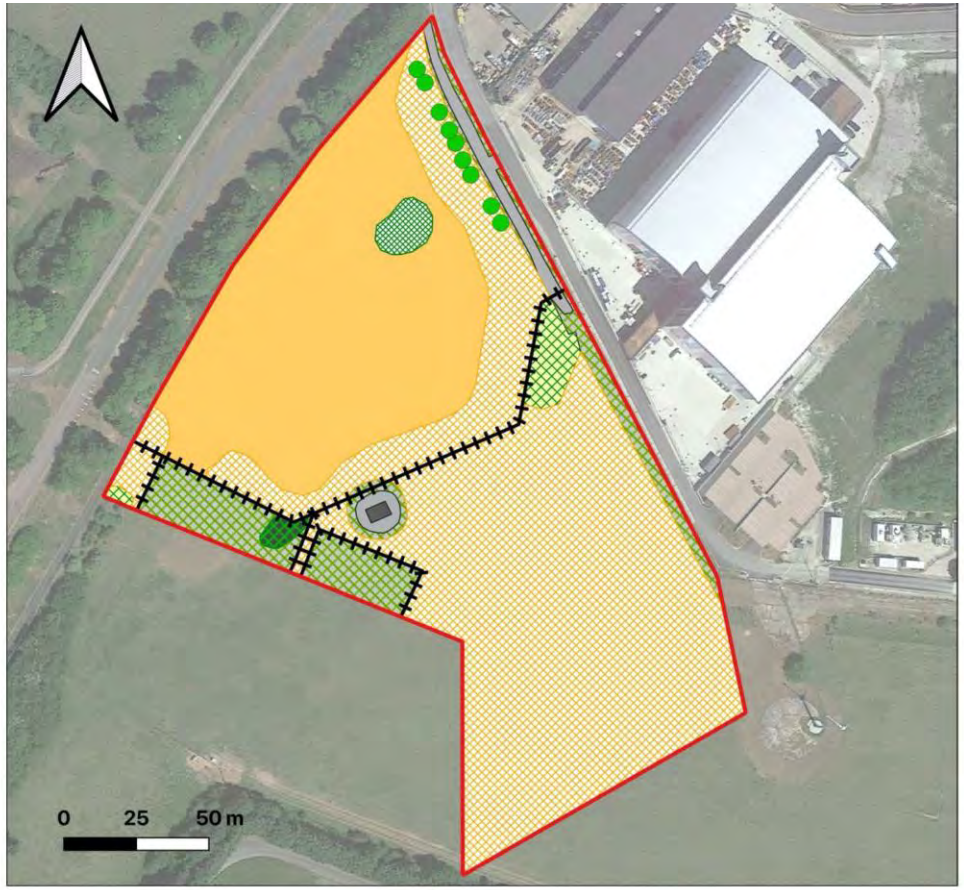
- **Ecology Survey** “A large portion of the site comprises lowland calcareous grassland priority habitat. Any loss of this habitat should be limited where possible, with any unavoidable loss to be compensated for through enhancement of retained areas. “
- RCC will keep existing trees where possible



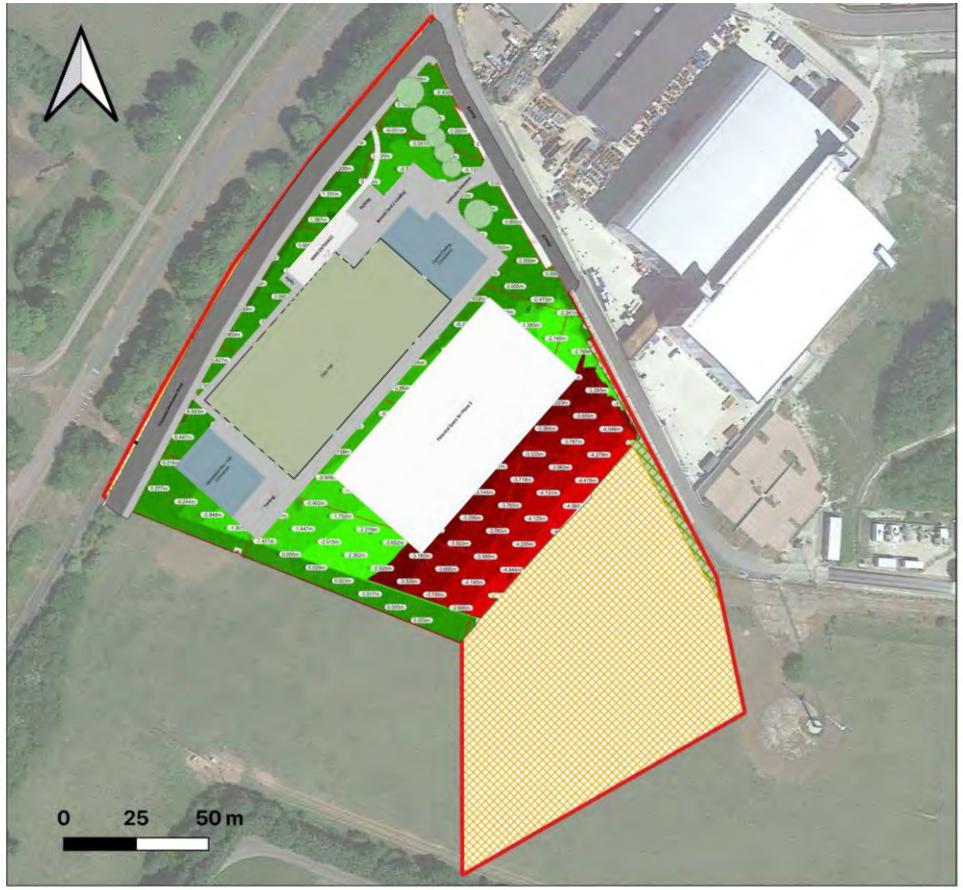
Site boundary
Other neutral grassland
Calcareous grassland
Buildings
Hardstanding
Scattered scrub
Ruderal/ Ephemeral
Blackthorn scrub
Street trees
Fence

BREEAM - Ecology

- **Ecology Survey** “A large portion of the site comprises lowland calcareous grassland priority habitat. Any loss of this habitat should be limited where possible, with any unavoidable loss to be compensated for through enhancement of retained areas. “
- RCC will keep existing trees where possible



Site boundary
Other neutral grassland
Calcareous grassland
Buildings
Hardstanding
Scattered scrub
Ruderal/ Ephemeral
Blackthorn scrub
Street trees
Fence



What does environmental sustainability mean for a datacentre?
In Operation

- The RCC will be able to support 6.6MW of Computing Load.

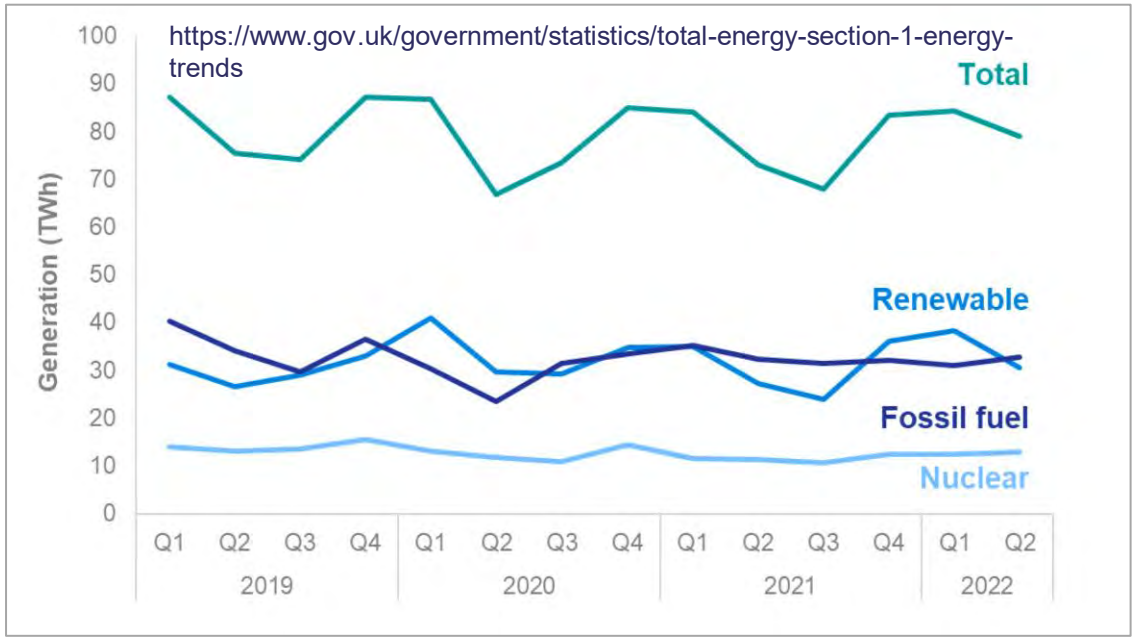
A typical UK house uses 241kWhrs per month

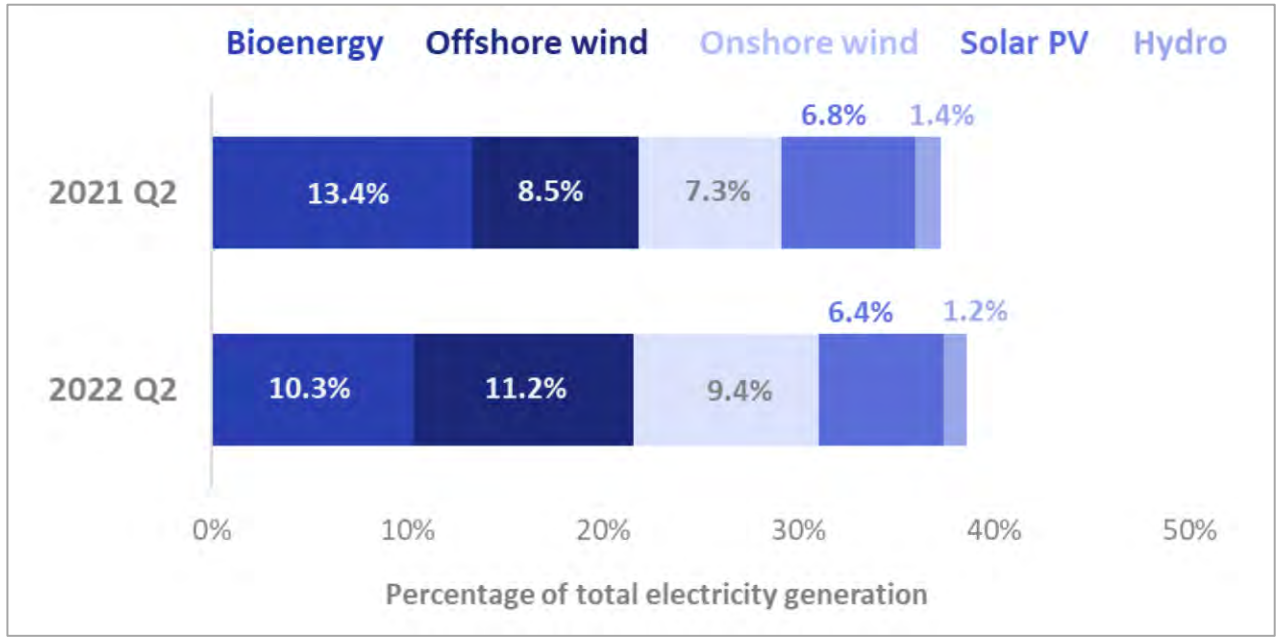
RCC 4,752,000kWhrs per month
~20,000 houses

Science and Technology Facilities Council
Scientific Computing

What does environmental sustainability mean for a datacentre?
In Operation

- Only ~38% of the UK's electricity is generated from renewables





- As we have only 1 electricity grid. The electricity you use comes from all sources even if on a "green tariff". We need to consider CO_2 emissions.

What does environmental sustainability mean for a datacentre?
In Operation

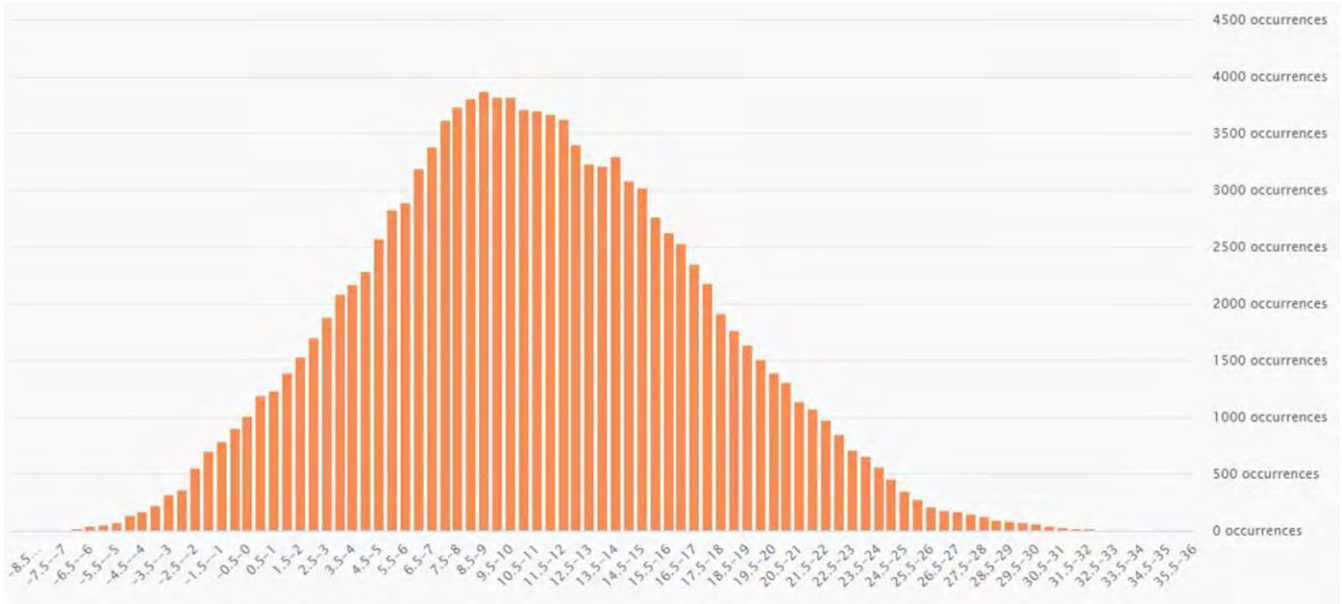
- The RCC needs to be as power efficient as possible.
- **Power usage effectiveness (PUE)** is a ratio that describes how efficiently a computer datacentre uses energy; specifically, how much energy is used by the computing equipment (in contrast to cooling and other overhead that supports the equipment).
- Typical datacentres have a PUE of ~1.5
  - (as does the existing R89 Datacentre at RAL)
- This means that for every 1MW of computing power you use 0.5MW for cooling and other overheads
- Increasing efficiency and reducing PUE clearly has benefits on power usage!
- The RCC is targeting a PUE of 1.1

$$PUE = \frac{Total\ Datacentre\ Power\ Usage}{IT\ Equipment\ Power\ Usage}$$

How to achieve a PUE of 1.1?

Understanding the Weather

Temp °C	# Occurrences
30 – 31	77
31 – 32	35
32 – 33	16
33 – 34	16
34 – 35	9
35 – 36	2



Hourly ambient air dry bulb temperature for Harwell 2008-2021

Climate change will affect these so need to plan for the future for a datacentre with a 20 year life

Science and Technology Facilities Council

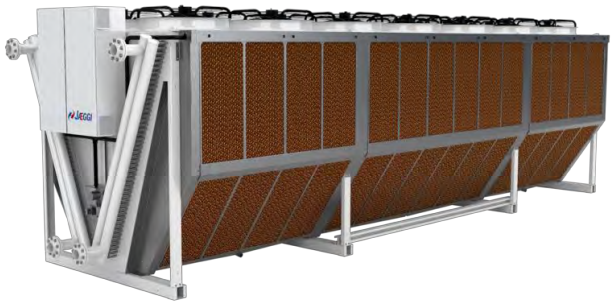
Scientific Computing

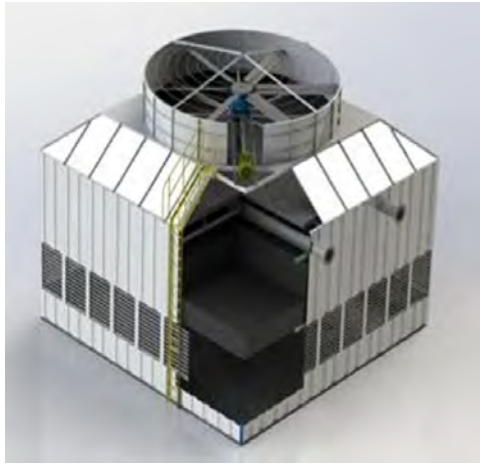
How to achieve a PUE of 1.1?

Understanding the Technology

Rear Door Heat Exchangers 100kW/rack

Type	Summary
Dry Air Coolers	Free Cooling, efficient average PUE, high peak PUE as needs chillers
Adiabatic cooling	Free Cooling, efficient average PUE, uses spray water, maintenance issues
Cooling Towers	Free Cooling efficient PUE, Uses LOTS of water, high maintenance
Hybrid Coolers	Free Cooling, efficient average and peak PUE, contained water, acceptable water usage as anticipated need low.





Images for illustrative purposes only

PUE, CO_2 and Power Costs (2028)

Power costs 37.9p/kWh (based on Crown Commercial Services estimates)

$kgCO_2e$/kWh 0.167 (UK Gov Green book)

Computing Load 6.6MW

PUE	Total MW	£m Electricity	tCO2e 1000's	# Houses 1000's
1	6.6	21.9	10	20
1.1	7.26	24.1	11	22
1.2	7.92	26.3	12	24
1.3	8.58	28.5	13	26
1.4	9.24	30.7	14	28
1.5	9.9	32.9	15	30
Diff 1.5-1.1	**2.6**	**8.8**	**4**	**8**

Environmental Sustainability is a double win, less CO_2 and smaller Electricity Bill!

Additional activities to increase efficiency of the RCC

There will be a large area for Solar power generation building on STFC investment solar panels across the RAL and Daresbury. The existing investment is estimated to deliver 3,450MWh of energy each year at RAL and 1,200MWh at Daresbury annually.

There are great examples of where the waste heat from datacentres has been used for domestic district heating, agriculture and other processes. STFC is exploring if the waste heat can be used for new campus buildings.

An Efficient Datacentre is not the end of the story.

The previous calculations have assumed all the racks and servers are performing “useful work”

However…..

- Need to optimise usage as an idle resource is not doing useful work
- Optimise applications
- That 10+ year old code could be re-engineered and perform more efficiently.
- Newer libraries
- Are you using the application in the best way?
- New approaches such as ML could drastically cut down time to solution (but training costs could be high)

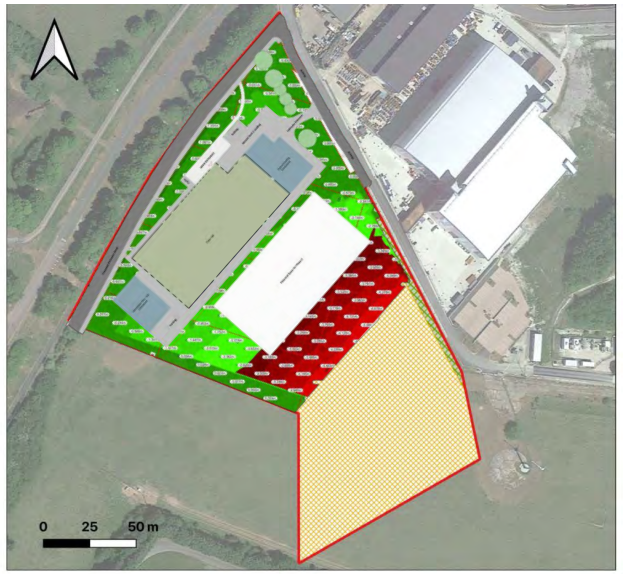
Science and Technology Facilities Council

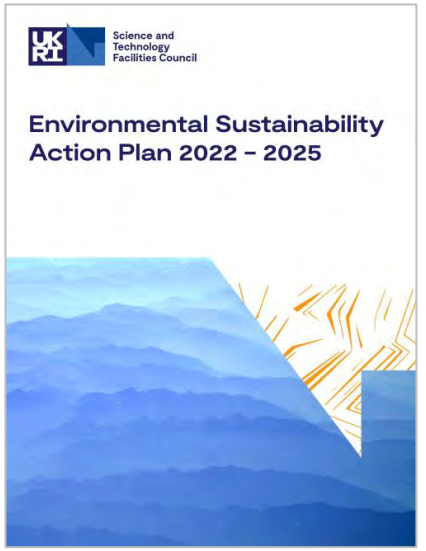
Scientific Computing

Summary

- Build and Operation matter
- Using standards such as BREEAM for the build minimises the environmental impact of the build
- Landscaping to minimise impact and enhance ecology and bio-diversity
- Cooling technology choices matter
- At a PUE of 1.1 can save up to £8.8m electricity compared to PUE 1.5!
- A low PUE **decreases tCO2e** and **decreases the electricity** costs

BREEAM®
EXCELLENT





Science and Technology Facilities Council
Scientific Computing

Scientific Computing

Questions?

Science and Technology Facilities Council

Scientific Computing

Thank you

scd.stfc.ac.uk

@SciComp_STFC

Pekka Lehtovuori (CSC - IT Center for Science Ltd)

Reaching zero carbon footprint in HPC operations

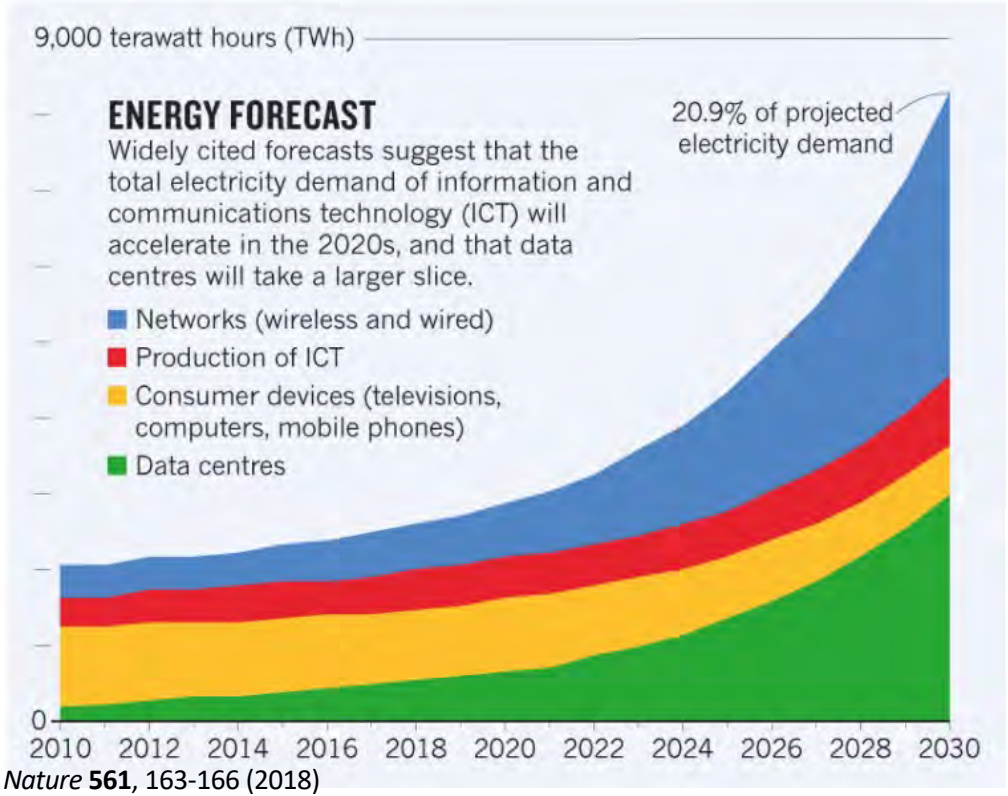
Abstract: Operating HPC data centers is very energy intensive business. In this presentation we explain how LUMI datacenter was designed to be one of the world most energy efficient data centers and how we achieved carbon negative foot print for our HPC operations.

Bio: Director Pekka Lehtovuori (PhD) leads the Services for Computational Research unit at CSC. He is responsible for CSC's national HPC services, scientific and user support as well as cloud and data intensive computing services. He is a member of the board for Nordic e-Infrastructure Collaboration. Dr Lehtovuori has extensive experience in development, implementation, and operation of national and European research infrastructure projects, such as FCCI, PRACE, EGI, ELIXIR, NeIC, EOSC, and EuroHPC.

Reaching Zero-Carbon Footprint in HPC Operations

Dr. Pekka Lehtovuori
Director, Services for Computational research
CSC – IT Center for Science, Finland

Data centers & CO_2 emissions



Nature **561**, 163-166 (2018)

- Data center operations already use more than 2% of the world's electricity, and contribute to 2% of world's CO_2 emissions.
  - Equivalent to the world's entire airline industry
- European Green Deal's goal is to make data centers climate neutral by 2030.

Considerations for a HPC system's carbon footprint

LUMI

- Data center and operations level choices
  - Power: used electricity, power-usage efficiency, PUE
  - Waste heat reuse, ERF
    - District heating, sorption cooling, water preheating, desalination, biomass processing, greenhouses,…
  - Construction/retrofitting of the data center
- System level choices
  - ICT manufacturing
  - Eco-efficiency of the hardware and software ("science per watt")

Case study: EuroHPC LUMI, Kajaani, Finland

LUMI: one of the fastest supercomputers in the world

- LUMI is owned by EUROHPC JU, operated by CSC in Kajaani
- HPL performance over **375 petaflop/s** makes the system one of the world's fastest

  Partial system listed 11/22 with 309 Pflop/s, #3 Top500, #7 Green500 (#2 of the big systems)

1 system

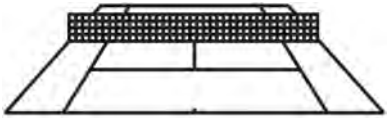
375 Pflop/s

Sustained performance

Computing power equivalent to

1 500 000

Modern laptop computers



Size of two tennis courts

Modern platform for

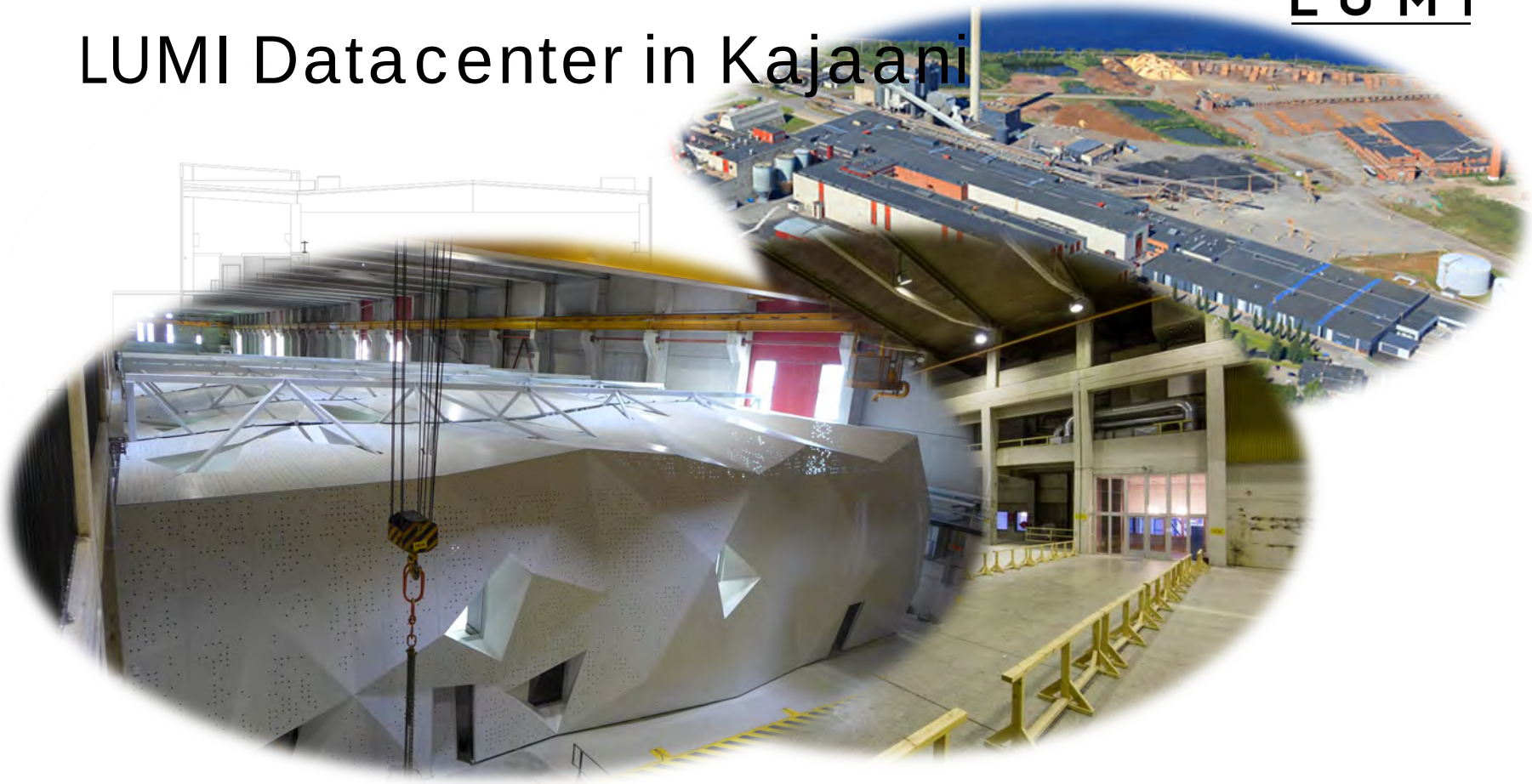
High-performance computing, Artificial intelligence, Data analytics

Based on GPU technology

LUMI Consortium

- Unique consortium of 10 countries with strong national HPC centers
- The resources of LUMI will be allocated per the investments
- The share of the EuroHPC JU (50%) will be allocated by a peer-review process (cf. PRACE Tier-0 access) and available for all European researchers
- The shares of the LUMI partner countries will be allocated by local considerations and policies – seen and handled as extensions to national resources

LUMI Datacenter in Kajaani




RENFORSIN RANTA
RATA
LUMI
KONE

Benefits of the brownfield solution

We assume having **reduced the CO2 footprint of LUMI data center construction by over 80%** with the brownfield solution vs. constructing an all-new building for LUMI

Materials - building shell 5,700 ft² (530 m²) office facility	Tonnes of CO_2	Percentage of total
Foundation (concrete)	4.7	4%
Flooring (concrete slab, insulation)	39.9	31%
Ceilings (plaster board)	2.3	2%
Structure (steel beams)	15.4	12%
External walls (brick, insulation)	32.1	25%
Internal walls (wood frame and plasterboard)	8.7	7%
Stairs (concrete)	1.1	1%
Windows (glass and frame)	0.59	0.4%
Internal doors (particle board)*	-0.4	-0.3%
External doors (plastic)	0.6	0.5%
Roof (wood, concrete, insulation)	23.4	18%
TOTAL	128.3	100%

For a 1 MW DC, source: Schneider-Electric white paper 66

Green electricity & waste heat utilization

LUMI

- **High capacity green power** is provided with **six links to the national grid**. Green energy production in the region, including three local hydro power plants, solar power and wind farms.
- LUMI uses **100% certificated hydro power (**with a Carbon Usage Effectiveness of close to zero) in all its data center production and office environments.
  - Kajaani DC area features green power up to 200 MW
  - Being green, not "buying green"
  - One outage during the last 39 years

Waste heat utilization

- **95% of LUMI's waste heat can be re-used in the district heating system of Kajaani**
  - Energy costs go down by 37% as local energy company pays for the waste heat.
  - As an alternative, 100% free cooling available, PUE 1.04
- With LUMI's heat, the local energy company can reduce the use of budned coal that corresponds in CO_2 emissions to removal of ~3000 cars from traffic

100 %
Carbon-neutral energy

CO_2 footprint
-12 400T
Co_2 eq/emissions

LUMI produces up to
20 %
of Kajaani's district heating needs

DRY AIR COOLING FOR BACK UP ~ 9 MW

DISTRICT HEATING NETWORKS ~ 10 MW

-Renforsin Ranta Business Park
-CITY of Kajaani

LUMI

+31 °C

+45 °C

+60 °C

+80 °C

Maximum heat production approximately 10 MW

HEAT EXCHANGERS

HEAT PUMPS

+32 °C

+44 °C

+32 °C

+44 °C

Electricity to heat pumps is CO_2 free.

HPC load Max. 9 MW

In addition of DLC there is approximately 1 MW of capacity for the air-cooled servers (e.g. storage). This IT is free-cooled.

LUMI DESIGN IMPACT

PUE 1.04 & 1.24

Support to Excess Heat Utilisation

Carbon Negative Operations

Use of Local Renewable Energy

20 % of annual need

40 % impact to total cost of energy

Footprint:
-12 400 tonnes

LUMI system level choices

- ICT equipment life cycle
  - Responsibility and sustainability required and rewareded in the CfT
- Operations and energy efficiency
  - Chosen GPU solution (and LUMI) will be at the top of Green500 over multiple lists to come
- Other considerations
  - LUMI is strongly positioned as an instrument for climate research, especially EU's Destination Earth programme
  - The "Climate Adaptation" digital twin will provide vital insight for how to mitigate the climate change

Concluding remarks

- Carbon footprint of ICT, HPC included, **does** matter!
- Green-ness of a HPC installation is fully dependent on the **data-center level choices**, especially contracted source of energy
- HPC systems **do not need to physically locate** in the country where it is being used or who owns it
- Carbon-neutral (even negative) HPC operations possible already today
  - Use 100% carbon-neutral energy (wind, nuclear, solar, hydro)
  - Reuse the excess heat, it is a big amount of energy (in big picture)
  - Repurpose existing buildings and use brownfield solutions instead of building new DCs

Acknowledgement:
Dr. Pekka Manninen ja Jukka-Pekka Partanen for the slides

Pekka Lehtovuori

DIrector,
Services for computational research

CSC – IT Center for Science Ltd.

phone.:09 457 2293
Firstname.lastname@csc.fi

PL 405, 02101 Espoo

www.csc.fi

facebook.com/CSCfi

twitter.com/CSCfi

youtube.com/CSCfi

linkedin.com/company/csc---it-center-for-science

github.com/CSCfi

Thomas Eickermann (Jülich Supercomputing Centre)

Towards more sustainable HPC at the Jülich Supercomputing Centre

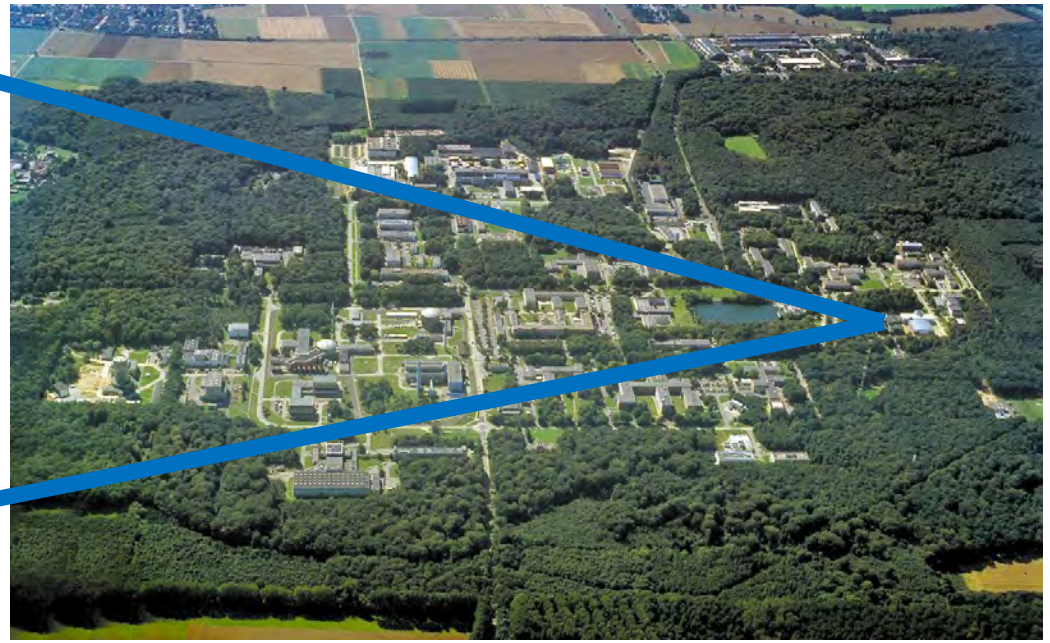
Abstract: With the increasing energy demand of high-end HPC systems, their efficient operation becomes ever more important both from an economic and ecologic perspective. JSC follows a holistic approach, ranging from the selection of HPC technology to the reuse of waste heat. The presentation gives an overview of its activities in this respect.

Bio: Thomas Eickermann is working at Jülich Supercomputing Centre (JSC) since he finished his PhD in Physics at the University of Düsseldorf in 1994. His activities cover system administration, Grid computing, and networking. In 2002, he became head of the communication systems division of JSC. Between 2008 and 2015, Thomas Eickermann has been project manager of PRACE preparatory and implementation phase projects and served on the PRACE aisbl Board of Directors. Currently, he is a deputy director of JSC and engaged in the preparations for a EuroHPC Exascale computer in Jülich.

Towards sustainable HPC at the Jülich Supercomputing Centre

CIUK 2022 | December 1, 2022 | Thomas Eickermann

Research Centre Jülich by Numbers



Research areas

- Information
- Energy
- Bioeconomy

- **Budget:** 861 Mio €,
  including 395 Mio € third party funding
  171 Horizon 2020 projects, 420 national projects
- **Employees:** 7.120
  incl. 2.626 scientists including PhD students
  934 guest scientists from 65 countries
- **Publications:** 3.081
  (source: fact sheet 2021)

Jülich Supercomputing Centre (JSC)

Facts and Figures

Staff:

220 Total (185 FTE)

160 Scientists

13 PhD Students (+13 external)

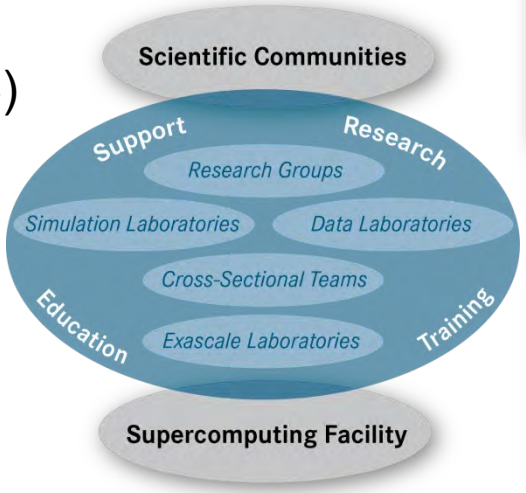
Budget:

30 Mio. € Institutional Funding (PoF)

15 Mio. € Third Party Funding

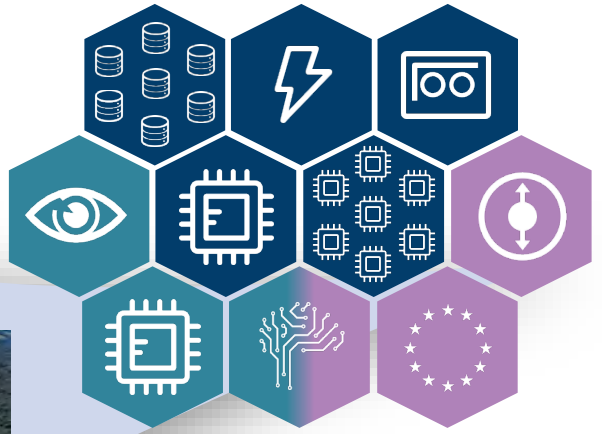
Jülich Supercomputing Centre at a Glance

- **Supercomputer operation for**
  - Centre – FZJ
  - Region – RWTH Aachen University
  - Germany – Gauss Centre for Supercomputing (GCS)
    John von Neumann Institute for Computing (NIC)
  - Europe – PRACE, EU projects, EuroHPC
- **Application support**
  - Unique support & research environment at JSC
  - Peer review support and coordination
- **R&D work**
  - Methods and algorithms, computational science, performance analysis and tools
  - Scientific Big Data Analytics with HPC
  - Computer architectures, Co-Design, Exascale Labs together with IBM, Intel, NVIDIA
- **Education and training**



Towards Sustainable HPC at JSC

Optimisation of Energy Usage



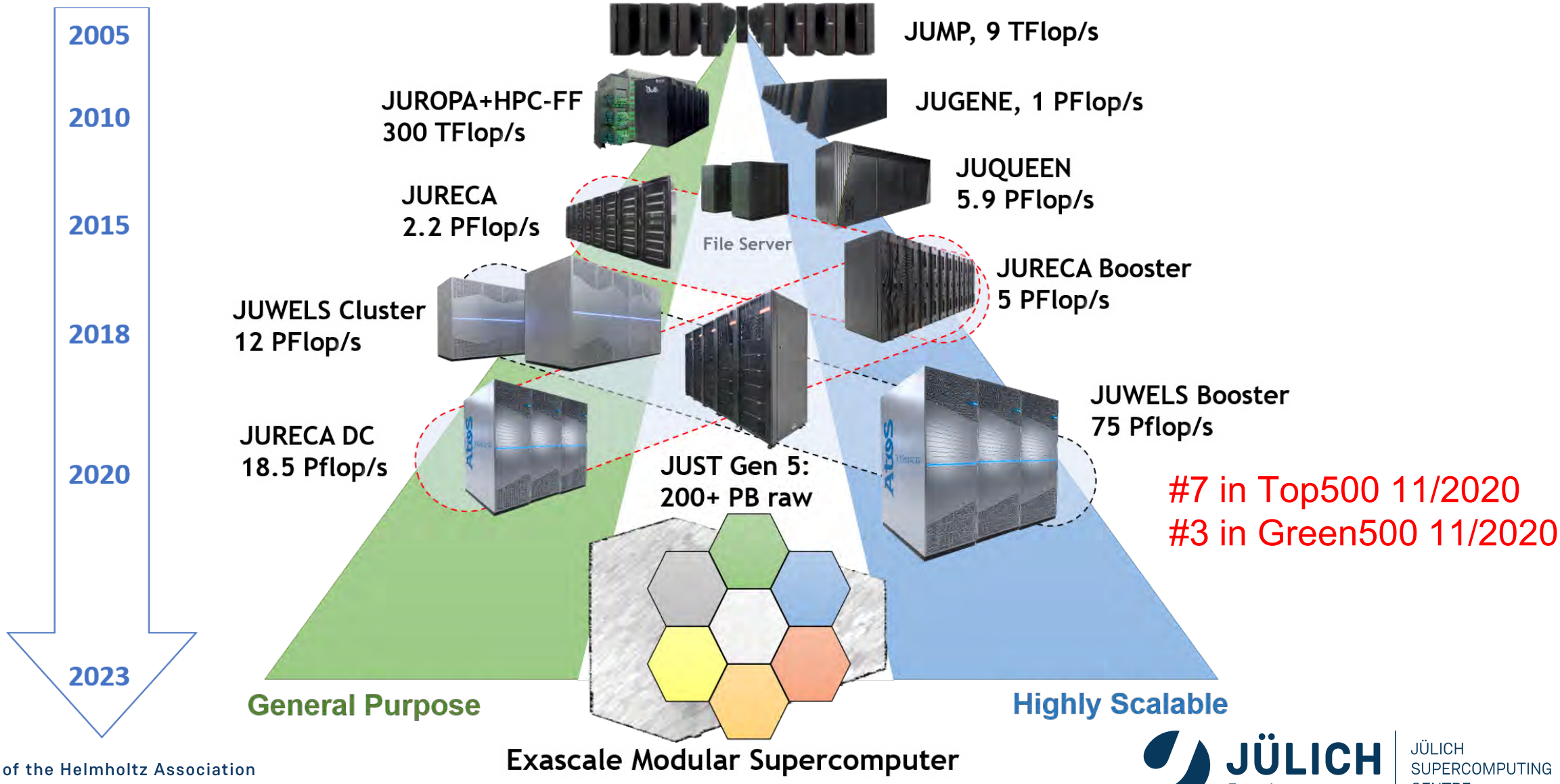
Future Plans

Campus Level

Data Centre Level

System Level

(DUAL) hardware strategy at JSC



System Level

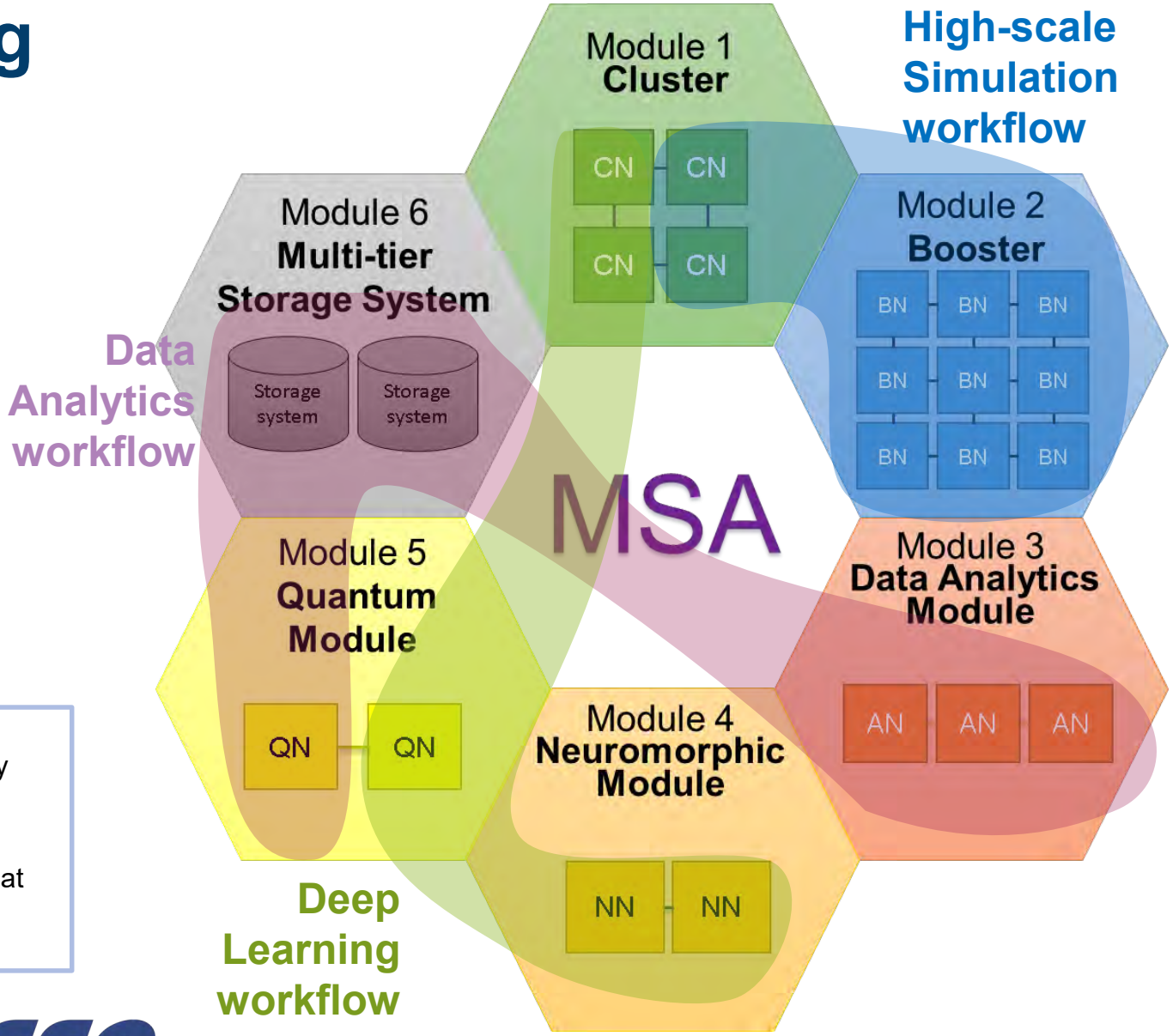
- **Energy-efficient compute nodes**
  - GPU accelerators boost Flops/W
- **Energy-efficient system architectures**
  - Many applications cannot benefit from GPUs (today)
  - Idle GPUs are not energy-efficient
  - Dual hardware strategy: General Purpose + Highly Scalable system for different demands and mixed workflows
  - 35% of JUWELS Booster projects have also allocations on the JUWELS Cluster
  - Modular Supercomputer Architecture: tight integration of heterogeneous resources

Modular supercomputing architecture

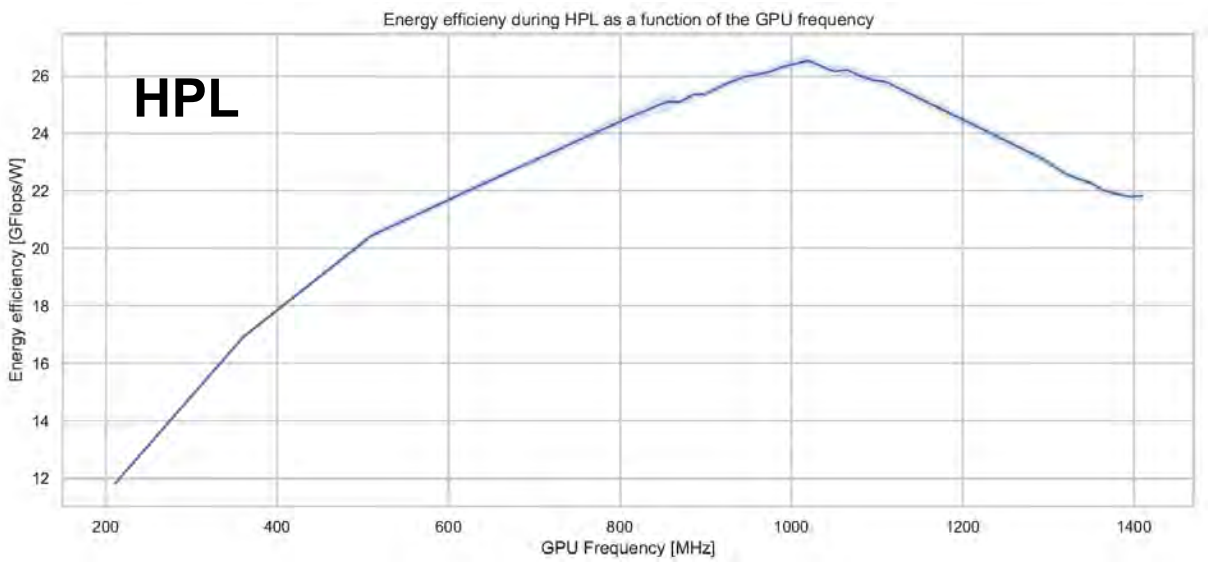
Composability of heterogeneous resources

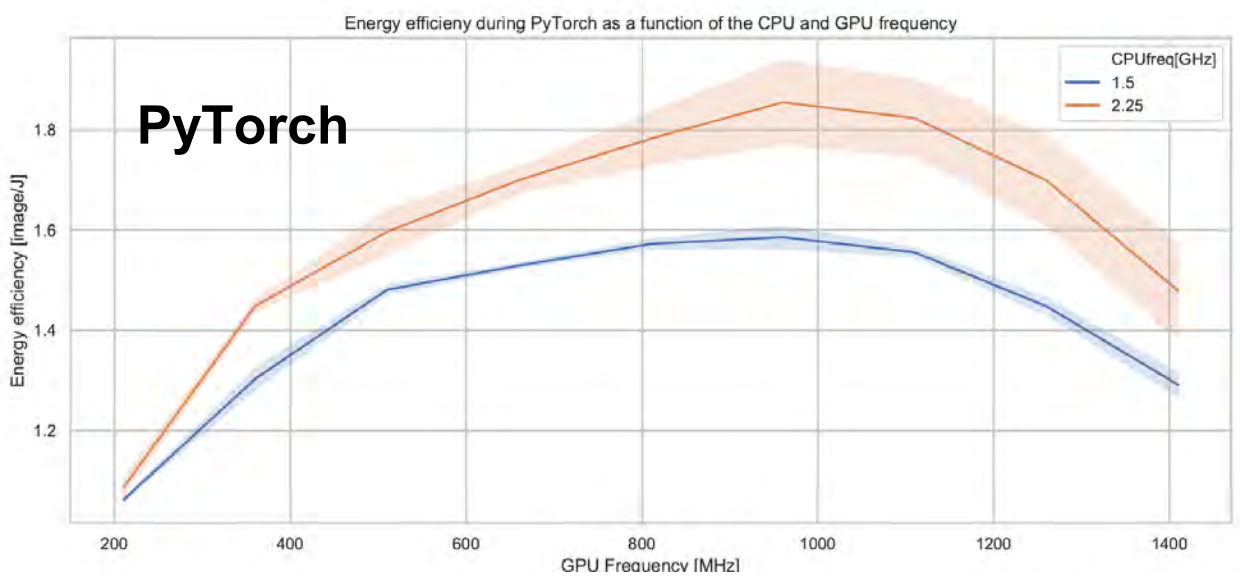
- Cost-effective scaling
- Effective resource-sharing
- Match application diversity
- Large-scale, complex workflows

- **E. Suarez**, N. Eicker, Th. Lippert, "*Modular Supercomputing Architecture: from idea to production*", Chapter 9 in Contemporary High Performance Computing: from Petascale toward Exascale, Volume 3, p 223-251, CRC Press. (2019)
- **E. Suarez**, N. Eicker, and Th. Lippert, "Supercomputer Evolution at JSC", Proceedings of the 2018 NIC Symposium, Vol.49, p.1-12, (2018)



System Operation: Adaptation of GPU / CPU Frequencies





GPU freq. [MHz]	Time-to-solution	Energy to solution	TCO to solution
1410	-	-	-
1095	7.82 %	- 15.59%	- 6,2%
1029	11.33 %	- 17.71%	- 6,1%

GPU freq. [MHz]	Time-to-solution	Energy to solution	TCO to solution
1410	-	-	-
1110	3.32%	18.85%	-10,0%
960	9.01%	20.18%	-8,5%

Measurements on JURECA-DC: 2x AMD EPIC 7742, 4x NVIDIA A100-SMX4-40GB

by Sebastian Achilles (JSC)

System Operation Cont.

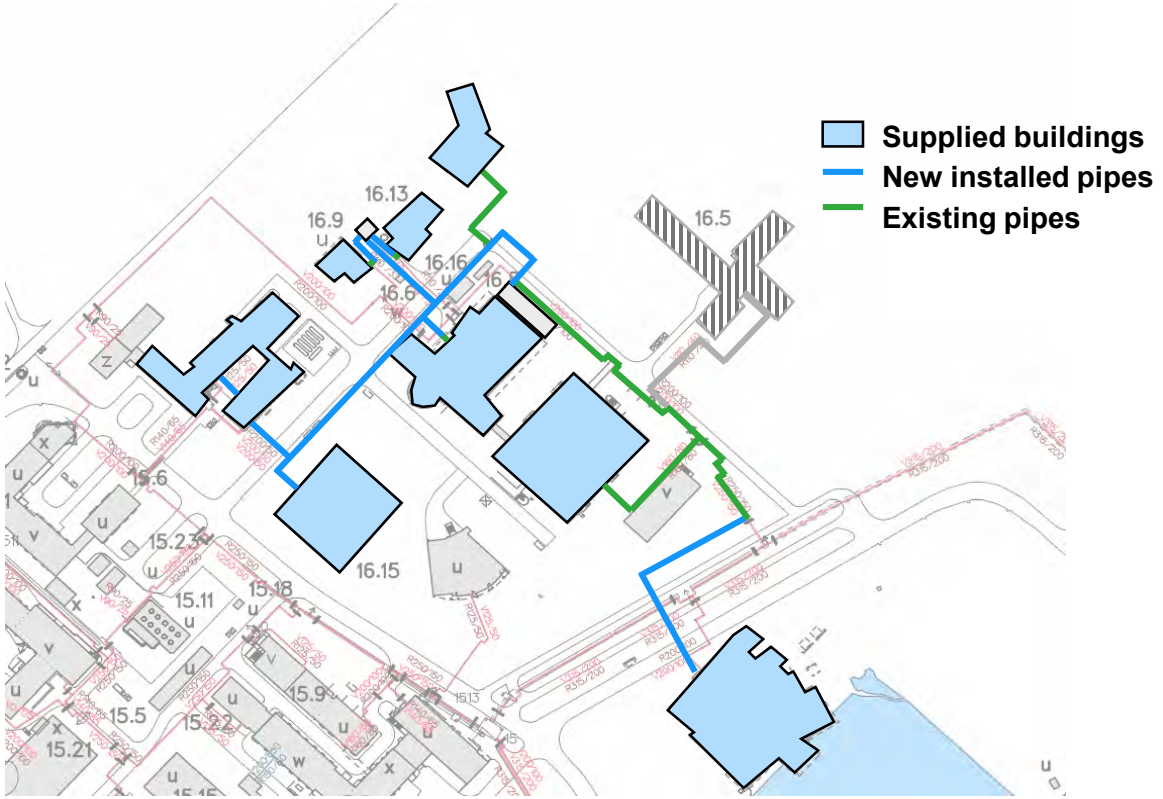
- **GPU Frequency adaptation**
  - Extended test opportunities provided for JUWELS users
  - No significant gain in energy-to-solution for many applications, 5-10% for some
- **Powering off idle nodes**
  - JUWELS is fully loaded, but …
    - Scheduling a mix of small and large node-count jobs leads to idle periods
  - Tested and put into production on smaller systems, incl. JURECA-DC
    - Reduced interconnect stability
    - Little impact on user experience

Data-Centre Level – Cooling

- **Until 2022**
  - JUWELS and JURECA-DC use direct liquid cooling
  - Chilled water is centrally supplied for the Jülich campus
  - Coefficient of Performance ~ 2.5
  - Supported by free cooling in winter
- **Since May 2022**
  - 1.8 MW Hybrid warm-water cooling:
    - inlet ~ 34 °C outlet ~ 42 °C
  - Free cooling + water evaporation in hot periods
  - PUE ~ 1.1
  - Extension to 3 MW is underway
  - Chilled water only for air cooled components: storage, network

Research Centre Level – Waste Heat Usage

- **Living Lab Energy Campus**
  - A project to develop and deploy an integrated campus-wide energy management
    - Renewable energy production and storage
    - Monitoring and predicting usage & steering energy production (e.g. gas-fired combined cooling, heating and power (CCHP) plant) and battery usage
- **Under Construction**
  - Low-temperature (~ 40 °C) district heating system powered by JSC waste heat
  - Temperature is sufficient to directly heat buildings fulfilling current German insulation standards
  - Heat pumps used to achieve temperature levels (~ 70 °C) required for older buildings – such as the JSC



JUPITER – The 1st European Exascale System

EuroHPC Joint Undertaking

SPONSORED BY THE

Federal Ministry of Education and Research

Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

- **EuroHPC Joint Undertaking**
  - Joint undertaking between EU, member states, private partners
  - Took over funding of HPC related projects from EC
  - Co-funds Petascale, and owns Pre-Exascale, and Exascale systems

- **JUPITER - JU Pioneer for Innovative and Transformative Exascale Research**
  - Selected on June 14, 2022 as the 1st EuroHPC Exascale system
  - Installation in Jülich targeted for end of 2023
  - 500 Mio. € Total Costs, equally shared between EuroHPC and Germany (federal and state of North Rhine-Westfalia funding)

JUPITER – Modular Supercomputer

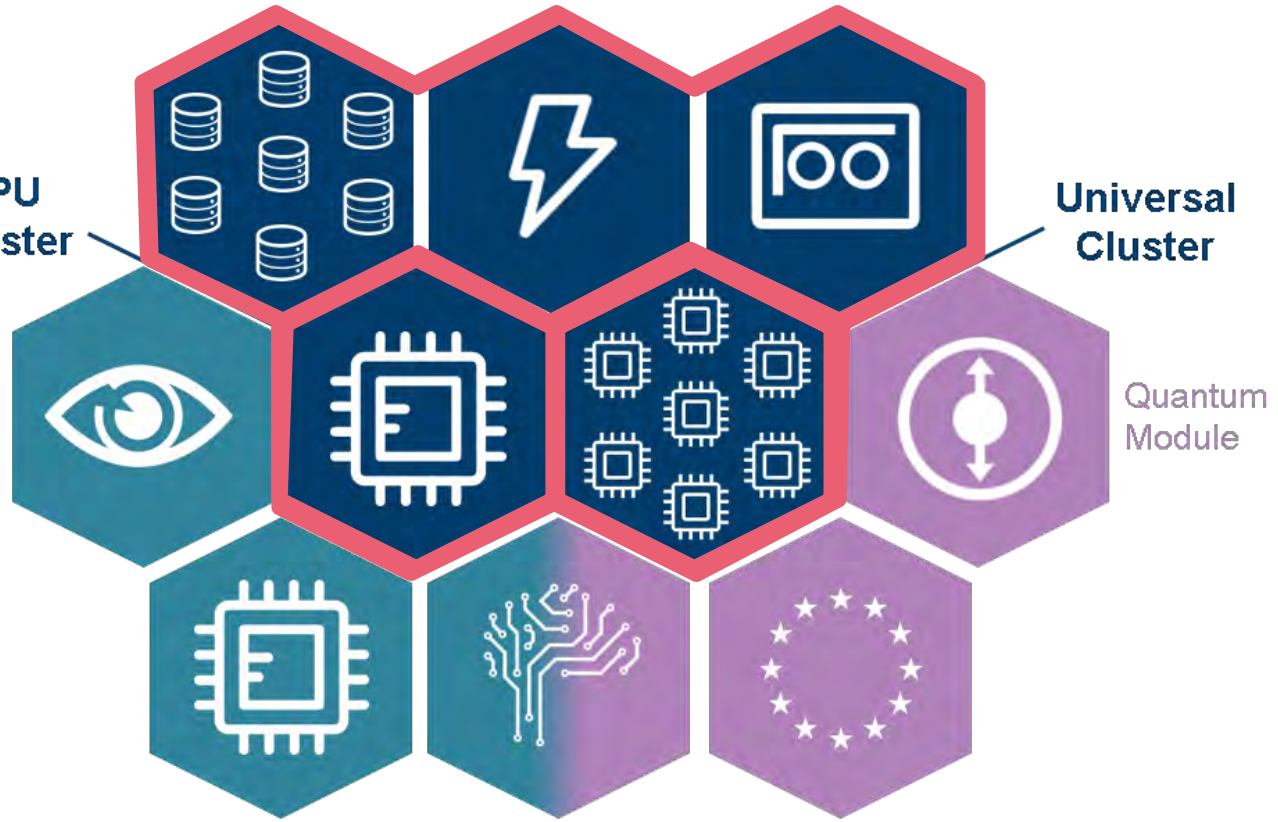
Target >20×
application performance compared to JUWELS Booster

> 1 EB
Parallel High Capacity Data System

Parallel High Bandwidth Flash Module

High Capacity Backup/Archive System

> 1 EF
GPU



7.5 PB/s
x86 or ARM

- **Basis Configuration**
- **Optional Modules**
- **Future Technology Modules**

JUPITER - Towards Sustainability

- **JUPITER will leverage all of the above:**
  - Modular Supercomputer Architecture
  - GPU-based booster
  - Operated with green electricity
  - Direct warm-water cooling
  - Waste heat usage: funding secured for
    - a high-performance heat pump (> 1 MW)
    - Measures on the campus that enable broader utilization of JUPITER's waste heat
  - Optimisation of energy supply
    - From: 110 kV → 35 kV → 10 kV → 480 V
    - To: 110 kV → 35 kV → 480 V

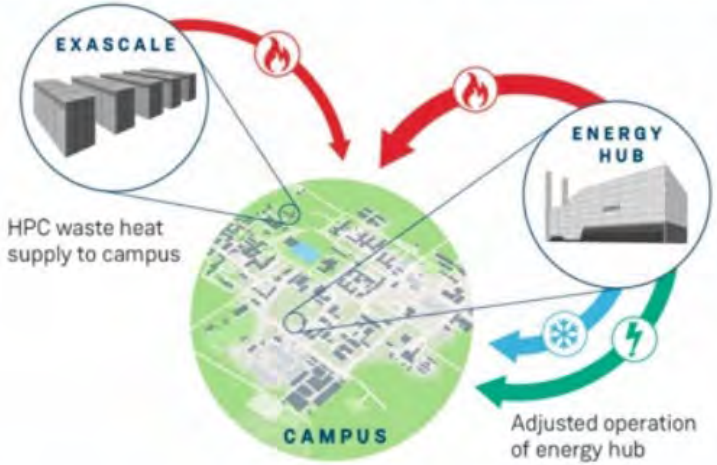
Location of Exascale Data Centre

SPONSORED BY THE
Federal Ministry of Education and Research
Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

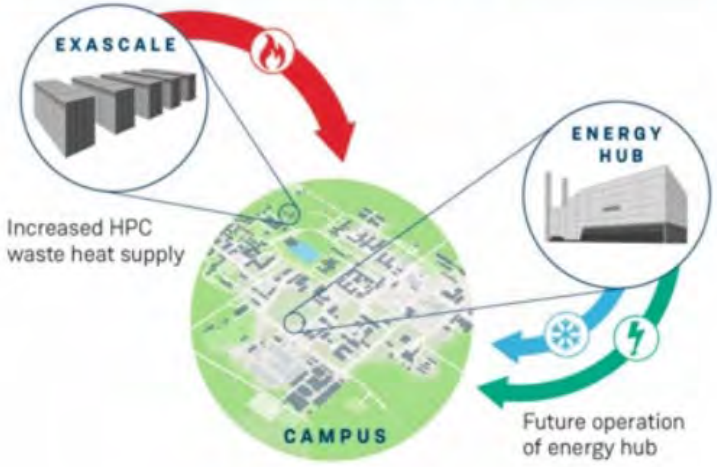
Waste Heat Usage – Long-Term Vision

1. Actual energy supply

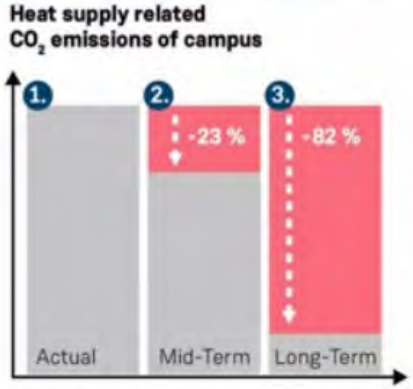
2. Mid-Term energy supply



3. Long-Term energy supply



Effects of waste heat integration



By using higher share of renewable electricity, the heat related emissions could further decrease

Mid-Term energy supply:
- Reducing operational costs by adjustments at actual energy hub operation

Long-Term energy supply:
- Main heat supply by HPC waste heat
- Actual energy hub mainly supply cooling

- **JUPITER average power ~ 15 MW**
- **Campus heat demand ~7.5 MW**

gas-fired combined cooling, heating and power (CCHP) plant

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

Towards Sustainable HPC at JSC

Optimisation of Energy Usage

Future Plans
JUPITER

Campus Level:
waste heat usage

Data Centre Level: free cooling

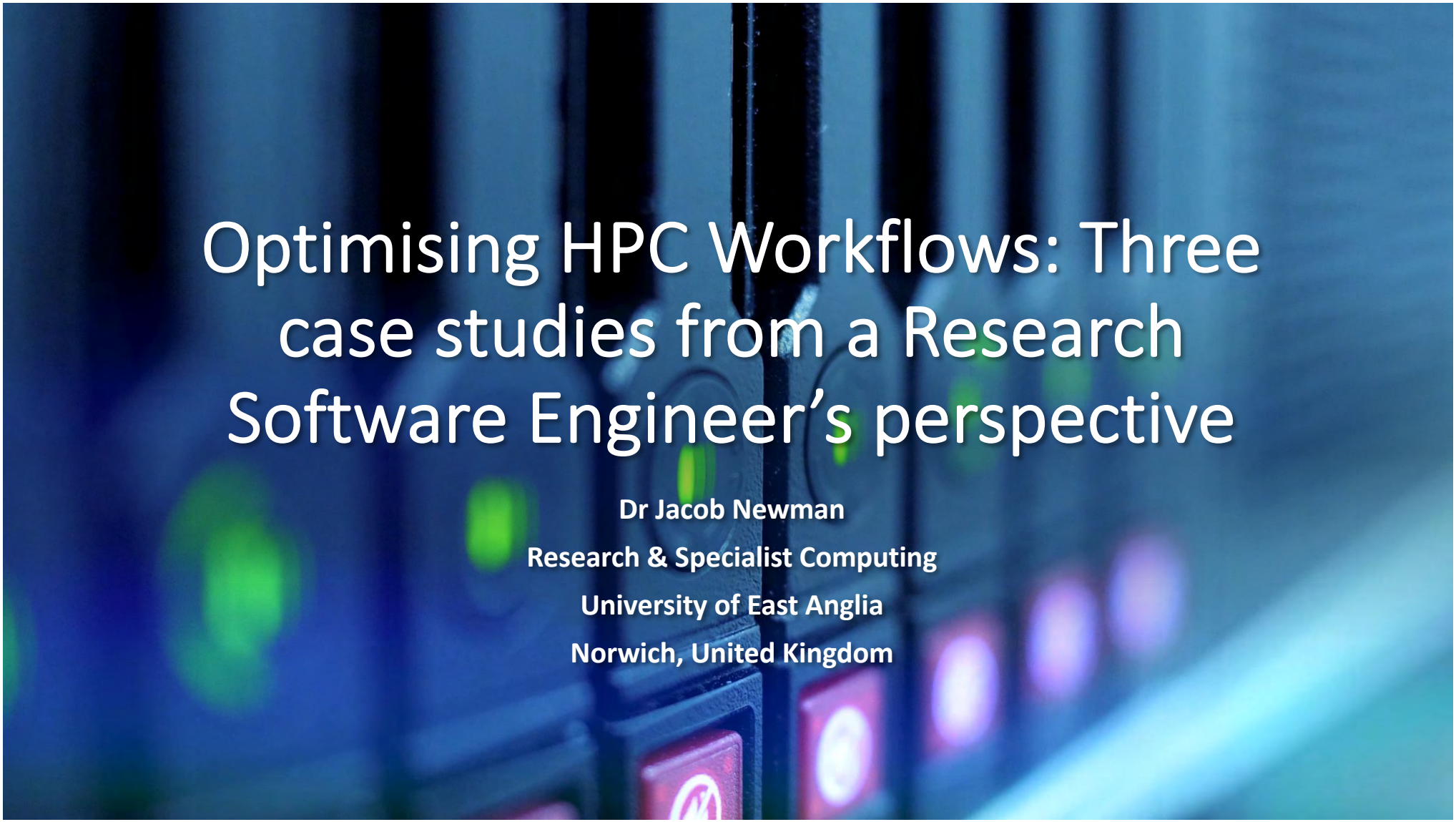
System Level: GPUs, Modular Architecture

Jacob Newman (University of East Anglia)

Optimising HPC Workflows: Three Case Studies from a Research Software Engineer's Perspective

Abstract: The University of East Anglia's HPC service has recently appointed a Research Software Engineer (RSE) to address the challenge of optimising users' jobs to make more efficient use of the available resources. In this talk, I will present three case studies of users requesting assistance to optimise their computational workflows on our HPC. I will describe how their requests were presented, the solutions considered and selected, and quantify the speed improvements obtained. Solutions explored will range from identifying simple idiosyncrasies due to software versioning, through to utilising GPU technology more effectively.

Bio: Jacob studied an undergraduate degree in Computing with Electronics, and a PhD in Computing entitled, "Language Identification Using Visual Features". He worked as a researcher for 11 years in the areas of speech recognition, computational biology, and medical computing. In 2021, Jacob joined the University of East Anglia's Research and Specialist Computing Support team, where his main responsibilities are to develop and optimise user workflows on the HPC.



Optimising HPC Workflows: Three case studies from a Research Software Engineer's perspective

Dr Jacob Newman

Research & Specialist Computing

University of East Anglia

Norwich, United Kingdom

Me: Research Software Engineer

- Previously a computing researcher
- Extensive user experience of HPC
- Advocate for HPC in research
- Appointed RSE in May 2021

Outreach

Optimisation

UEA's HPC: Ada

- Centos 7
- Slurm 19.05
- About 10,000 cores
- Across 173 CPU nodes
- 57 GPUs
- Including 30 Nvidia Quadro RTX6000s
- InfiniBand 100Gb/sec
- GPU/Compute/IB/Hi-Mem/Vis partitions

Our users

- About 1000 users
- Range of disciplines
- Simulations, modelling, machine learning, data processing, code development, etc.
- Associated levels of computing proficiency


3 Case Studies

User A

Senior lecturer in Chemistry

Researching molecular dynamics

GROMACS

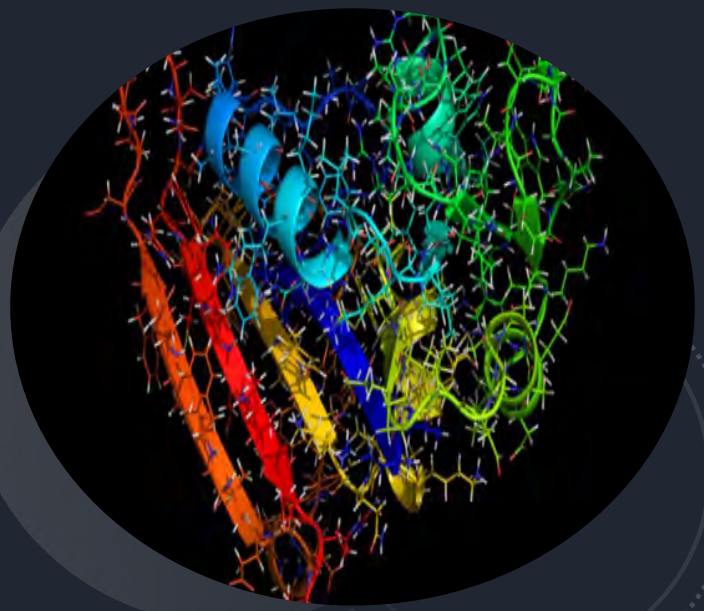
User A

“Currently I am trying to run molecular dynamics calculations using **gromacs** on the **compute** queues. These show **0.32 ns/day** on a shared node, **1.31 ns/day** on an exclusive node. Perhaps I am not using the queue properly because on my desktop **at home** (Intel i5-2500 @3.3GHz) I can achieve over **8 ns/day**. Anyway, **I need** a simulation of **200 ns** or so, hence I am wondering whether the other hardware available will offer better performance. If so, can I be given access to the appropriate queue. gromacs is **GPU aware** though whether this offers a significant improvement depends on the balance of calculations and the advice is to test the system.”

User A

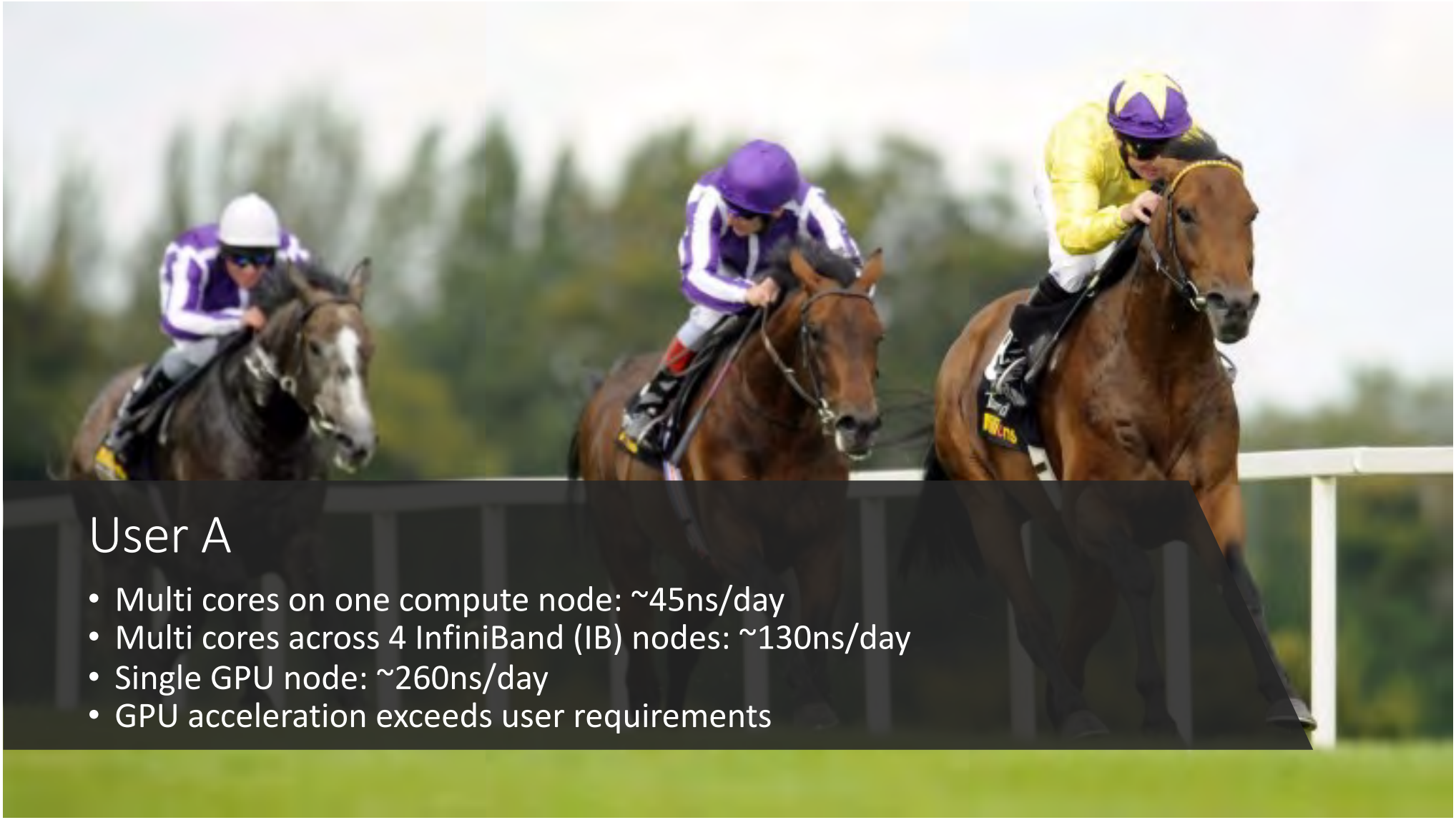
- Replicate results
- Now what?
- 2019.4
- 2020.4
- 34x speed increase
- ~45 ns/day, which is still not close to 200 ns/day

```
   gromacs/2019.4/gcc                    (D)
   gromacs/2020.4/gcc-gpu-cuda11.0
   gromacs/2020.4/gcc-openmpi-eth        (D)
```



MPI

GPU



User A

- Multi cores on one compute node: ~45ns/day
- Multi cores across 4 InfiniBand (IB) nodes: ~130ns/day
- Single GPU node: ~260ns/day
- GPU acceleration exceeds user requirements

User B

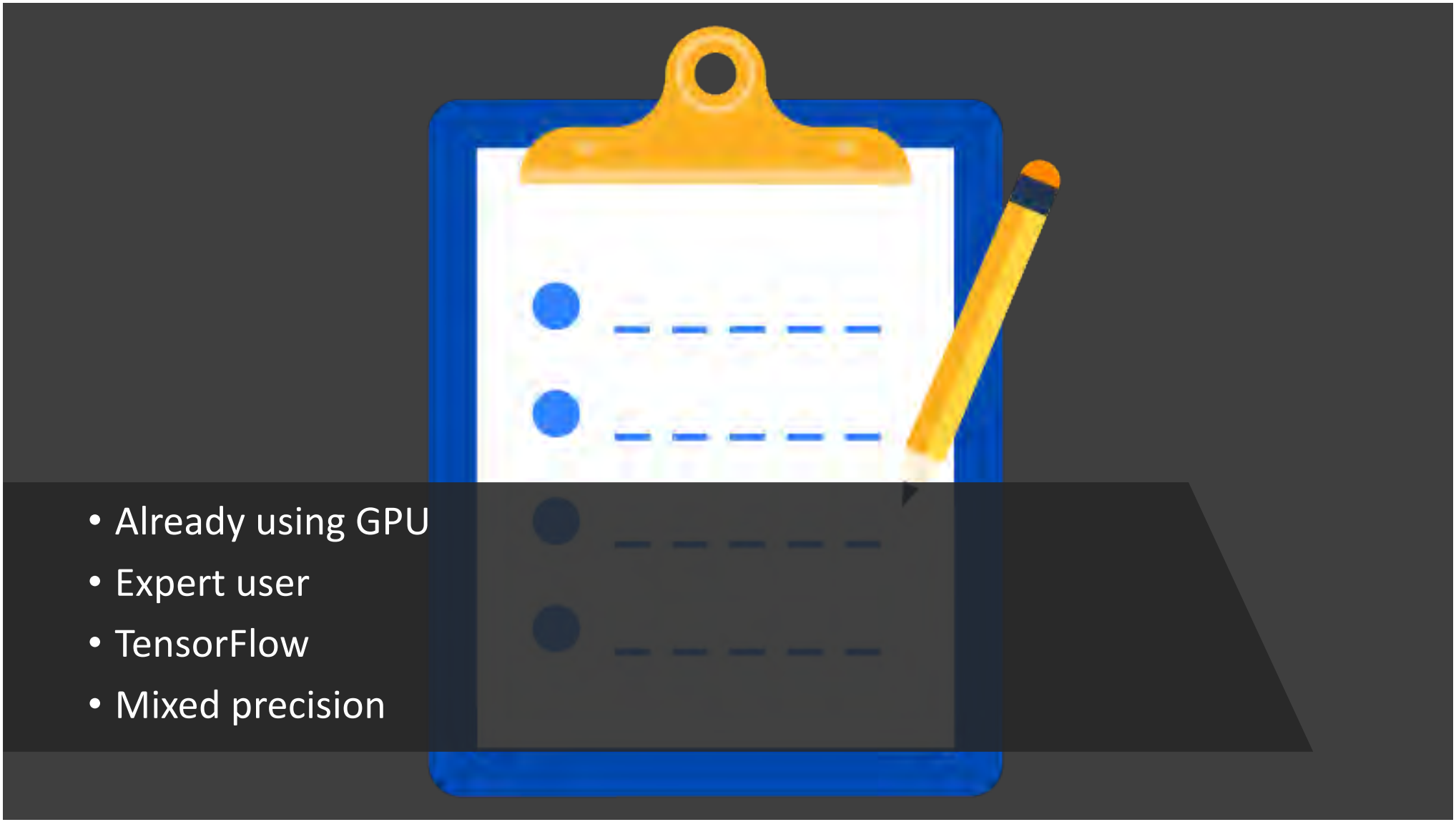
Postgraduate student in Computing

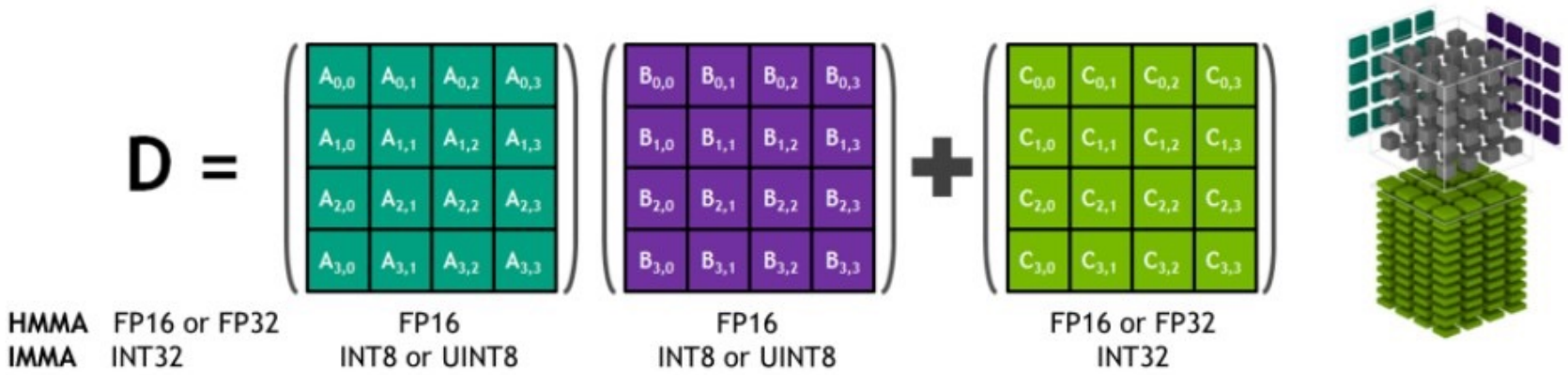
Time-series classification

TensorFlow

User B

Are there any way you can think of that might improve efficiency/speed up GPU jobs?


• Already using GPU
• Expert user
• TensorFlow
• Mixed precision



Tensor Cores

- Reduces memory usage
- Speeds up data transfer
- Performance increases, shortening job duration

Mixed precision

```
from tensorflow.keras import mixed_precision

# The feature is experimental in TensorFlow < 2.4. >= 2.4
it is a main feature.
mixed_precision.experimental.set_policy('mixed_float16’)

# The output layer needs to be float32 for numerical
stability
dense_1 = Dense(nclass, activation=activations.softmax,
dtype='float32', name="dense_3_ptbdb")(dense_1)
```

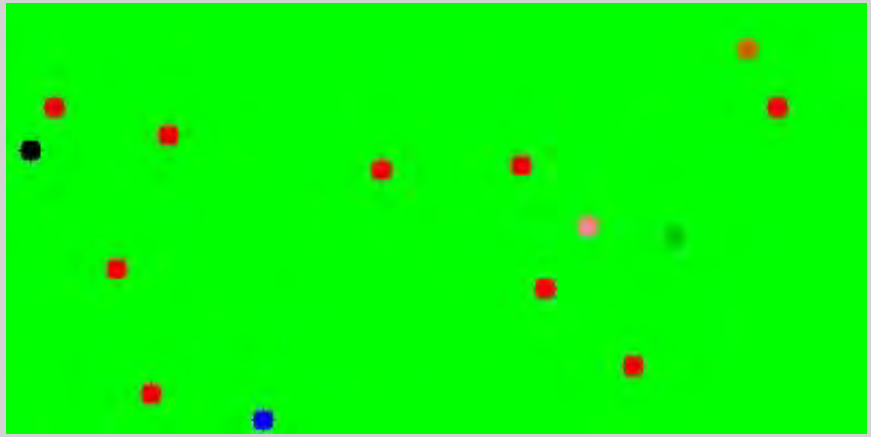
My own tests

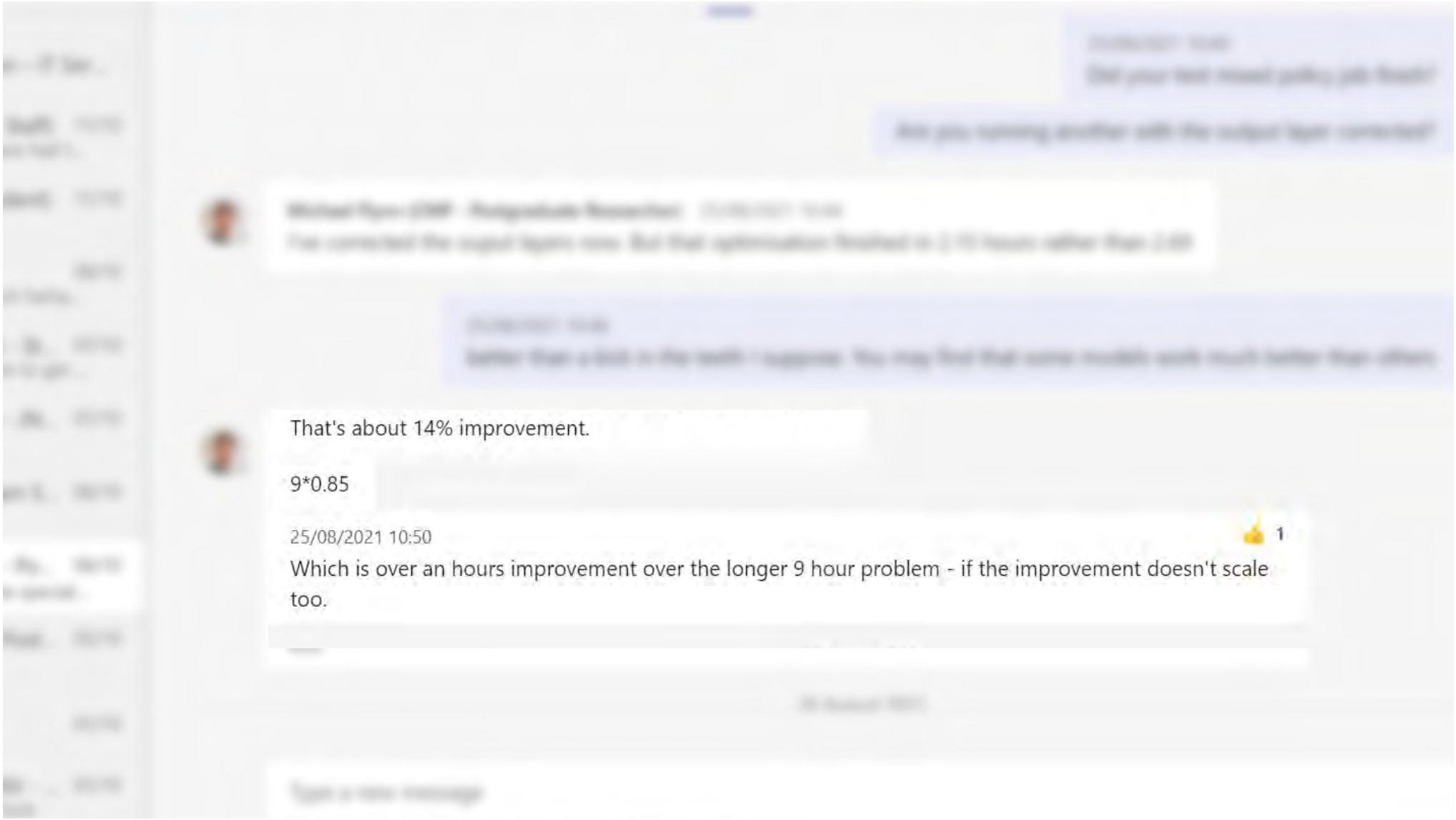
Without the mixed policy…

```
Epoch 1/100
1490/1490 [======] -
1124s 754ms/step
```

And with the mixed policy…

```
Epoch 1/100
1490/1490 [======] -
485s 325ms/step
```




That's about 14% improvement.
9*0.85
25/08/2021 10:50
1
Which is over an hours improvement over the longer 9 hour problem - if the improvement doesn't scale too.

User C

Researcher in Biology

Quantifying bird abundance

R

User C

Hi Jake,

I hope you are well. I have been using the code you made for Claire to run some analysis on the cluster, however after 2 weeks it timed out.

I think the best idea is to try and split it up using the site index but I have no idea how to do 2 array indexes or if that is even possible? I have attached all the code. Do you have some time to talk about this soon?

Many thanks,
Cat

Current workflow

- For 16 datasets
  - For 15k sites
    - Process 330k records
- Takes at least two weeks to complete
- Current implementation submits a job for each dataset
- Bash script to submit an array job from idx 1 to 16

- Suggestion to process the 15k sites in smaller batches

New workflow

		Dataset ID		
		1	2	3
Site ID	1	Array ID 1	Array ID 2	Array ID 3
	2	Array ID 4	Array ID 5	Array ID 6
	3	Array ID 7	Array ID 8	Array ID 9

New workflow

```
#!/bin/bash
SBATCH_ARRAY_INX=$(ls ./datasets/*.csv | wc -l)
SBATCH_ARRAY_INX=$((SBATCH_ARRAY_INX*10))
sbatch --array=1-$SBATCH_ARRAY_INX pairwise_site.sh
```

New workflow

```
idx <- strtoi(args[1],base=10)
num_datasets <- strtoi(args[2],base=10)

dataset_idx <- ((idx-1)%%num_datasets)+1
site_idx <- ceiling(idx/num_datasets)
```

Worked example, where idx=19 and num_datasets=16:

dataset_idx = ((19-1)%%16)+1 = 3

site_idx = ceiling(19/16) = ceiling(1.1875) = 2

Results

- Longest job completed in 4.5 days compared to "over two weeks"
- The cumulative time is 389 days
- But more work to be done…
- Could increase granularity
- Could ensure the jobs are distributed equally
- But in this case, five days is acceptable

Benefits of an RSE service

Since you started helping I've managed to get 1,200 gpu jobs completed

[redacted] Jake was brilliant and worked with me for weeks very patiently to fix this, I can't thank him enough. [redacted]

Hi Jake,

This is looking great. I think it is exactly what I want.

it work, thaaaaaank you

Thanks a lot Jacob for your kind help!

Benefits of an RSE service

- Relationship building
- Recognition of RSCS
- Freeing time
- Proactive mitigation
- Jobs complete more quickly
- Improved productivity: global problems solved faster

Increased efficiency?

“REDUCE POWER CONSUMPTION”?

INCREASE RESEARCH OUTPUT

POWER MANAGEMENT

Conclusions

- No two jobs are the same
- Impressive improvements can be made using simple solutions
- It is worth trying several approaches
- Productive users and good relationships

Thank you for listening

Gabryel Mason-Williams (Rosalind Franklin Institute)

DisTRaC: Accelerating High-Performance Data Processing

Abstract: Clusters in high-performance computing (HPC) are composed of four main components: a job scheduler, compute nodes, networking, and storage - each of which plays a role in the performance and sustainability of the system. DisTRaC is a Ceph deployment tool for creating temporary distributed RAM-disk file/object stores across job compute nodes. DisTRaC can remove the need for a parallel filesystem and the out-of-cluster network IO, reducing competition between concurrent jobs. Increasing efficiency while lowering external network and storage utilisation lowers CO2 emissions and costs associated with HPC, helping move towards Net Zero. In this talk, we introduce DisTRaC as a mechanism to increase the efficiency of applications without needing modification and showcase benchmark results against real-world HPC tasks.

Bio: Gabryel Mason-Williams is a computer science graduate from the University of Plymouth. During their studies, they did a year in industry with Diamond Light Source conducting research into "High-Performance Object Stores for Big Data Processing" and a dissertation project titled "ALaBDaC-Automated Lab Book Data Collection" in collaboration with the Rosalind Franklin Institute. After this, Gabryel worked at the Rosalind Franklin Institute, building upon the success of the research conducted at Diamond Light Source and focusing on novel approaches to HPC, cloud, compression and storage. They are currently studying an MSc in Artificial Intelligence at Queen Mary University of London whilst continuing to work with the Rosalind Franklin Institute. Following the MSc, they plan to do a PhD in Machine Learning.

DisTRaC: Distributed Transient Ram Ceph

Accelerating High-Performance Data Processing

Gabryel Mason-Williams

gabryel.mason-williams@rfi.ac.uk

About the Rosalind Franklin Institute and me

- The Rosalind Franklin Institute:
  - A United Kingdom Research Institute dedicated to developing new technologies to tackle important health research challenges. Based in Harwell Campus, Didcot and funded by the UKRI ESPRC.
  - **5 Themes**: Artificial Intelligence and Informatics, Biological Mass Spectrometry, Correlated Imaging, Next Generation Chemistry and Structural Biology
- Me:
  - Currently studying an MSc in Artificial Intelligence at Queen Mary University of London and working as a Research Software Associate at The Rosalind Franklin Institute
  - Looking for a PhD

UKRI Engineering and Physical Sciences Research Council

diamond

University of BRISTOL

The Rosalind Franklin Institute

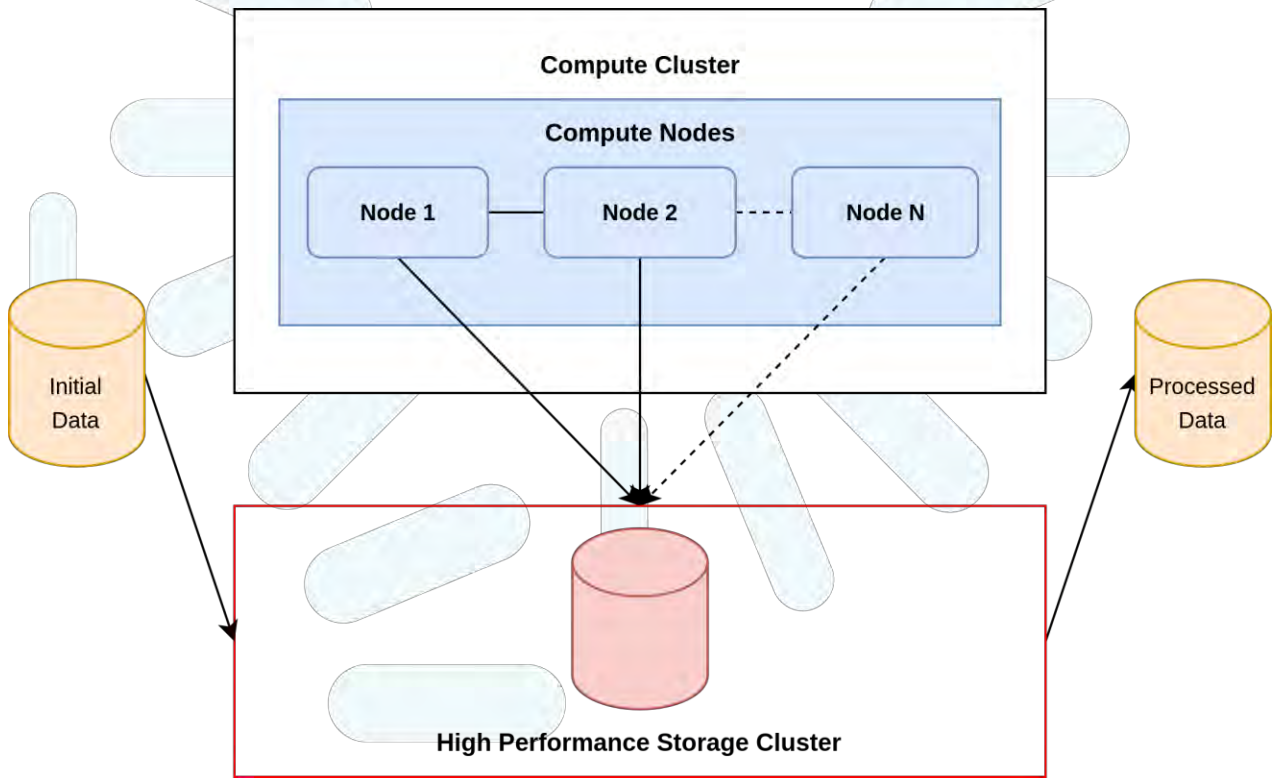
Contents

- Background
- The Problem
- What is DisTRaC
- Solving the Problem
- Case Studies
- Ongoing Work
- Conclusion
- Acknowledgments
- Questions

Background

High Performance Compute Cluster

- Job scheduler
- Compute – lots of RAM
- Storage
- Networking – Storage and Interconnect



Traditional High Performance Compute Cluster Setup

What is the Problem

Engineering and Physical Sciences Research Council

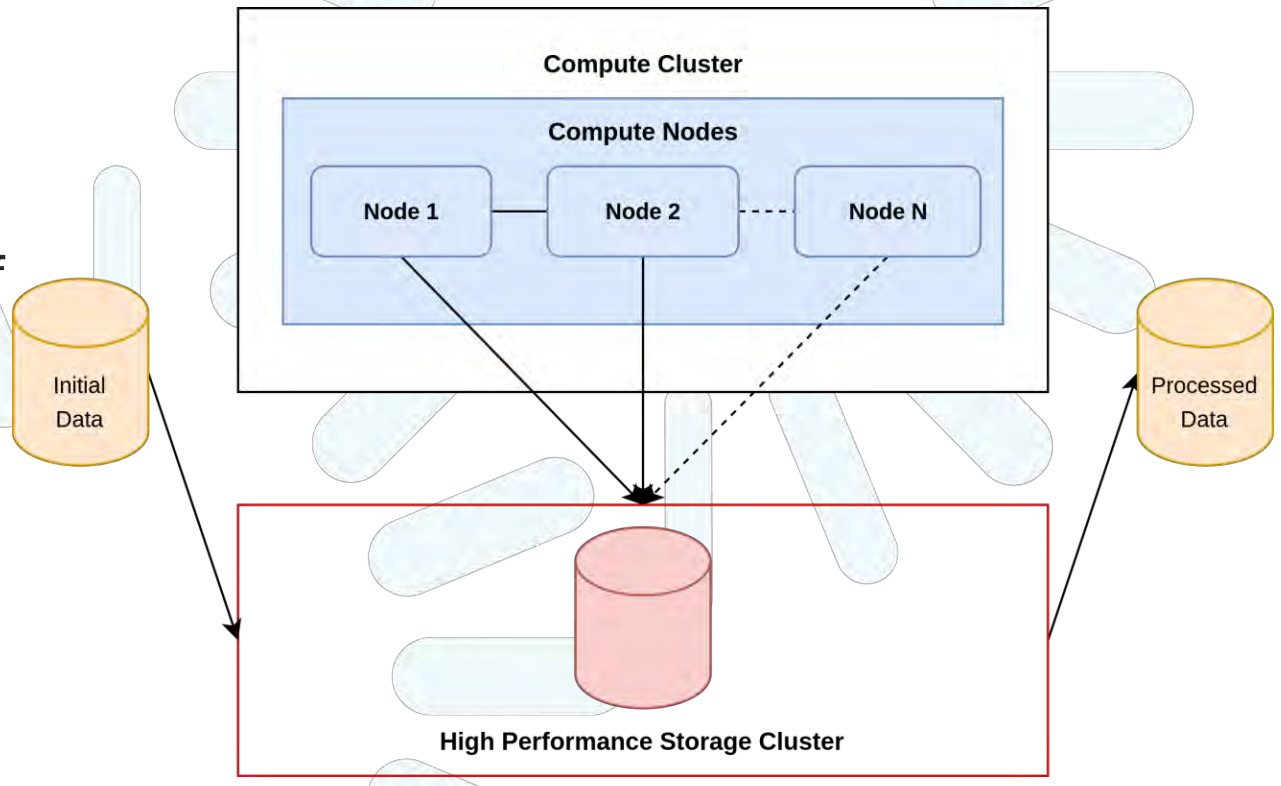
diamond

University of BRISTOL

The Rosalind Franklin Institute

Problem

- Network connection limits IO Bound Applications
- Shared Storage Resourced
  - Users' can affect the performance of others
- Storing of Intermediate Data
- Storage clusters are expensive, hard to maintain and set up, especially in cloud
- Inefficient use of resources



Traditional High Performance Compute Cluster Setup

Solution?

Engineering and Physical Sciences Research Council

diamond

University of BRISTOL

The Rosalind Franklin Institute

DisTRaC

https://github.com/rosalindfranklininstitute/DisTRaC

Engineering and Physical Sciences Research Council

What is DisTRaC?

- Distributed Transient Ram Ceph
- A program for deploying a transient Ceph [1] cluster onto HPC infrastructure utilising RAM in a scalable and efficient manner.
- Creating a job persistent and isolated in-memory file/object store for HPC applications.

Why DisTRaC and not another deployment tool

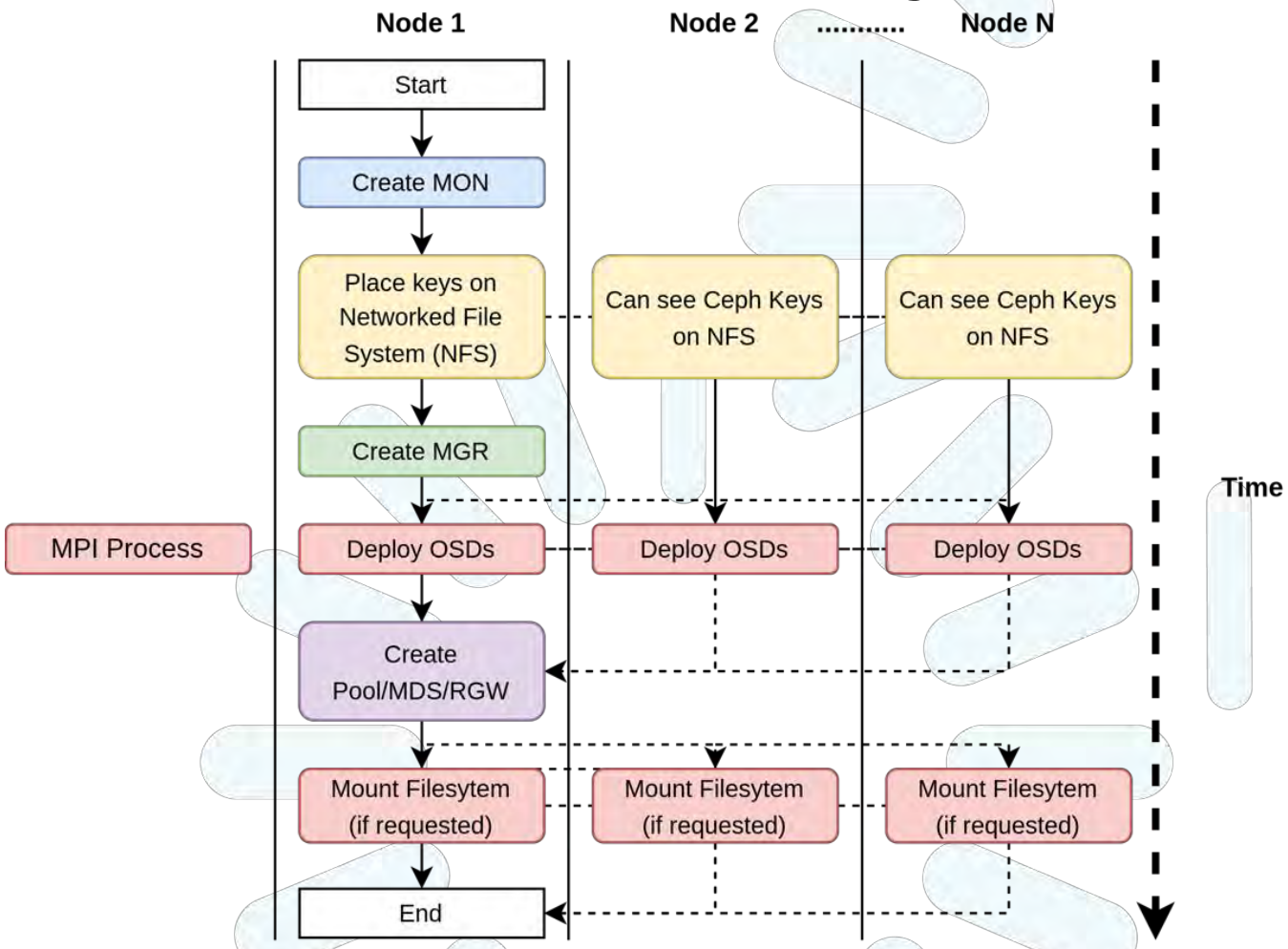
Current Deployment tools

- Designed to build long-lasting maintainable clusters
  - Lots of safety checks
  - Slow to deploy and remove clusters
- Sequential
- Require passwordless SSH

We need something quick and efficient

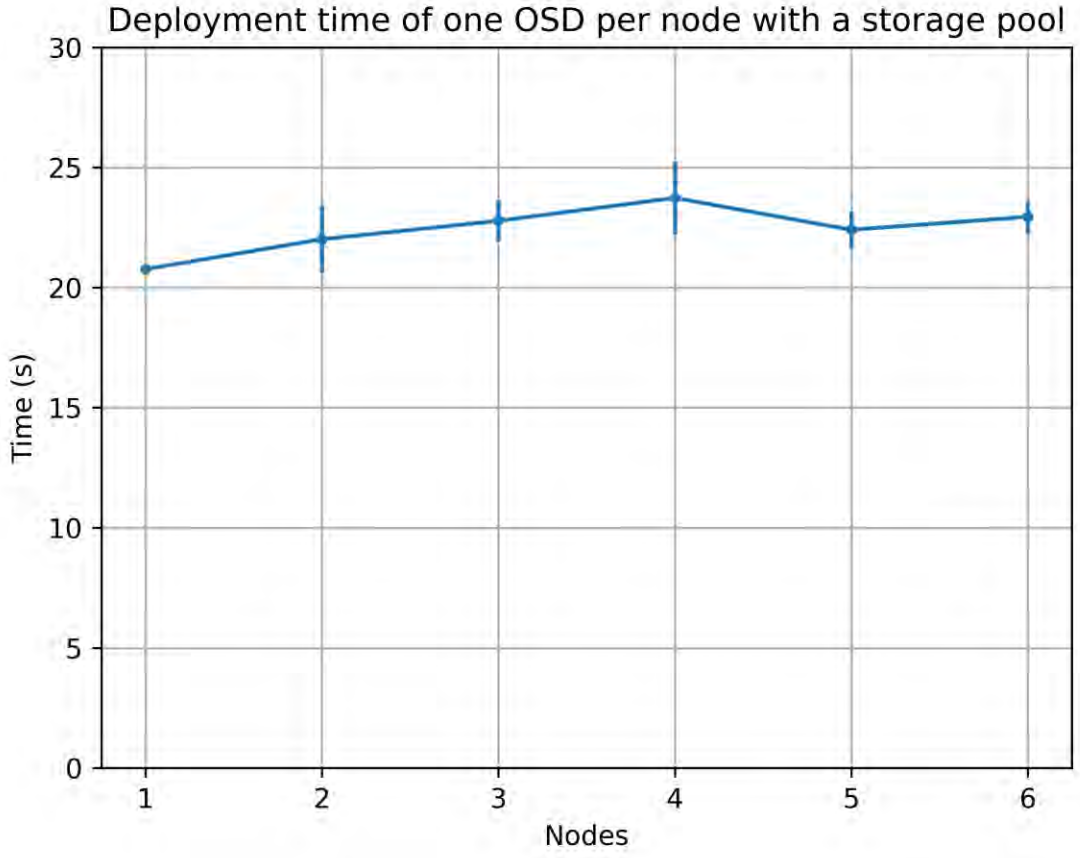
- Compute should be used for compute not setting up storage

DisTRaC Deployment



Engineering and Physical Sciences Research Council
diamond
University of BRISTOL
The Rosalind Franklin Institute

Deployment Time



Version 1 of DisTRaC, Ceph Luminous

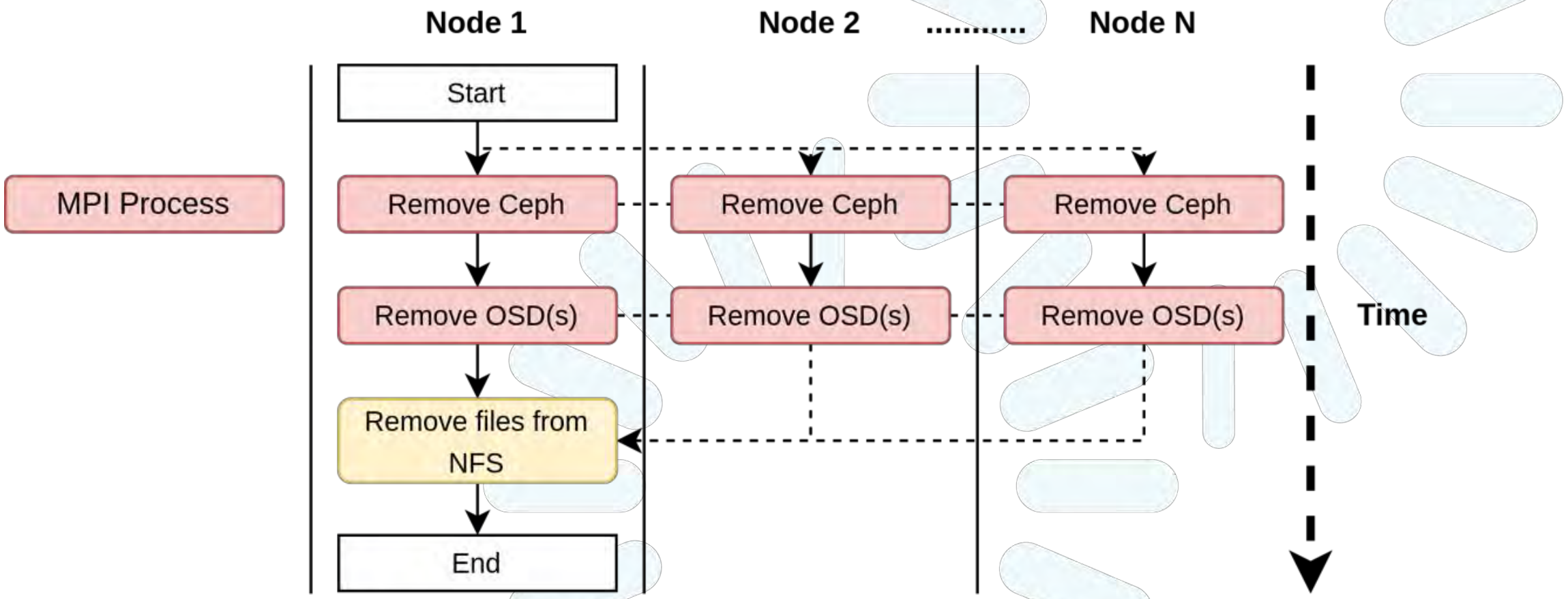
Engineering and Physical Sciences Research Council

diamond

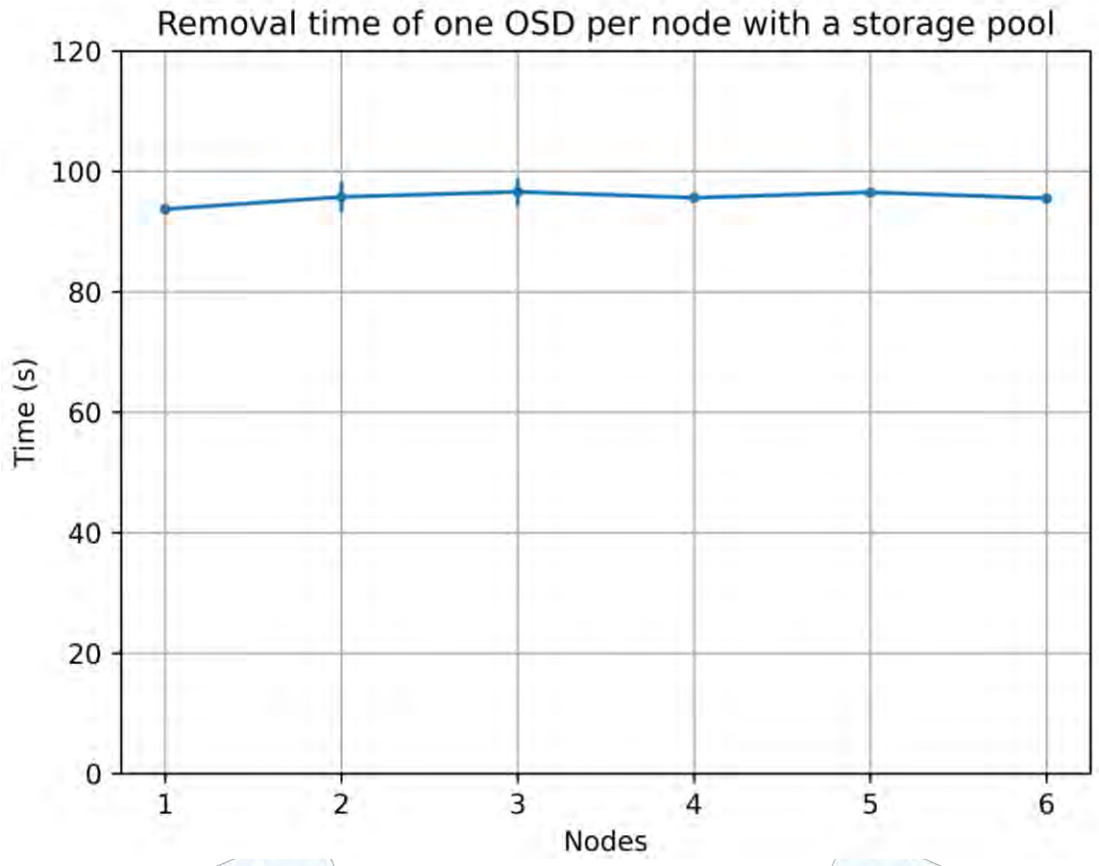
University of BRISTOL

The Rosalind Franklin Institute

DisTRaC Removal



Removal Time



Version 1 of DisTRaC, Ceph Luminous

Engineering and Physical Sciences Research Council

diamond

University of BRISTOL

The Rosalind Franklin Institute

Example Deployment Script

```
1   #!/usr/bin/env bash
2   #SBATCH --nodes=3
3   #SBATCH --ntasks-per-node=32
4   scontrol show hostnames > hostfile.txt
5   HOSTS=$PWD/hostfile.txt
6   ...
7   # Deploy DisTRaC
8   distrac.sh -i=$INTERFACE -s=$OSD_SIZE -n=$NUMBER_OF_OSDs -t=$TYPE_OF_RAM -pn=$POOL_NAME -hf=$HOSTS
9   # Run HPC Application
10  srun $HPC_Application
11  # Remove DisTRaC
12  remove-distrac.sh -t=$TYPE_OF_RAM -hf=$HOSTS
```

DisTRaC deploy and remove

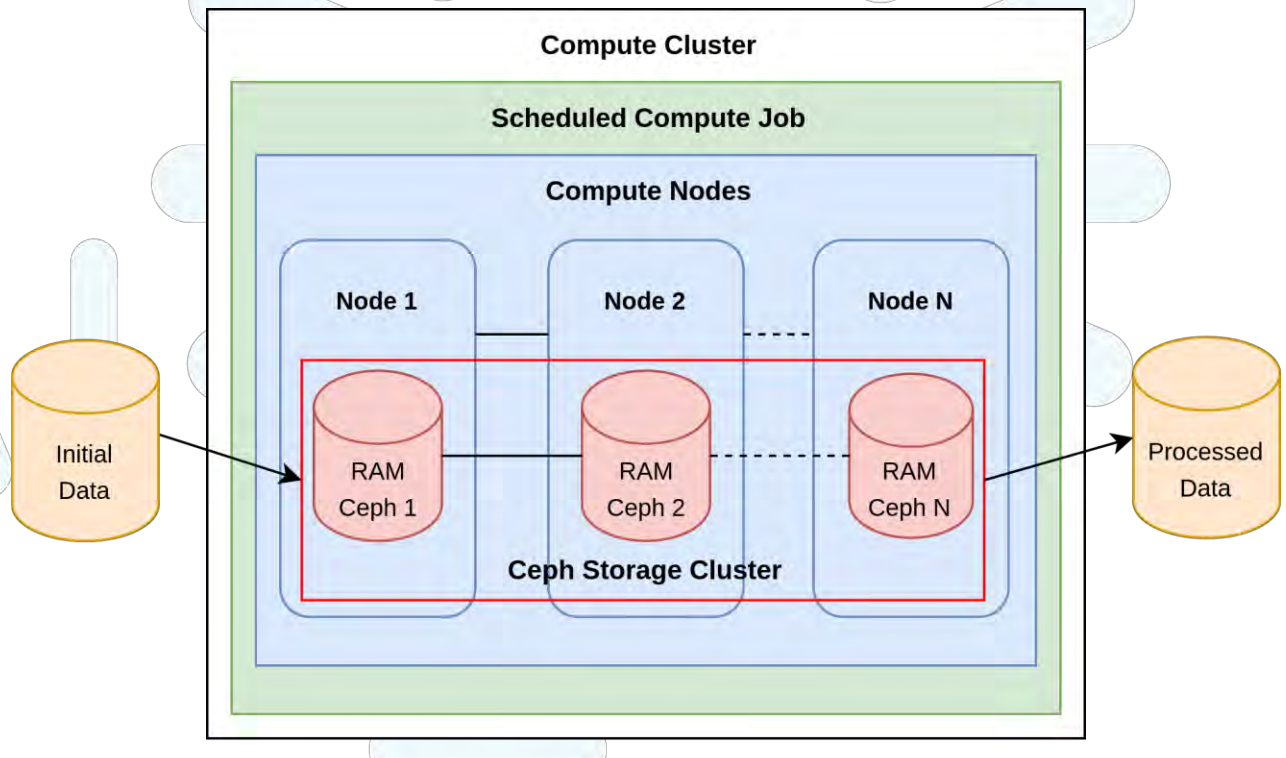
Engineering and Physical Sciences Research Council
diamond
University of BRISTOL
The Rosalind Franklin Institute

Recap

- We can create a Ceph cluster in fast and scalable way
- We can use DisTRaC deployment and removal within a job submission script
- But how does this solve the problem?

How DisTRaC solves the problem

- The IO bottleneck is now the node interconnect
- Isolated resources
- Takes pressure off HP storage
- Can remove the need for HP storage
- Reduces HPC cluster costs, especially in the cloud.
- Helps HPC facilities move towards Net-Zero



DisTRaC Deployed High Performance Compute Cluster Setup

Engineering and Physical Sciences Research Council

diamond

University of BRISTOL

The Rosalind Franklin Institute

Case Studies

- RELION [2]
- SAVU [3]

Case Study: RELION

- RELION: A cryo-microscopy structure determination program used at The Rosalind Franklin Insitute
- Compute Bound Application
- Runs using whole node cluster allocation
- Produces small intermediate files.
- Can negatively impact other users' jobs

RELION: Benchmark Setup

- Dataset provided by Cambridge[4]
- Baseline: 2xg4dn8xlarge nodes utilizing the EBS file system provided by AWS
- DisTRaC: 2xg4dn8xlarge nodes utilizing 96 Gib of RAM split into 6-16 Gib OSDS 3 on each host

RELION: Results

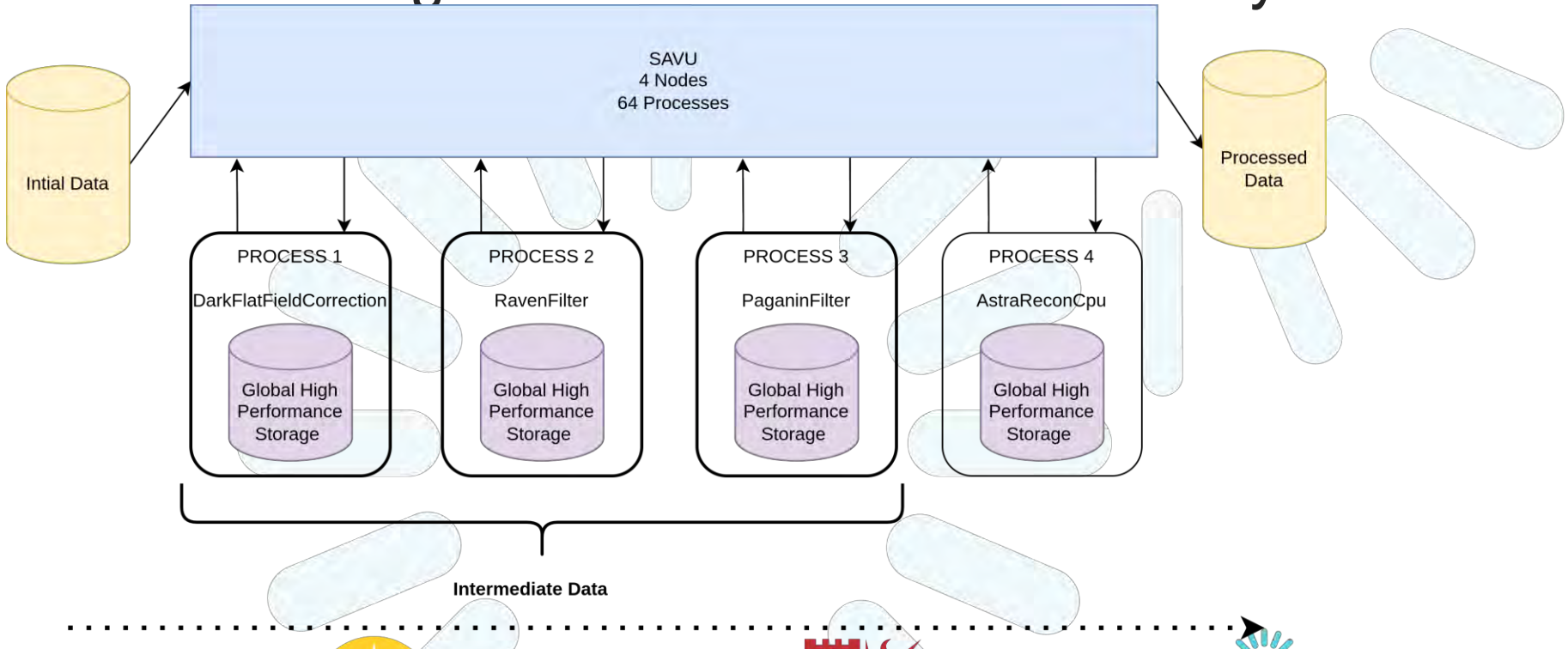
- Reduction of processing time: **5.51%**
- Total time reduction: **4.37%**
- Reduction in IO overhead: **100%**
- Removed the cost and need for running of running HP file system in the cloud
- More efficient usage of existing hardware
- Helping towards Net-Zero Goals

Case Study: SAVU

- SAVU: Tomography Reconstruction and Processing Pipeline used at Diamond Light Source and The Rosalind Franklin Insitute.
- Runs using whole node cluster allocation
- Can run at network rate
- Produces intermediate files.
- Capable of saturating access to a parallel file system
- Can negatively impact other users' jobs

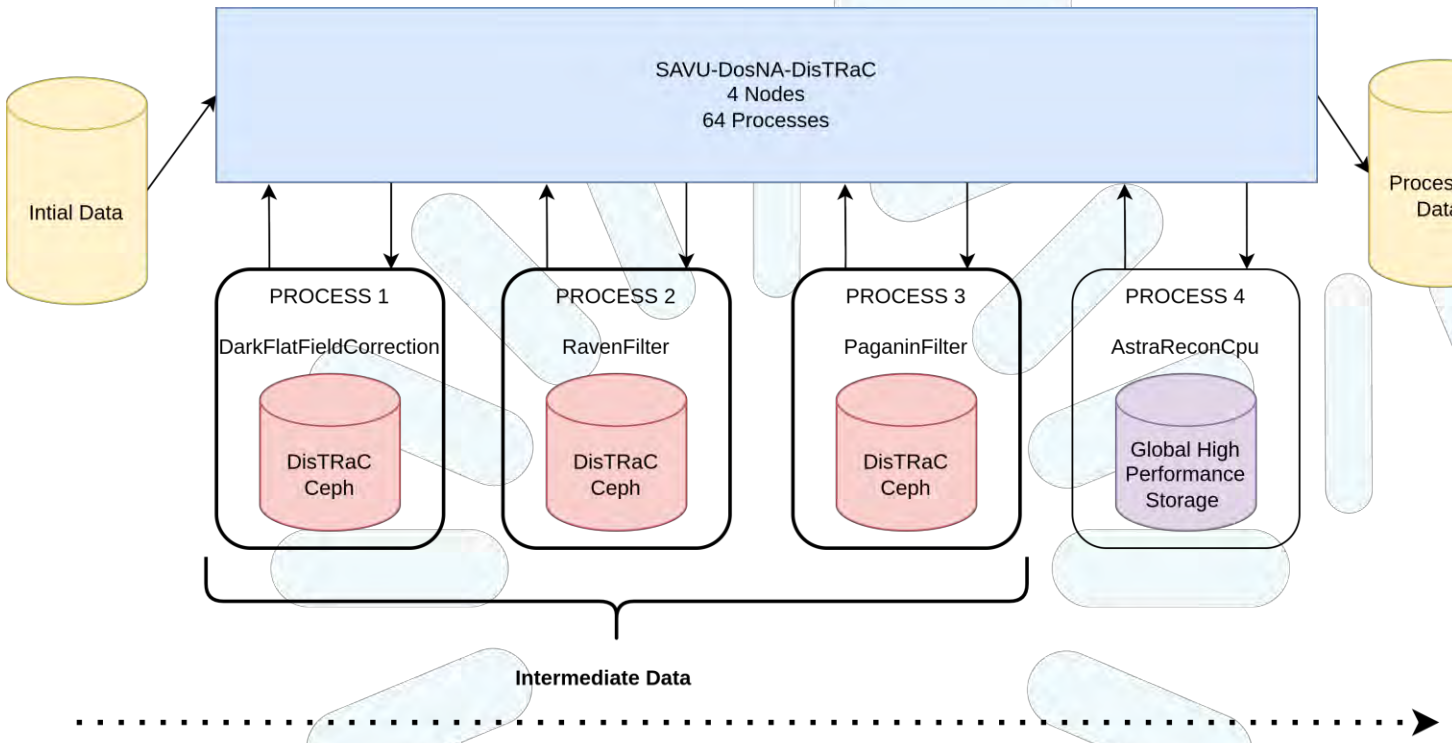
SAVU: Setup At Diamond Light Source

- Dataset: Diamond Light Source Visit NT23252 Dataset [5]
- Baseline: 4 nodes utilizing the GPFS Central HP File system



SAVU: Setup At Diamond Light Source

- DisTRaC: 4 nodes utilizing Ceph via DosNA(1)



(1) https://github.com/rosalindfranklininstitute/DosNA

Engineering and Physical Sciences Research Council

diamond

University of BRISTOL

The Rosalind Franklin Institute

SAVU: Results

- Total time reduction: **8.32%**
- Reduction in IO overhead: **81.04%**
- Reduce impact of SAVU on other users
- Prevented storing of intermediate data.
- More efficient usage of existing hardware
- Helping towards Net-Zero Goals

SAVU: Benchmark Setup At AWS

- Dataset: Diamond Light Source Visit NT23252 Dataset
- Baseline: 4 nodes utilizing the EBS AWS File system
- DisTRaC: 4 nodes utilizing Ceph Via DosNA

SAVU: Results

- Total time reduction: **67.53%**
- Reduction in IO overhead: **81.04%**
- Reduce costs of AWS
- Makes the cloud more viable for HPC
- Helping towards Net-Zero Goals

Ongoing work

- DisTRaC – Intergration into Cluster-In-The-Cloud(1)
- Adding support for Heterogenous Clusters
- Adding support for NVME deployment
- DisTRaX- removing the Ceph requirement making it extensible to other storage mechanisms.

(1) https://cluster-in-the-cloud.readthedocs.io/en/latest/

Conclusion

- DisTRaC is a Ceph deployment tool that creates a hyper-converged HPC cluster for the duration of the job by utilising the RAM of the Compute Nodes.
- DisTRaC reduces the I/O overhead of the networked filesystem and offers a potential data processing performance increase.
- Helps better utilise existing hardware to improve the performance of HPC applications
- Moves us closer to sustainable and net-zero HPC

Thank you

This project has spanned many years and has had many people involved:

Diamond Light Source - Scientific Computing:

- Dave Bond
- Mark Basham

STFC Ceph User Group:

- Tom Byrne

Rosalind Franklin Institute Artificial Intelligence theme:

- Mark Basham
- Laura Shemilt
- Joss Whittle

University of Bristol:

- Matthew Williams
- Christopher Woods

References

- [1] Weil, S.A., Brandt, S.A., Miller, E.L., Long, D.D. and Maltzahn, C., 2006, November. Ceph: A scalable, high-performance distributed file system. In *Proceedings of the 7th symposium on Operating systems design and implementation* (pp. 307-320).
- [2] Scheres, S.H., 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology*, *180*(3), pp.519-530.
- [3] Wadeson, N. and Basham, M., 2016. Savu: a Python-based, MPI framework for simultaneous processing of multiple, N-dimensional, large tomography datasets. *arXiv preprint arXiv:1610.08015*.
- [4] Wong, W., Bai, X.C., Brown, A., Fernandez, I.S., Hanssen, E., Condron, M., Tan, Y.H., Baum, J. and Scheres, S.H., 2014. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife*, *3*, p.e03080.
- [5] Mark Basham, Nghia Vo, Avery Pennington, Win Tun, Olly King, & Gabryel Mason-Williams. (2020). Diamond Light Source Visit NT23252 Dataset [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4030687

UKRI Engineering and Physical Sciences Research Council

diamond

University of BRISTOL

The Rosalind Franklin Institute

Thank you for listening

Questions?

Simon Atack (University of Bristol)

Creating A Cluster - Going it Alone

Abstract: In 2019 Bristol University went live with its homemade cluster. This is our experience of the trials and tribulations encountered during this endeavour from a sysadmin perspective. I will discuss the technical aspects and design considerations (including networking, deployment, scheduling [software stack] etc). And will finish with our reflections on this experience.

Bio: Simon is the HPC Team Leader at the Advanced Computing Research Center(ACRC), University of Bristol for the last 5 years, and has been with the center for the last 8 years. Involved in a wide variety of technical areas of HPC, sysadmin, networking and storage. Previously employed at the University of Nottingham for many years in a variety of roles, including software development, HPC, software licencing, system deployment, user support etc.

CREATING A CLUSTER
-
GOING IT ALONE

Simon Atack
University of Bristol
Simon.Atack@Bristol.Ac.Uk

OVERVIEW

In 2019 Bristol University went live with its homemade cluster. This is our experience of the trials and tribulations encountered during this endeavor from a sysadmin perspective. I will discuss the technical aspects and design considerations (including networking, deployment, scheduling [software stack] etc). And will finish with our reflections on this experience

BACKGROUND

Vision of the system

- A system that can be expanded (that could replace small standalone individual clusters
- researchers can buy into the cluster
- researchers could buy at every new cluster eg 3 or 4 years
- Central irregular top ups
- Add in new equipment test things

INITIALLY

- Initially dev clusters created from scrap equipment
- Initial eval hardware purchased

CLUSTER STAGES

In order to satisfy end users

- stopped being a development
- became testing
- ultimately sneaking into production before

continuing to develop & fine tune while a production system

MGMT SYSTEMS

- we went from classical HA systems
- to HV VMs doing HA

PROVISIONING/DEPLOYMENT

- foreman - as previously in system (non HPC)
- self developed wrapper round templating - as a quick lightweight solution

NETWORKING

- not aimed for low latency/high performance
- Initially as a 1G
- investigation of higher speeds
- settled on higher 100G/25G backbone
- 10G to nodes

- Previously Cumulus OS – Do we go Dell OS/Sonic?

QUEUING

- PBS was used to start (due to experience)
- switch to slurm to consolidate

STORAGE

- Initially basic NFS storage
- multiple namespaces
- moved to over to GPFS multicluster
- aimed as a storage for HPC rather than a storage for a cluster

SOFTWARE

- considering software management
- easybuild considered - due to time constraints couldnt go that route

JACK OF ALL TRADES

- We became on top hpc sysadmin
- Devops
- Developers
- Network Engineers
- continual racking & settingup

ISSUES OCCURED

- MGMT Stack wasnt stable reboot due to deployment shortcuts in building (chicken/egg issue)

OUTCOMES

- Upskilled team in understanding of clusters, how connected
- Gained working knowledge
- Single points of failure in experience/knowledge -- due to deadline
- Buddy development Work
- Documentation fell behind
- Minimum Viable product - missing features
- increased usage of VMs
- Investigating alternative finance structure for 'selling' resources ot machines

LESSONS LEARNT

- We are not developers! (skills)
- Have a development system support structure
- This takes time (much more that you expect)
- Success has depended on goodwill of staff going above and beyond
- System is as you want it (you choose what you want)
- Youve taken on the full support -- buck stops here.
- Scaling - lots of work for one institution to support

THANKS TO ACRC HPC TEAM

- Callum Wright
- Amaurys Avila Ibarra
- Isaac Prior
- Tom Batstone
- Dianaimh Green
- Ethan Williams

QUESTIONS

Ed Threlfall (UKAEA)

Project NEPTUNE - sustainable software for sustainable fusion energy

Abstract: Modelling the edge region of magnetically-confined nuclear fusion plasmas brings many computational challenges including plasma turbulence and the significant kinetic effects. Project NEPTUNE, part of the UK's ExCALIBUR programme, aims to apply modern numerical methods and software engineering techniques to address these issues using current and next-generation (i.e. exascale) HPC. One key theme is the application of spectral / hp finite element methods, which, due to their intrinsically large number of arithmetic operations per unit of data, are well-suited to today's exascale architectures - this contrasts with existing finite-difference codes, which are not expected to scale well. A second key theme is the use of particle methods for matter that is out of thermal equilibrium owing to low collisionality.

This presentation will outline applications of spectral / hp methods and particle methods to fusion-relevant problems, including heat transport and the dynamics of charged and neutral particles, intended as the initial steps toward a full-featured plasma edge simulation code. These applications serve to illustrate the NEPTUNE project's aims for software sustainability (software capable of significant evolution during a 30-year lifecycle) and embody an approach to multi-site opensource software development encompassing `separation of concerns' and `co-design'.

Bio: Ed Threlfall holds a degree and a Masters in physics from Cambridge University and a PhD in theoretical particle physics from the University of Southampton. Following a postdoctoral position at Durham University, he joined a software company to pursue the development of advanced physics simulation codes and wrote a commercially successful finite-element simulation engine for solving problems in photonics and nonlinear optics, using C++.

After ten years' experience in industry, he joined UKAEA in 2020, in order to explore the challenge of numerical fusion simulations. Ed has extensive experience with implementing efficient finite-element codes, particularly Discontinuous Galerkin methods.

PROJECT NEPTUNE

SUSTAINABLE SOFTWARE FOR SUSTAINABLE FUSION ENERGY

Ed Threlfall, UKAEA (on behalf of the NEPTUNE team)

CIUK 2022, Manchester, 1st December 2022

Met Office

UK Research and Innovation

UK Atomic Energy Authority

NEPTUNE – why? (1)

ITER (International Thermonuclear Experimental Reactor)

""We say that we will put the sun in a box. The idea is pretty. The problem is, we don't know how to make the box."

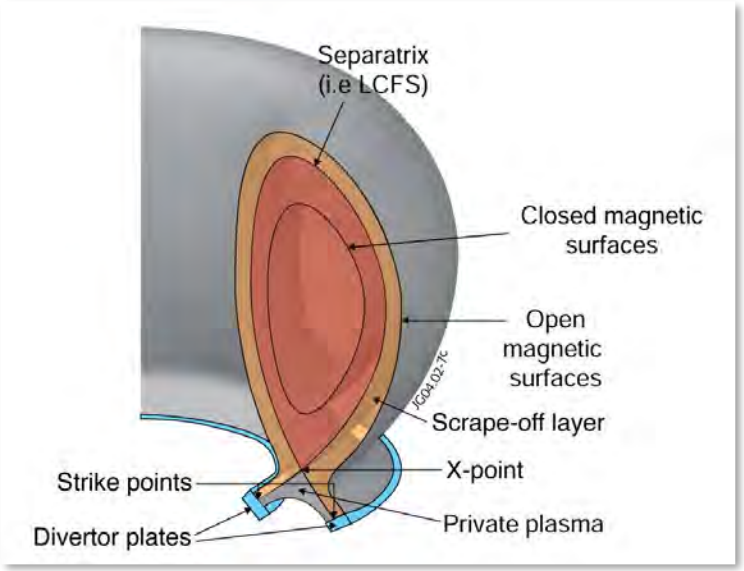
Pierre-Gilles de Gennes

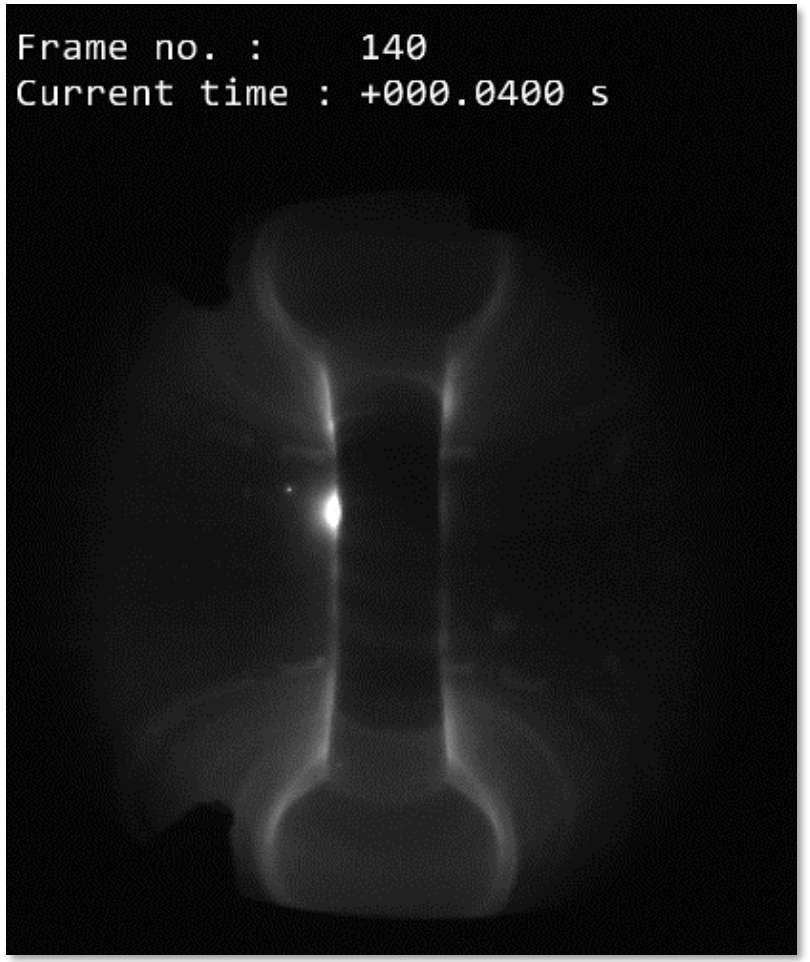
- Fusion is a potential sustainable energy solution.
- Magnetically-confined approach – 100 million degrees plasma.
- Contained in a tokamak - a toroidal `box'.
- Plasma touches side – heat loads ~ spacecraft re-entry …

NEPTUNE – why? (2)

Modelling the plasma `edge' or exhaust:

- A long-established exascale grand-challenge multi-physics, multi-scale problem.
- Complexity: turbulence (hard – outstanding Millennium Prize!), many species, atomic physics, etc.
- Kinetic effects: out-of-thermal equilibrium matter (few collisions), requires coupled fluid and particles.





~1s long H-mode MAST-U pulse

NEPTUNE – what?

Why digital models?	Tokamaks expensive! (ITER ~€20bn, initiated 1988).
Existing software limitations	Finite-difference methods, not best suited to new HPC architectures.

NEutrals and Plasma TUrbulence Numerics for the Exascale

- *'Once-in-a-generation' opportunity for UK to improve fusion plasma physics software processes in order to develop 'actionable' code i.e. code suitable for incorporation in an engineering design workflow as required for STEP.*
  (Wayne Arter, Technical Lead.)
- Interdisciplinary 'rainbow' team; 60-70% of work defrayed to UKRI.
- Proxyapps development philosophy (each solves part of the problem).
- Sustainability – re-usable, reliable, efficient, scalable.

STEP (Spherical Tokamak for Energy Production) – fusion power on UK grid by 2040s.

NEPTUNE – how?

- Fluid physics module - C++ finite element library, modelling plasma turbulence.
- Particles module – C++ library for out-of-equilibrium matter.
- Uncertainty quantification for actionable code – non-intrusive.
- Sub-modules interchangeable with reduced-order models / surrogates.

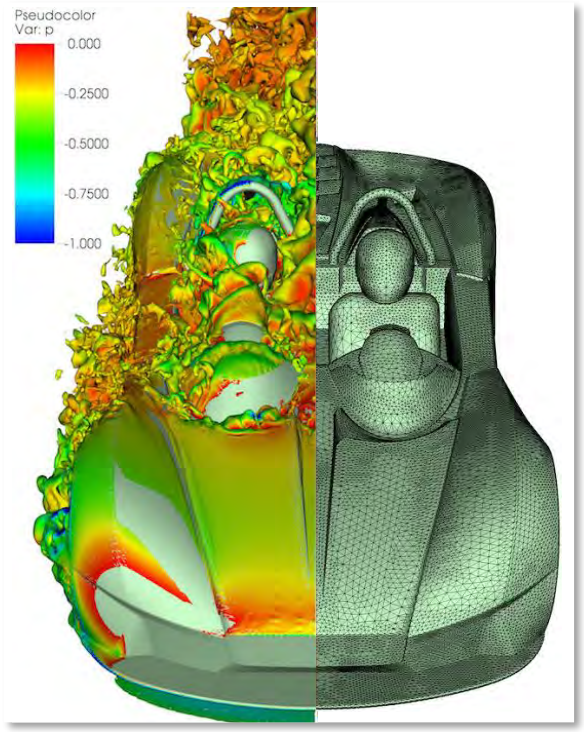
Use	• Engineers, plasma physicists, software developers. • Run on anything from single laptop to exascale. • Aim for 30-year life-cycle.
Domain-specific language	DSL for user interaction - compose library components.
Languages	C++17 / SYCL 2020 for main code, Python for user DSL.

Nektar++ : spectral / hp elements

Nektar++ spectral / hp code [1] for solving plasma fluid equations.

- Arbitrary convergence order p – error $\propto h^p$ (h~element size).
- Arithmetic intensity – increased number of operations on same data counters HPC data movement bottleneck.
- Supports complicated geometries, curved elements.

Structure	• Set of libraries. • C++ code with MPI parallelism for CPUs. • Refactoring for performance portability / GPUs / C++17.
Provenance	• Proven scaling to c.100k cores. • Well-tested code. • Established community of developers / users.
Benefit	*More rapid convergence means more energy-efficient calculations.*



CFD simulation of Elemental RP1 track car.

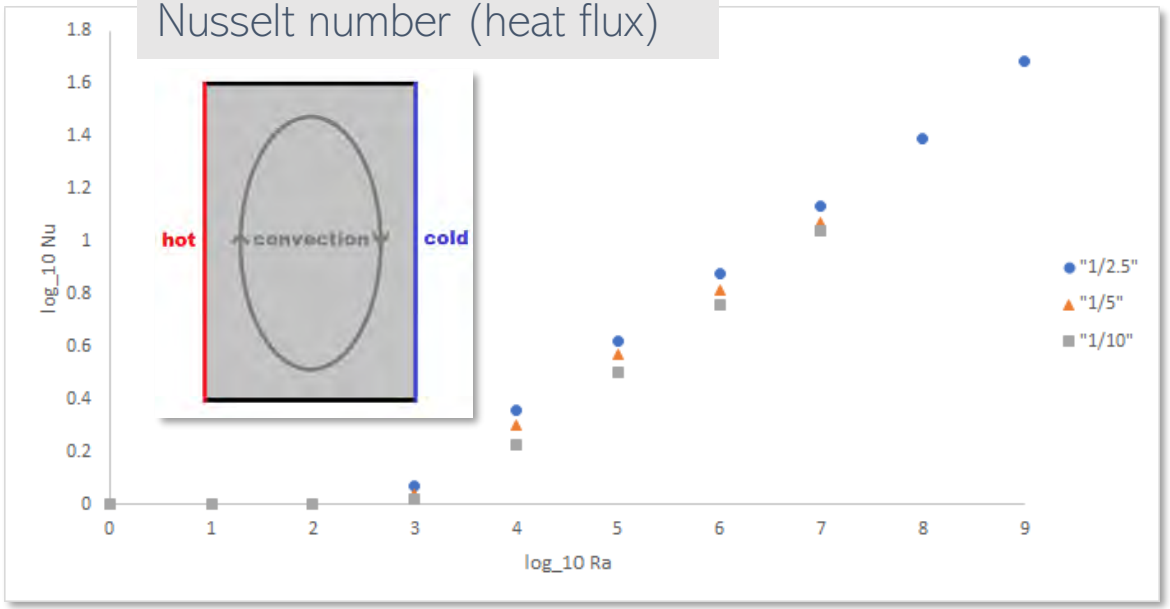
1. https://www.nektar.info

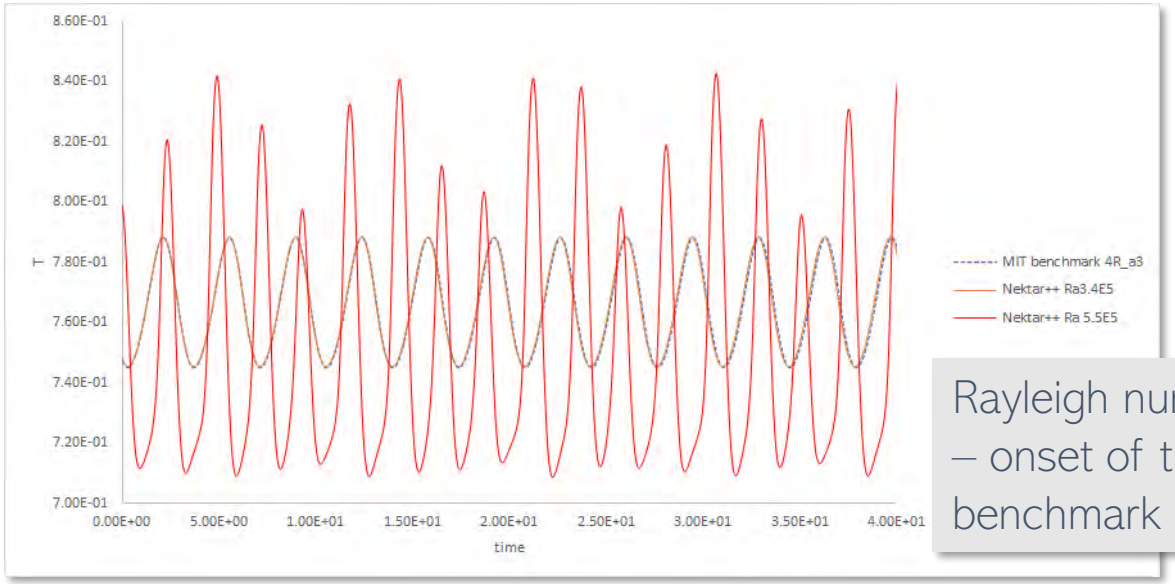
D. Moxey (King's College London); C.D. Cantwell, S.J. Sherwin (Imperial College London)

Nektar++ convection proxyapp

Temperature field - Rayleigh number (Ra) small

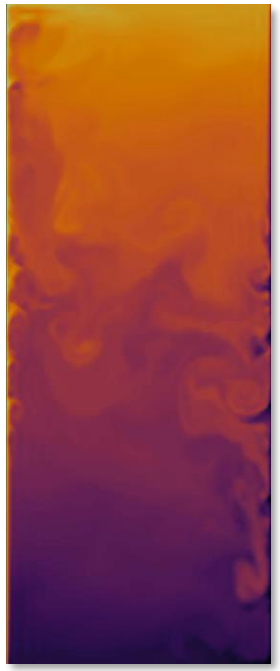
Nusselt number (heat flux)





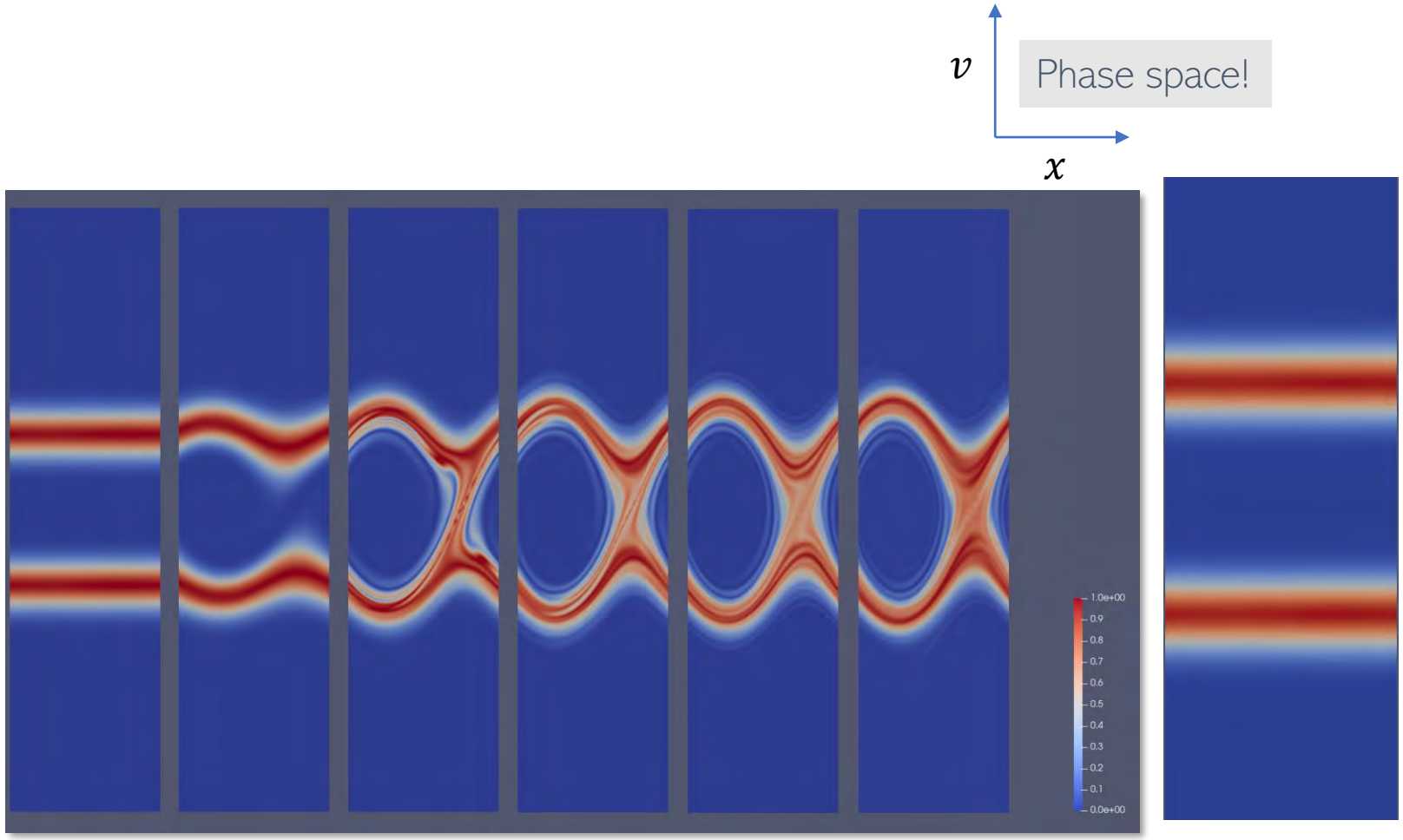
Rayleigh number intermediate – onset of turbulence (MIT benchmark [1])

1. https://www.mathematik.tu-dortmund.de/~featflow/en/benchmarks/cfdbenchmarking/mit_benchmark.html

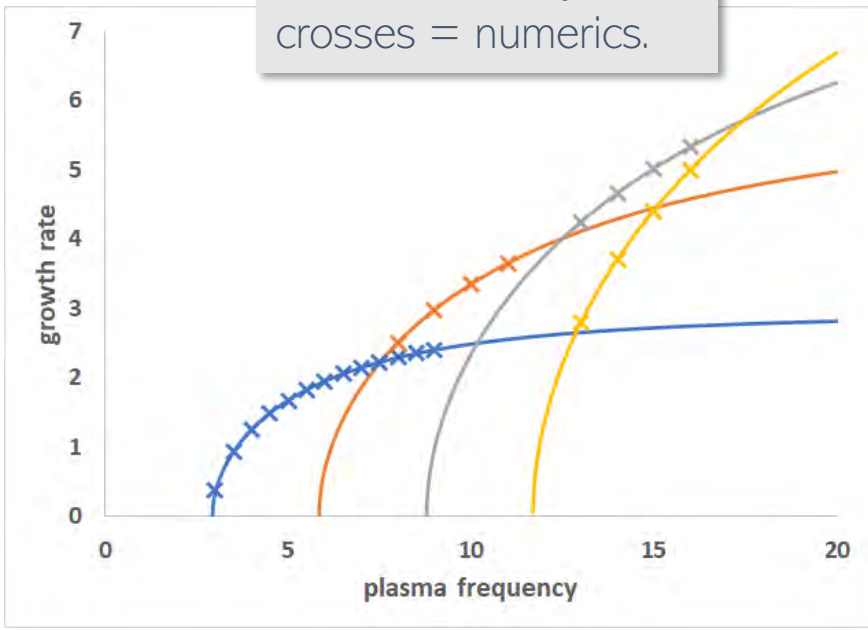


Temperature field - Rayleigh number large

Nektar++ kinetic proxyapp



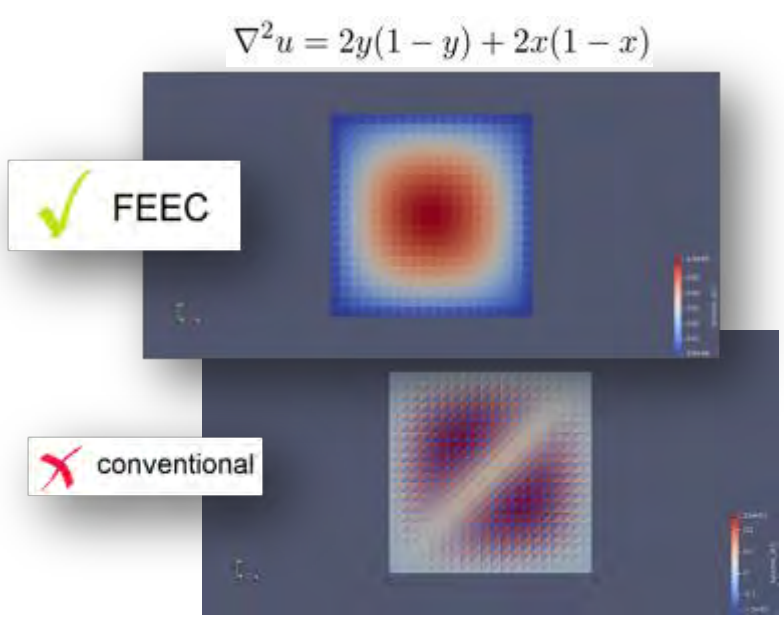
Instability growth rates for $k_n = \frac{2\pi n}{L}$
Curves = theory,
crosses = numerics.



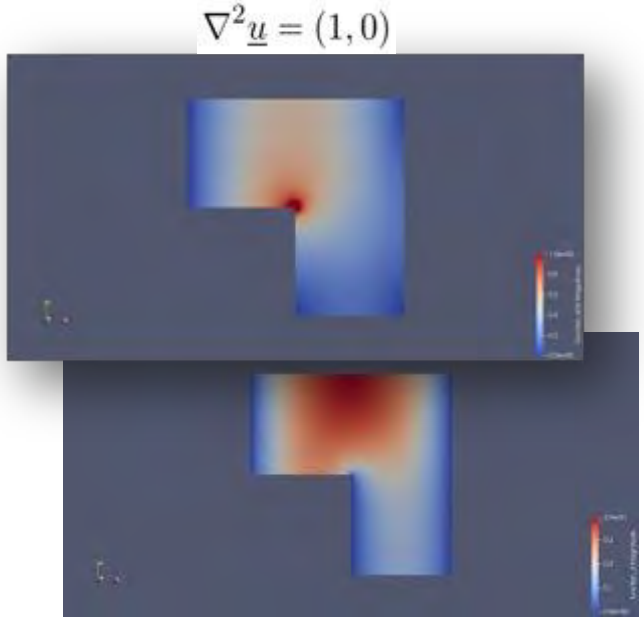
Two-stream instability – nonlinear evolution of initial Gaussian charged particle beams

Advanced techniques can guarantee numerical stability

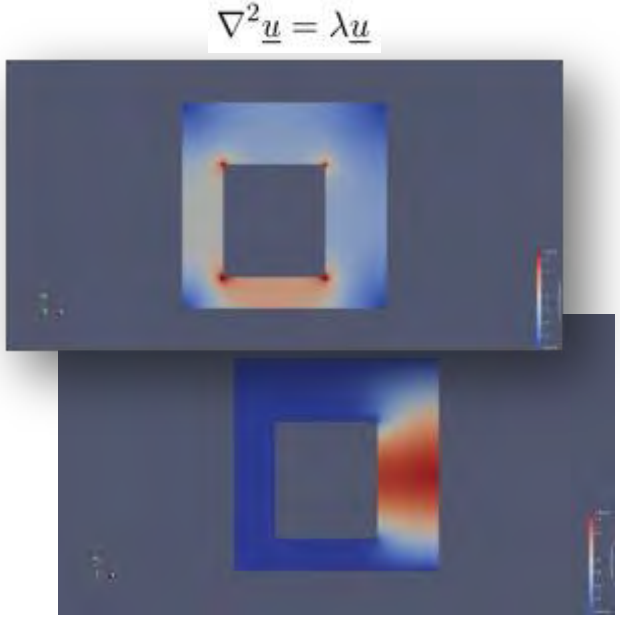
Finite element exterior calculus (FEEC)	Properties of PDE can survive discretization – conservation laws, stability, which elements to use …



Unstable numerics



Re-entrant corner field singularity



Non-trivial topology

Challenge	Implement in *Nektar++* without loss in performance …
References	See textbook by D.N. Arnold [1], Firedrake ([2]) implementations of Arnold's examples [3].

1. D.N. Arnold, *Finite Element Exterior Calculus*, CBMS-NSF regional conference series in applied mathematics **93**, SIAM.
2. https://www.firedrakeproject.org
3. https://github.com/ethrelfall/Finite-element-exterior-calculus

Advanced techniques can improve convergence rates

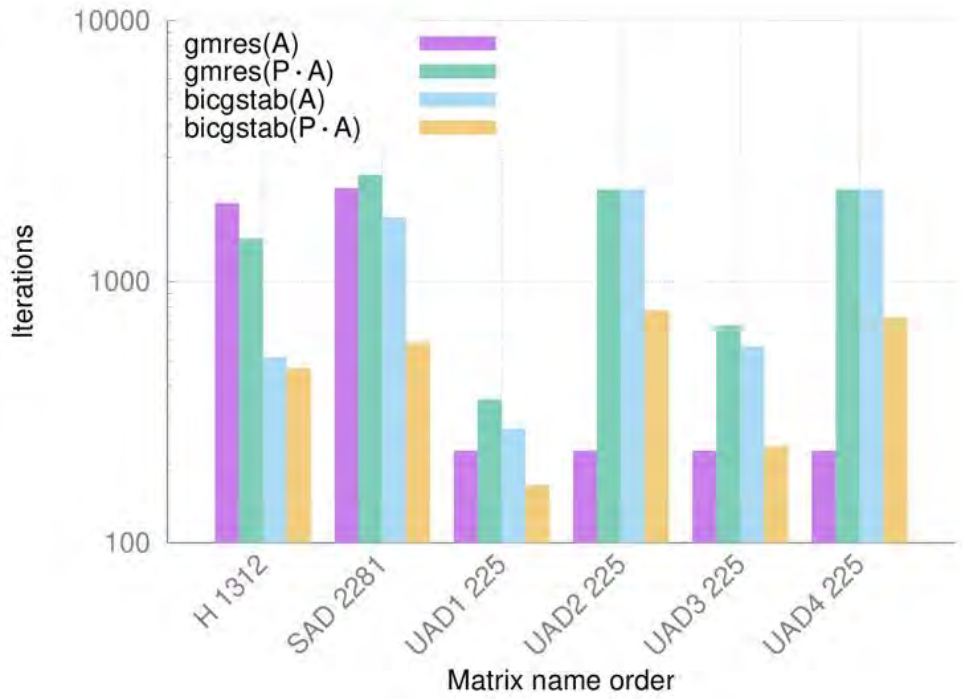
Preconditioners	Improve efficiency of $Ax = b$ matrix inversion by applying approx. solution before iterating.

Operator-based preconditioner

	N_y	None	P_0	P_1
RHS Evals	200	1466071	92682	56630
Time (s)	200	1925	151	95
RHS Evals	400	-	252708	169830
Time (s)	400	-	615	413

BOUT++ / *SD1D* plasma-neutrals simulation: ~37% reduction in runtime over original preconditioner.

Markov-chain Monte Carlo preconditioner

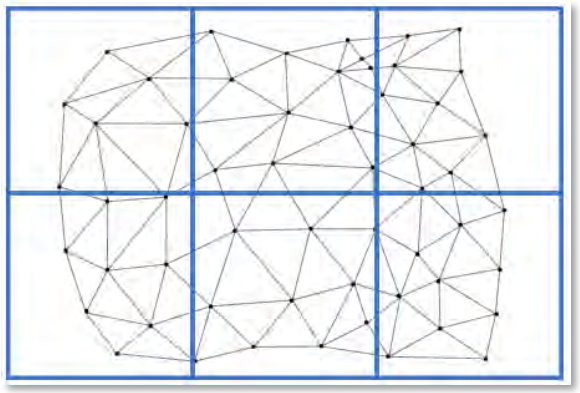


Nektar++ cases: order of magnitude speed-up.

S. Thorne, STFC Hartree Centre

DSL - enabled, performance-portable MPI particle library

NESO-PARTICLES [1] (NESO = NEPTUNE Experimental SOftware)	• UKAEA library for particle data and moving particles between MPI ranks on unstructured meshes. • Particle-mesh interface abstract - different mesh implementations possible (including *Nektar++* meshes). • Charged and neutral particles. • Header-only library.
Coupling	Initial implementations tightly-coupled to the finite element code - coupling to surrogate models anticipated.
Dependencies	• CMake • SYCL 2020 (tested with hipSYCL 0.9.2 and Intel DPCPP 2022.1.0) • MPI 3.0 (tested with MPICH 4.0 and IntelMPI 2021.6) • HDF5 (optional, if particle trajectories required)

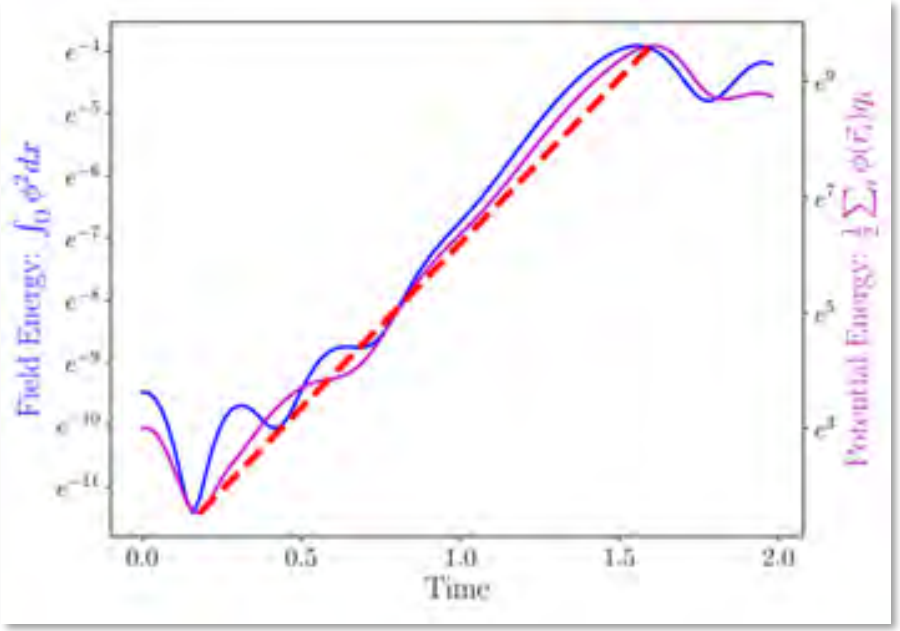


oneAPI hipSYCL

1. https://github.com/ExCALIBUR-NEPTUNE/NESO-Particles

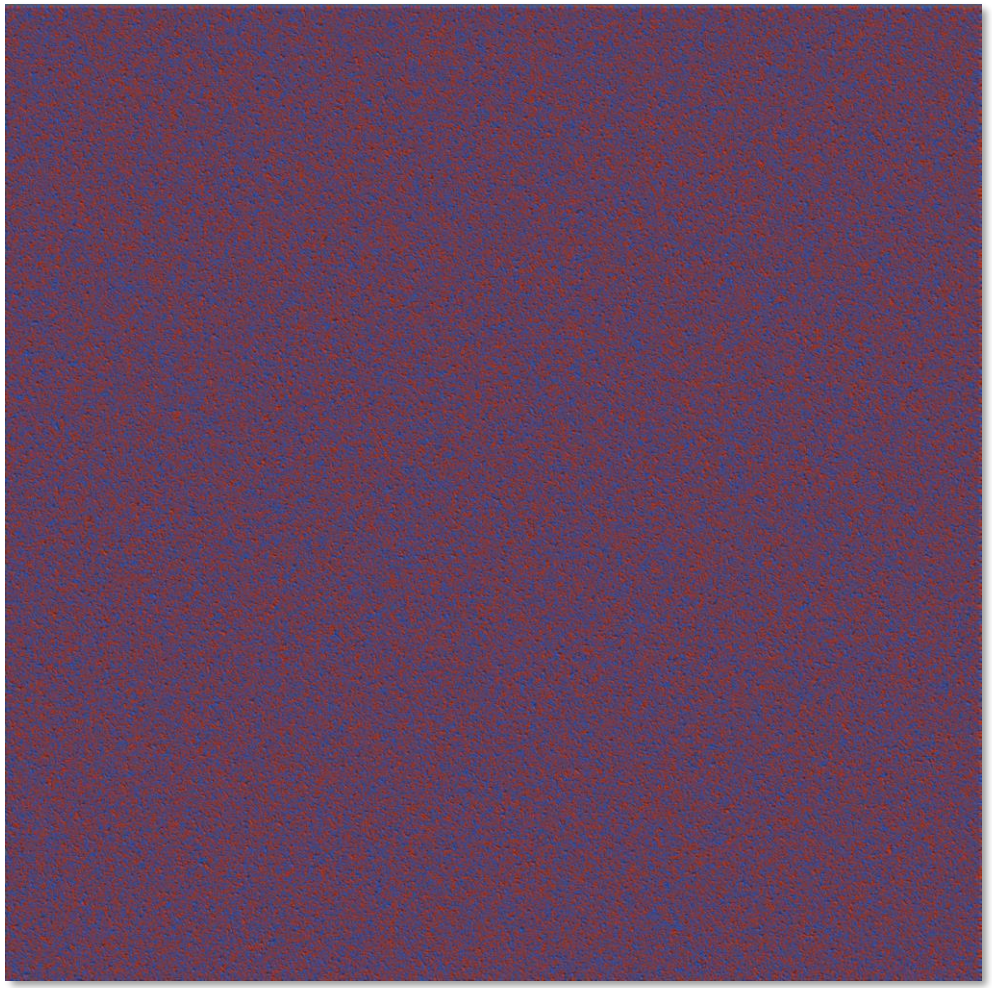
Particles proxyapps

NESO [1]	• Test implementations integrating particle capabilities and FEM. • Can be built using Spack package manager. • 2D2V electrostatic particle-in-cell solver. • Nektar++ provides Poisson solve.
Tests	• Linear growth rates of unstable modes. • Energy conservation.



Instability growth rate vs theory

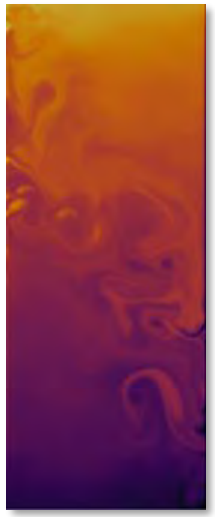
Time evolution of 512k interacting particles



1. https://github.com/ExCALIBUR-NEPTUNE/NESO

Actionable code requires Uncertainty Quantification

VVUQ	• Meaningful bounds on error in code outputs given statistical uncertainty in inputs. • Ensemble-based execution patterns. • Non-intrusive UQ – separation of concerns.
SEAVEA project synergy, SEAVEA toolkit [1]	• UQ campaigns on HPC. • Construction of surrogate models e.g. Gaussian process models, machine-learning. • Modern data science techniques e.g. data assimilation, Bayesian analysis.
FabNEPTUNE [2]	• FabSim3 plugin created specifically for NEPTUNE. • Easy execution of NEPTUNE simulations, integrated with SEAVEA toolkit. • Currently drives *Nektar++* 2D and 3D convection proxyapps - but more specific plasma applications very soon.



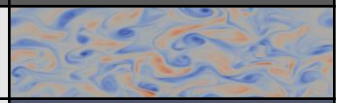
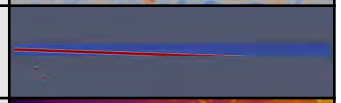
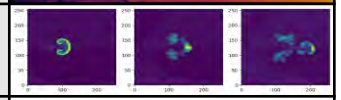
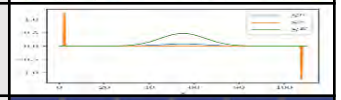
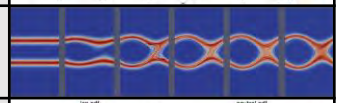
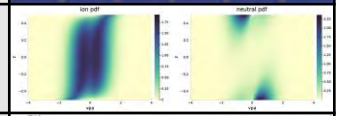
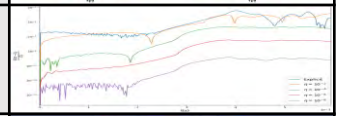
Nektar++ convection proxyapp + VVUQ to support *Smallab* experiments (Arter, Buta – Univ. Leeds)

FabNEPTUNE

1. https://www.seavea-project.org/seaveatk
2. https://github.com/UCL-CCS/FabNEPTUNE

S. Guillas, P.V. Coveney, K. Bronik (UCL)

Proxyapps inventory

Proxyapp	Framework	Language	Comments	Sample output
nektar-driftwave	*Nektar++*	C++	2D Hasegawa-Wakatani equations	
nektar-diffusion	*Nektar++*	C++	strongly anisotropic diffusion	
vertical natural convection in spectral / hp, 2D and 3D	*Nektar++*	C++	incompressible Navier-Stokes with buoyancy	
2D plasma turbulence equations in spectral / hp	*Nektar++*	C++	*Hermes-3* equation system	
1D fluid solver with UQ and realistic boundary conditions	*Nektar++*	C++	1D model of scrape-off layer	
Vlasov-Poisson kinetic solver in spectral / hp	*Nektar++*	C++	due Dec 2022	
moment-kinetics	new code (Univ. Oxford)	Julia	moment-kinetic gyro-averaged code	
minepoch	*EPOCH* (Univ. Warwick)	Fortran	used for testing particle implementations	
electrostatic PIC proxyapp	NESO-Particles	C++ / SYCL	due Dec 2022	
2D3V coupled fluids-neutral particles proxyapp	NESO-Particles	C++ / SYCL	due Mar 2023	coming soon

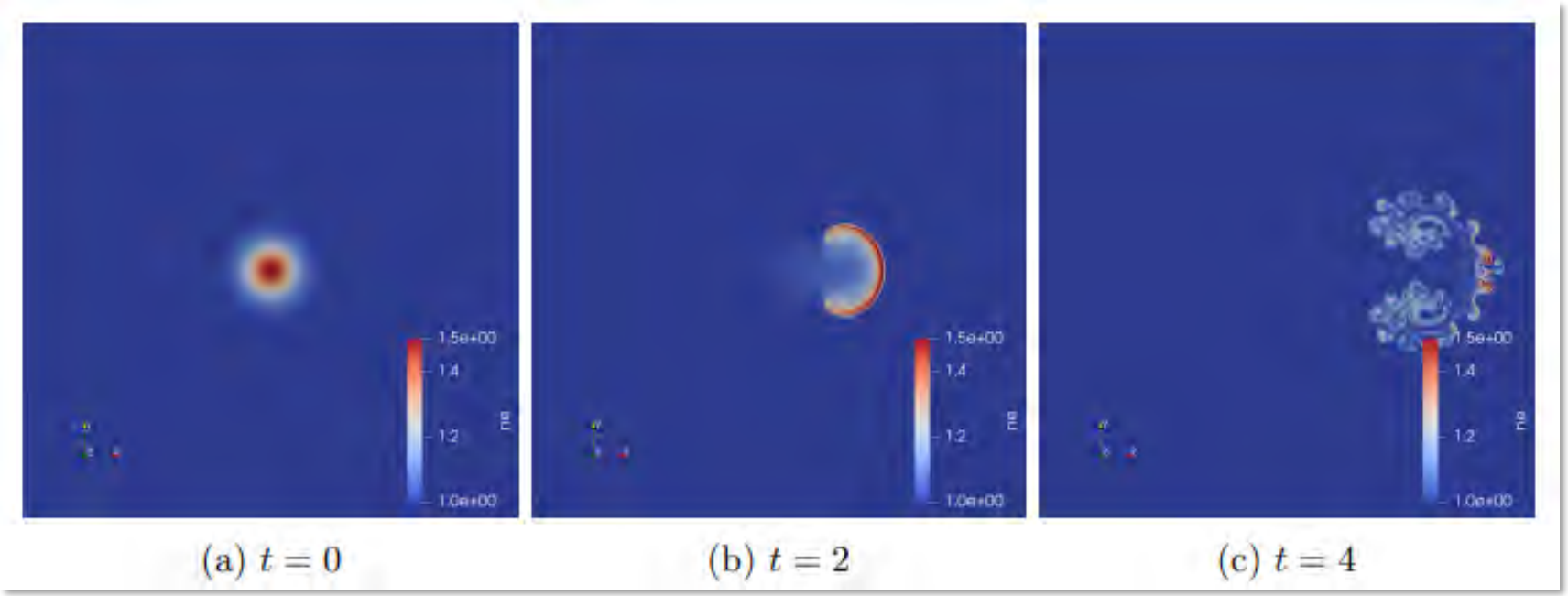
Community overview

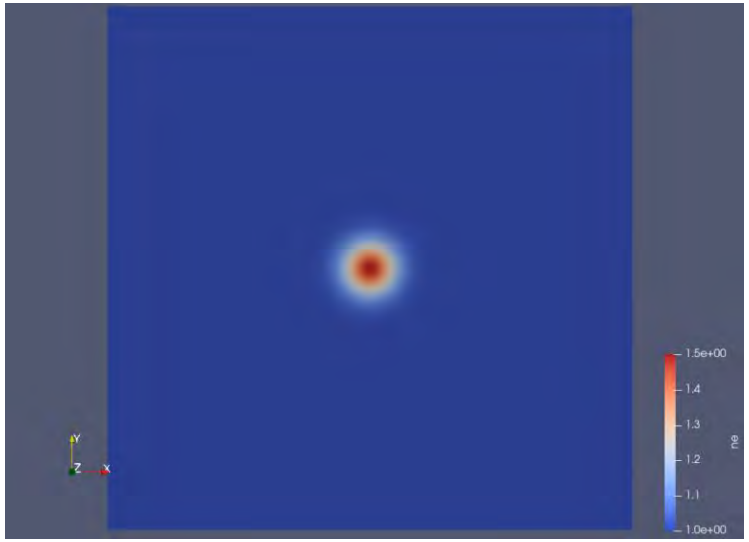
UKAEA TEAM	Rob Akers, Wayne Arter, Matthew Barton, James Cook, John Omotani, Joseph Parker, Owen Parry, Will Saunders, Ed Threlfall.
UKRI GRANTS	• University of Exeter (VVUQ, surrogate models): Peter Challenor, Tim Dodwell, Louise Kimpton. • King's College London (Nektar++): Mashy Green, David Moxey. • Imperial College London (Nektar++): Chris Cantwell, Bin Liu, Spencer Sherwin. • University of Oxford: Michael Barnes, Patrick Farrell, Michael Hardman. • STFC Hartree Centre: Vasil Alexandrov, Hussam al-Daas, Tyrone Rees, Emre Sahin, Andrew Sunderland, Sue Thorne. • University College London (VVUQ): Kevin Bronik, Peter Coveney, Matt Graham, Serge Guillas, Tuomas Koskela, Yiming Yang. • University of Warwick (DSLs): Gihan Mudalige. • University of York (plasma physics, support & coordination, DSLs): David Dickinson, Ed Higgins, Chris Ridgers, Steven Wright.
ALUMNI	• University of Oxford: Felix Parra-Diaz. • University of Warwick (EPOCH): Ben McMillan, Tom Goffrey. • University of York: Ben Dudson.
OUTPUT (INC. CODE)	• Proxyapps code (MIT licence): see repositories on https://github.com/ExCALIBUR-NEPTUNE (some, inc. NESO and NESO-Particles, are public). • Large body of supporting documents and reports – https://github.com/ExCALIBUR-NEPTUNE/Documents (currently private). • Developer website in development.

Participation welcomed!

Next steps: 2D3V plasma proxyapp

2D plasma turbulence with neutral particle source terms	• Tight-coupled integration of the spectral / hp and particles. • Kinetic neutral species in plasma background. • Due by end Mar 2023.
Plasma turbulence in *Nektar++*	*Nektar++* [1] implementation of equations from existing *Hermes-3* code (finite difference) [2].
Neutral particles	Neutral particles do not feel confining magnetic field, but ionize as they interact with plasma – source terms in fluid equations (= coupling).





1. https://github.com/ExCALIBUR-NEPTUNE/nektar-driftplane
2. https://github.com/bendudson/hermes-3

Joseph Hickson, Lewis Sampson and Victoria Smart (Met Office)

Preparing the Met Office for the next generation of supercomputers

Abstract: As part of ExCALIBUR funded projects at the Met Office there have been projects exploring multiple approaches towards a sustainable computing future, from general compute optimisations to GPU studies, all to enable Exascale weather & climate modelling. In this talk we hear about the results of some of these technical developments and the experiences of two early career software engineers who have been developing the skills needed to work on these projects.

Bios: Having spent most of his working career developing cloud-based business software, Joe joined the Met Office in 2019 as a Scientific Software Engineer working on developing the next generation modelling system. Since then, he's worked on the ubiquitous Covid Modelling projects before moving into managing the ExCALIBUR Pool of Scientific Software Engineers in 2021.

Lewis started higher education at the University of Plymouth, completing a PhD in "Ocean modelling with novel data assimilation techniques" under Professor Georgy Shapiro. Lewis joined the Met Office as a foundation scientific software engineer during lockdown in 2020, working for the Ocean Forecasting Research and Development team, while still being part of the ExCALIBUR Pool. His work focuses on the analysis of the WaveWatch III model, using the options surrounding parallel processing such as OpenMP, OpenACC, MPI, and hybrid implementations.

Victoria joined the Met Office in 2020 as a member of the ExCALIBUR pool. Since joining, Victoria has been deployed into multiple teams, previously working on improving the testing infrastructure for the IMPROVER post-processing code and investigating the method of launching coupled models on the supercomputer. Now Victoria is working on porting the ocean data assimilation code, NEMOVAR, to GPU.

PREPARING THE MET OFFICE FOR THE NEXT GENERATION OF SUPERCOMPUTERS

Abram

Joe ~~Hickson~~
[he/him/his]
joseph.hickson@metoffice.gov.uk

Victoria Smart
[she/her/hers]
victoria.smart@metoffice.gov.uk

Lewis Sampson
[he/him/his]
lewis.sampson@metoffice.gov.uk

Weather & Climate Modelling

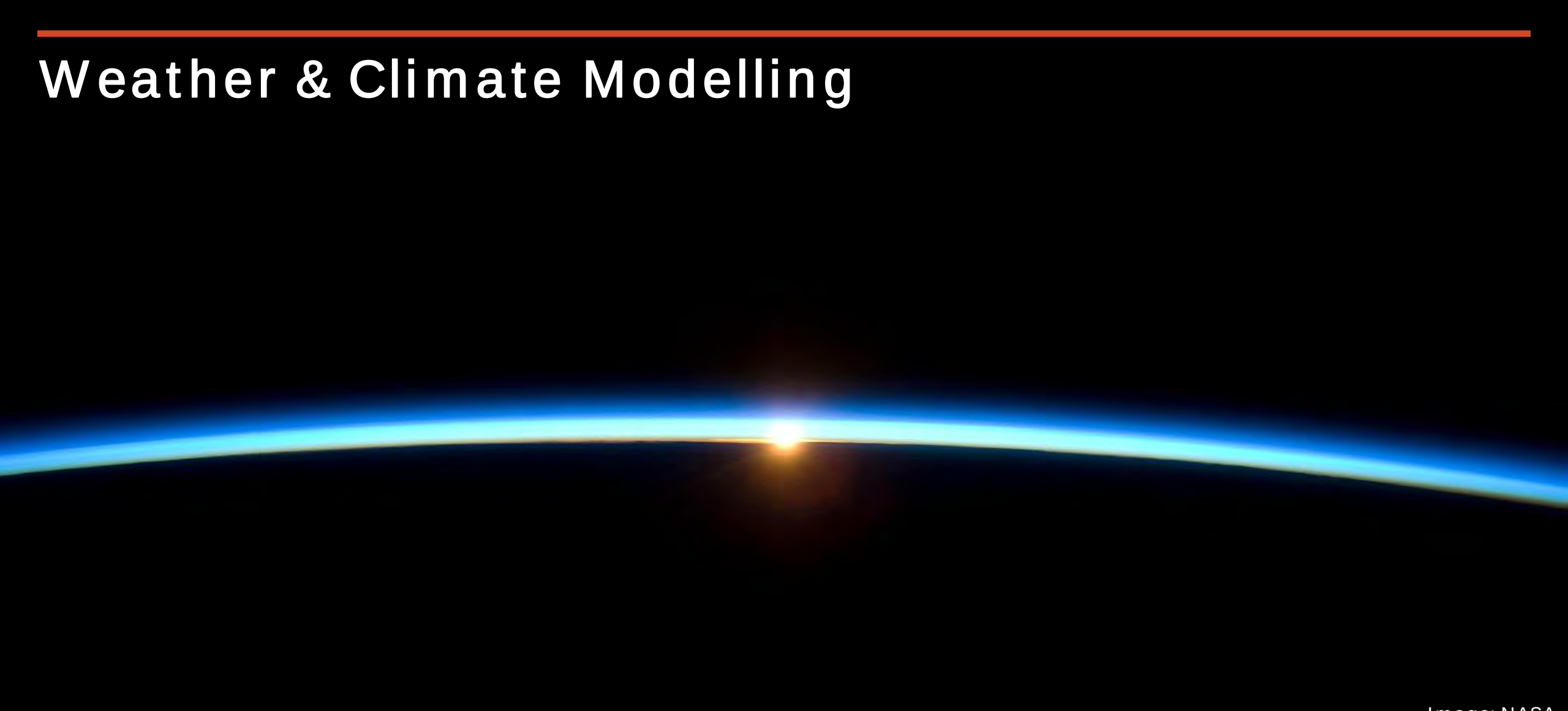
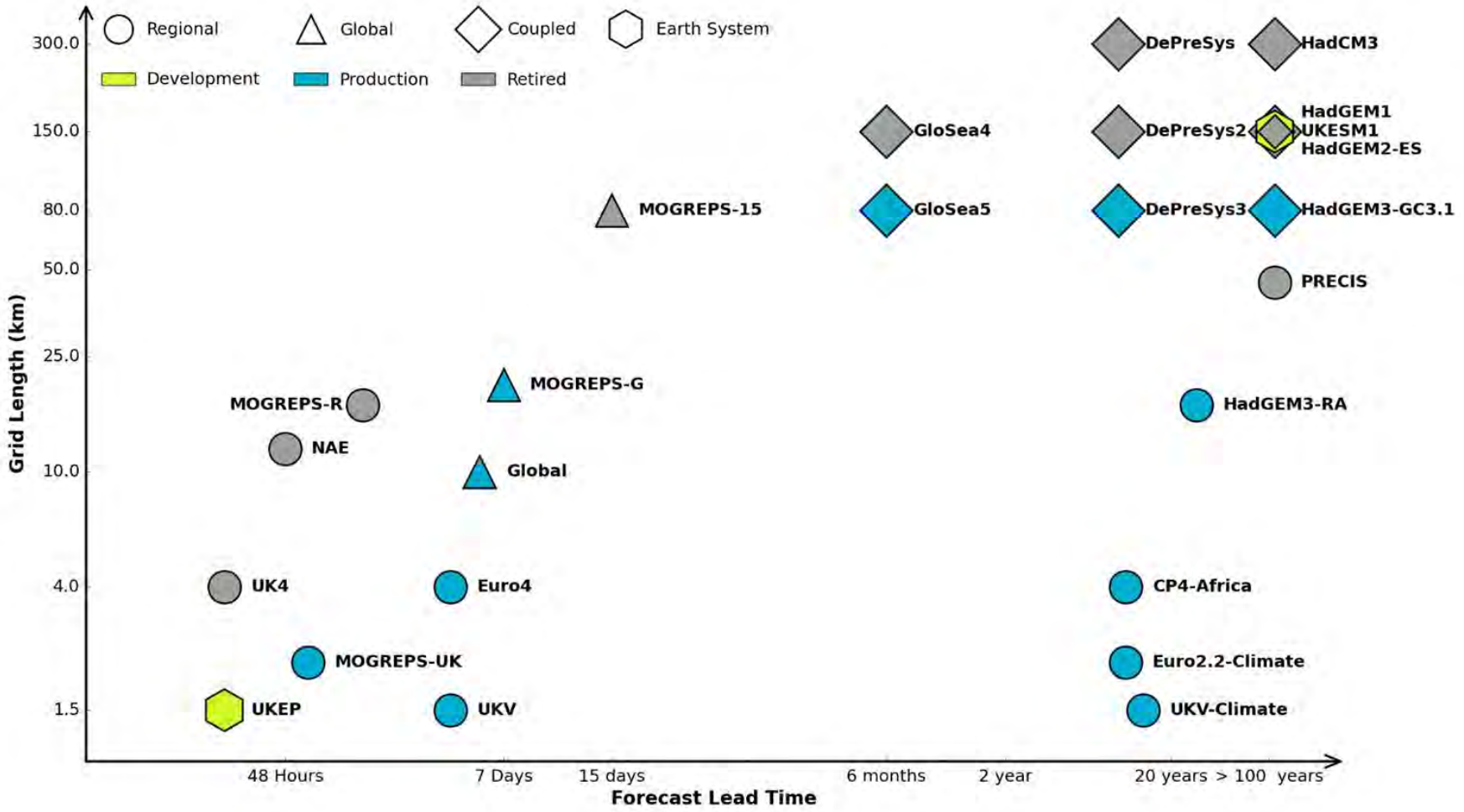


Image: NASA

Weather and Climate Modelling

Current Unified Model Configurations



June 2022 Top500

75	Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect , HPE United Kingdom Meteorological Office United Kingdom
207	Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect , HPE United Kingdom Meteorological Office United Kingdom
208	Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect , HPE United Kingdom Meteorological Office United Kingdom

June 2022 Top500

75	Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect , HPE United Kingdom Meteorological Office United Kingdom

207	Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect , HPE United Kingdom Meteorological Office United Kingdom
208	Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect , HPE United Kingdom Meteorological Office United Kingdom

PREPARING THE MET OFFICE FOR THE NEXT NEXT GENERATION OF SUPERCOMPUTERS

Abram

Joe ~~Hickson~~
[he/him/his]
joseph.hickson@metoffice.gov.uk

Victoria Smart
[she/her/hers]
victoria.smart@metoffice.gov.uk

Lewis Sampson
[he/him/his]
lewis.sampson@metoffice.gov.uk

CIUK 1-2 December 2022

PREPARING THE MET OFFICE FOR THE NEXT NEXT NEXT NEXT NEXT NEXT GENERATION OF SUPERCOMPUTERS

Joe ~~Hickson~~ Abram
[he/him/his]
joseph.hickson@metoffice.gov.uk

Victoria Smart
[she/her/hers]
victoria.smart@metoffice.gov.uk

Lewis Sampson
[he/him/his]
lewis.sampson@metoffice.gov.uk

CIUK 1-2 December 2022

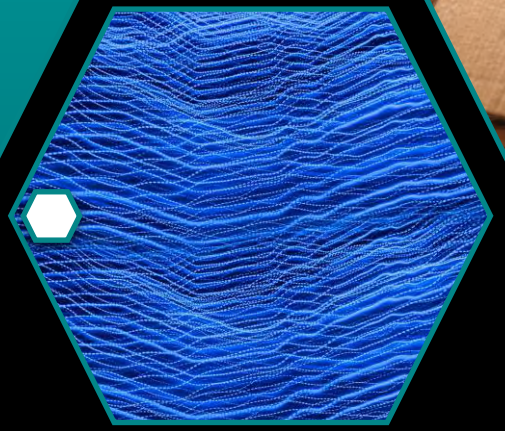
Major Themes

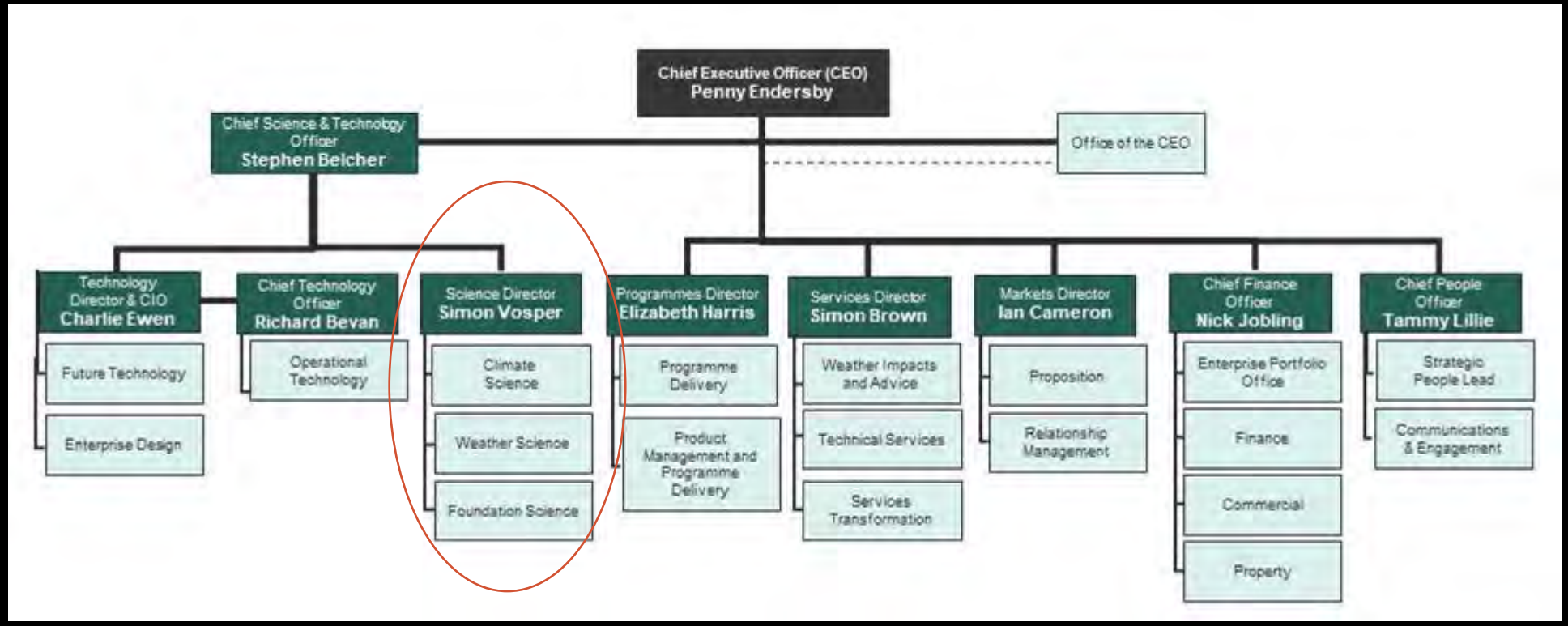
Resilience

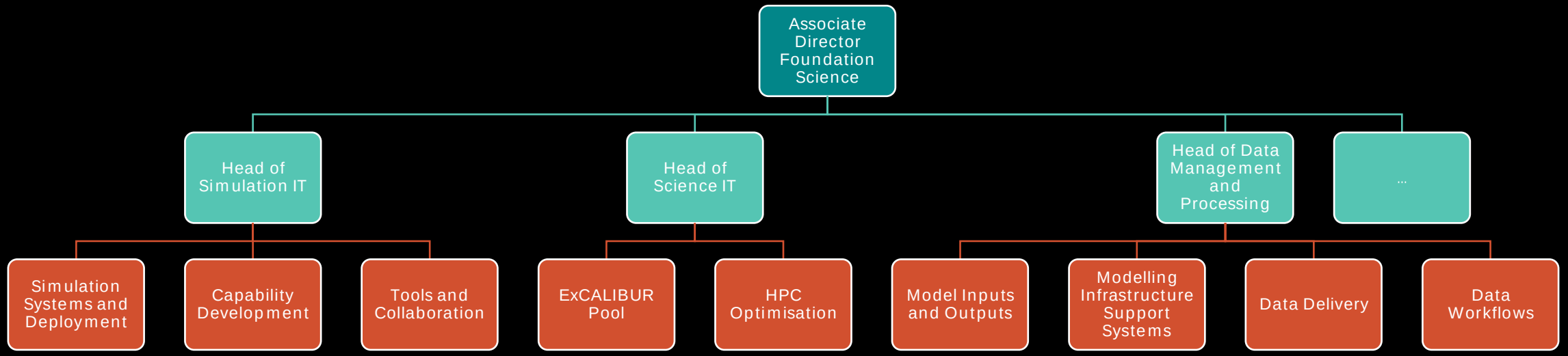
Portability

Scalability

Performance








EXCALIBUR
10
MIXED PRECISION
ALGORITHMS
Met Office
UK Research
and Innovation
UK Atomic
Energy
Authority

Photo by Hendrik Morkel on Unsplash

Mixed Precision Algorithms - Key Contributors:

Ben Shipway[1]

Steve Mullerworth[1]

Richard Malone[1]

Tom Bendall[1]

Mike Hobson[1]

Chris Maynard[1]

Tom Melvin[1]

Iva Kavcic[1]

Andrew Porter[2]

Rupert Ford[2]

Sergi Siso[2]

1: Met Office

2: UKRI Science and Technology Facilities Council

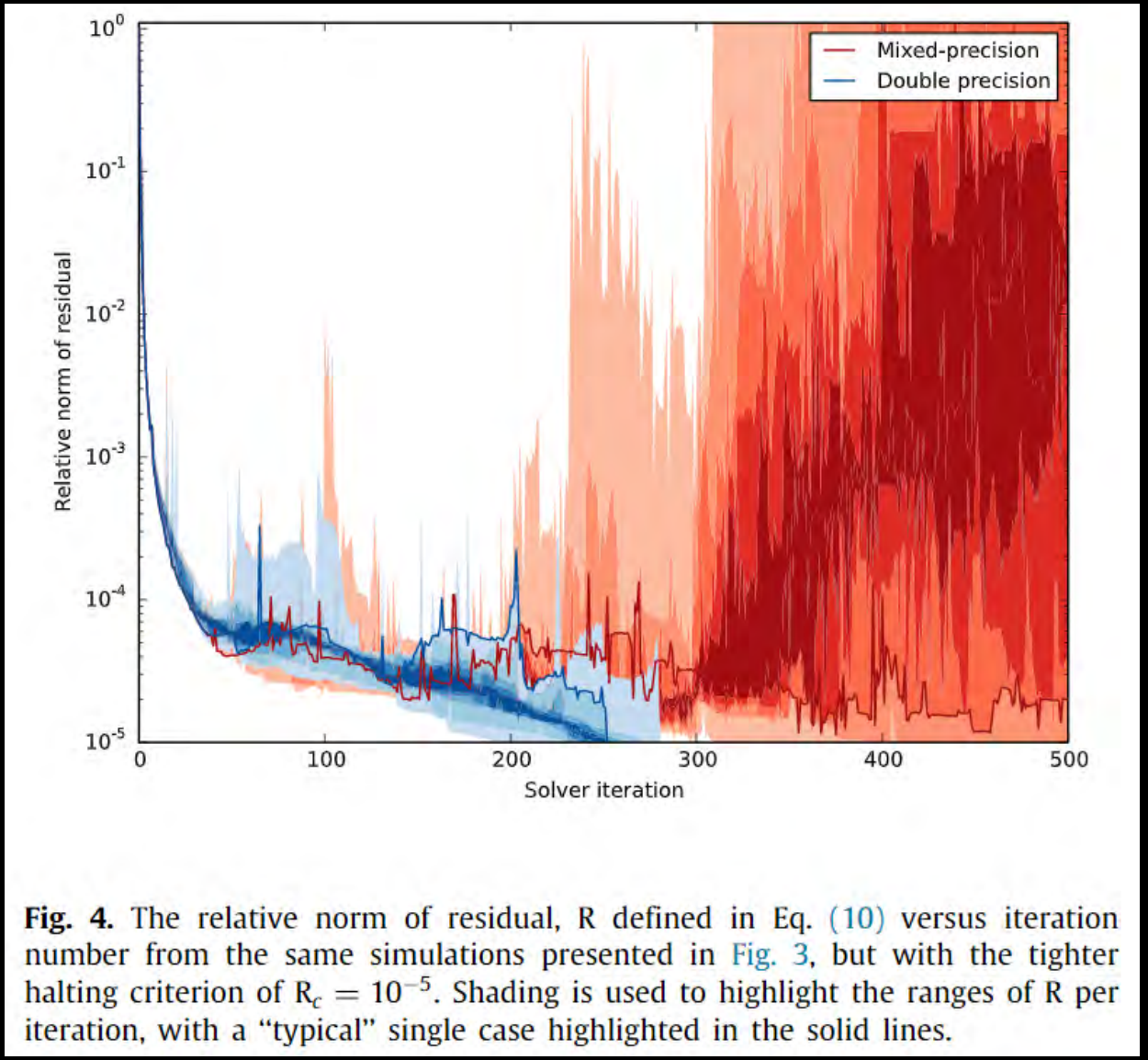


Fig. 4. The relative norm of residual, R defined in Eq. (10) versus iteration number from the same simulations presented in Fig. 3, but with the tighter halting criterion of $R_c = 10^{-5}$. Shading is used to highlight the ranges of R per iteration, with a "typical" single case highlighted in the solid lines.

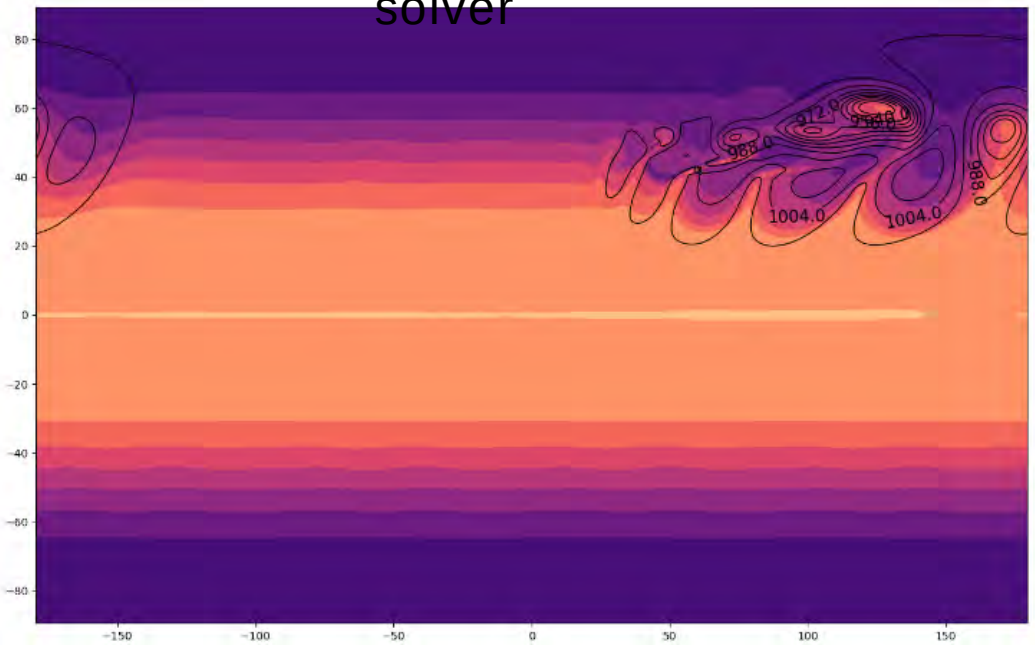
Mixed-precision arithmetic in the ENDGame dynamical core of the Unified Model, a numerical weather prediction and climate model code

C.M. Maynard and D.N. Walters / Computer Physics Communications 244 (2019) 69–75

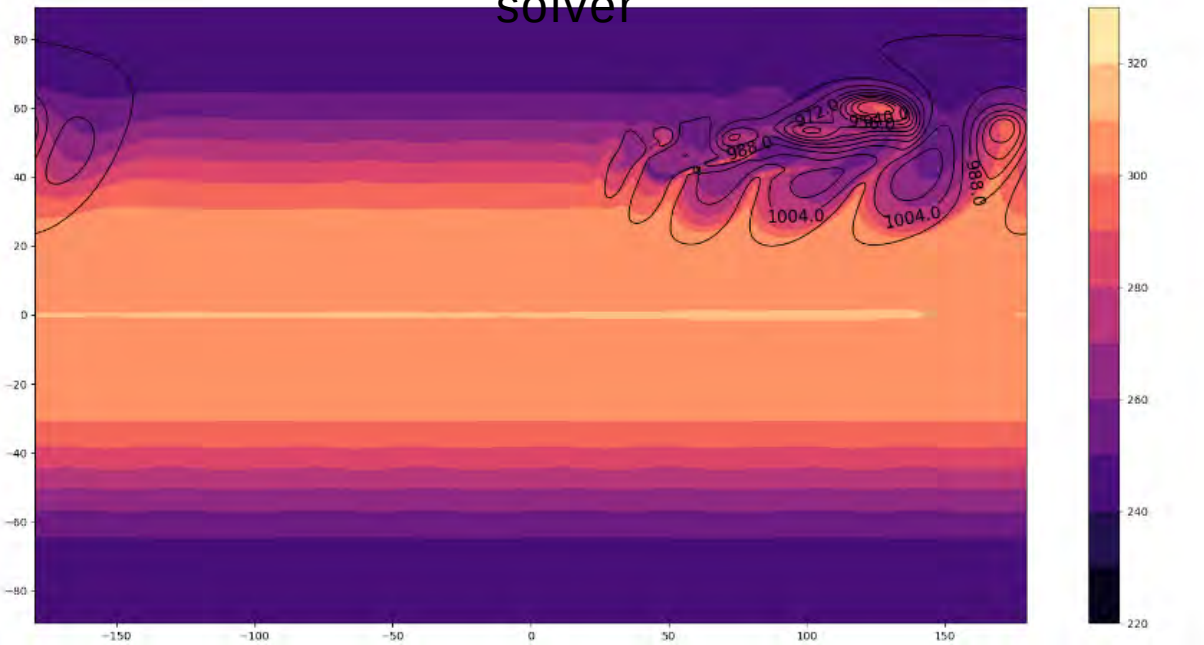
Scientific Evaluation

GungHo Baroclinic wave test

64-bit solver



32-bit solver



```
#if (R_TRAN_PRECISION == 32)
  use field_r32_mod, only: r_tran_field_type         => field_r32_type,       &
                           r_tran_field_proxy_type   => field_r32_proxy_type, &
                           r_tran_field_pointer_type => field_r32_pointer_type
#else
  use field_r64_mod, only: r_tran_field_type         => field_r64_type,       &
                           r_tran_field_proxy_type   => field_r64_proxy_type, &
                           r_tran_field_pointer_type => field_r64_pointer_type
#endif
```

GPU EXPLORATIONS

Victoria Smart
Lewis Sampson

Met Office

UK Research and Innovation

UK Atomic Energy Authority

Marine Systems GPU Projects - Key Contributors:

Lewis Sampson[1]

Victoria Smart[1]

Chris Bunney[1]

Martin Price[1]

Mike Bell[1]

Matthew Martin[1]

Muhammad Asif[1]

Joseph Wallwork[1]

Wayne Gaudin[2]

David Norton[2]

Seth Camp[2]

Andrew Porter[3]

Rupert Ford[3]

Sergi Siso[3]

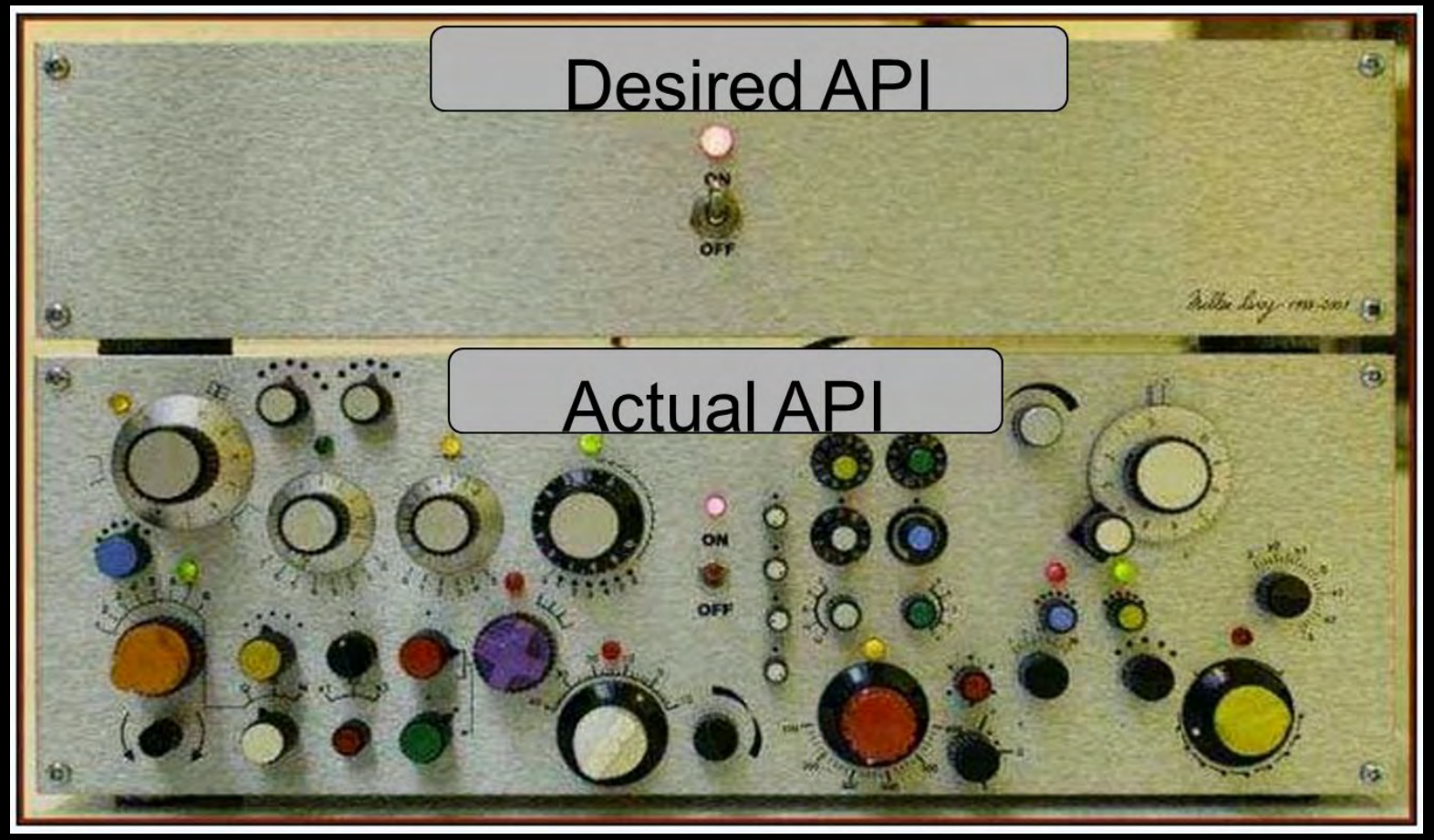
1: Met Office

2: NVIDIA

3: UKRI Science and Technology Facilities Council

NEMO

NOAA



Ben Rogers (University of Manchester) and Phil Hasnip (University of York)

PAX-HPC - Modelling particles at exascale: from atoms to galaxies

Abstract:

Bios: Prof Benedict D. Rogers (Manchester) is Chair of Computational Hydrodynamics and is leader of the Smoothed Particle Hydrodynamics (SPH) specialist group in the Department of Mechanical,

Aerospace and Civil Engineering at the University of Manchester. He is a founder of the SPH rEsearch and engineeRing International Community (SPHERIC), the international organisation for SPH and was chair from 2015-2021. He is a core developer of the open-source GPU-accelerated code DualSPHysics with more than 100,000+ downloads and has been at the heart of the development of massively parallel incompressible SPH solvers. He is a co-Investigator on 2 projects for the UK's ExCALIBUR exascale HPC programme including the Particles At eXascale (PAX-HPC) project. He has twice received the Thomas Telford Premium Award from the Institution of Civil Engineers (ICE) for his work on SPH applied to tsunami-structure interaction.

Phil Hasnip is a physicist and computer programmer in the Department of Physics at the University of York. He grew up in the 1980s, where he learned physics at school and computer programming on his Sinclair ZX Spectrum. He is an EPSRC Research Software Engineering Fellow, and writes computer software to tackle problems in physics research, with a particular interest in making scientific software user-friendly, scalable, efficient and reliable. Phil is a lead developer of the quantum mechanical materials modelling program CASTEP, chairs the UKCP High End Compute Consortium, and is the Knowledge Exchange Coordinator for the Particles At eXascale (PAX-HPC) project for the UK's ExCALIBUR exascale HPC programme, working to ensure that key UK modelling methods are ready for the next generation of HPC machines.

Engineering and Physical Sciences Research Council

ExCALIBUR

PAX-HPC - Modelling particles at Exascale: from atoms to galaxies

Benedict D. Rogers
University of Manchester

Phil Hasnip
University of York

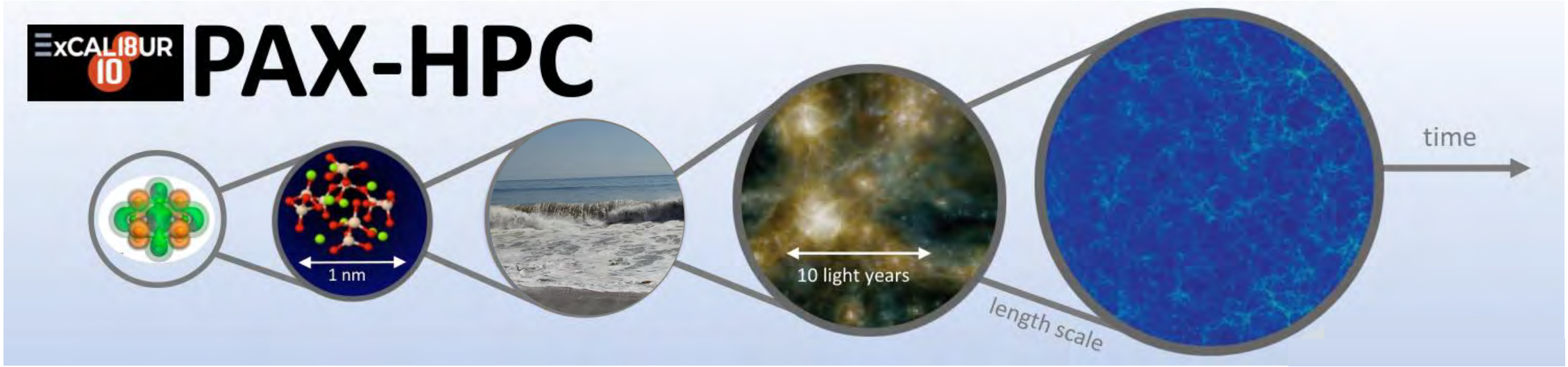
University of York · Durham University · University of Lincoln · Lancaster University · Universiteit Leiden, The Netherlands · UKRI Science and Technology Facilities Council

UCL · University of Hertfordshire UH · King's College London · Manchester 1824 · University of Southampton · epcc

Contents

1. Introduction to PAX-HPC
2. Progress on exascale development for atoms to meso-scale
3. Progress on exascale development for continuum fluids on Earth to Galaxies
    - Massively Parallel Particle Hydrodynamics (MPPH) Working Group
4. Discussion Points
5. Perspectives

Particles make up the universe



Particle-based methods in Science & Engineering

- Conventional modelling has often used mesh-based methods
- Particles make up everything we see and experience
- Particle-based methods are an intuitive to simulate scientific and engineering phenomena
- Particle-based methods are **ideally suited** to massive parallelisation

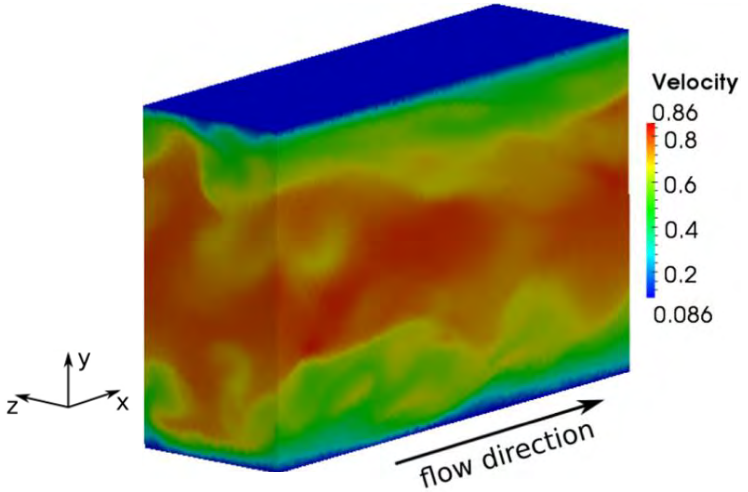
Why is Exascale Computing needed for Particles? #1

- There are three broad categories of calculations at atomic scales
- High-throughput calculations
  - Use a loosely-coupled ensemble of many simulations
  - Examples include materials discovery, using global optimisation methods (e.g. Genetic Algorithms) to find new materials with improved properties
- "Hero" calculations
  - A small number of simulations on large, complex systems
  - Examples include large-scale dynamics, e.g. liquid lithium for fusion reactor cooling, and sophisticated quantum mechanical simulations
- Complex workflows, e.g. multiscale, multiphysics
  - Coupling different methods and software together
  - Examples include embedding quantum mechanical simulations within otherwise classical simulations

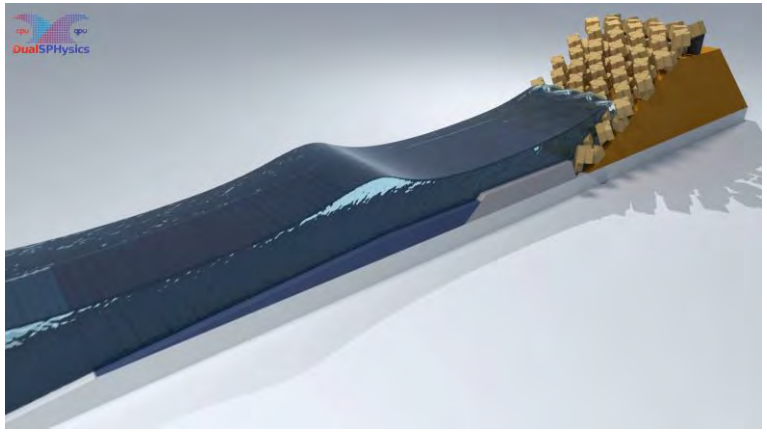
Why is Exascale Computing needed for Particles? #2

- Smoothed Particle Hydrodynamics (SPH) is a very attractive method for applications with high nonlinearity: Astronomy, planet collision, Fluid-Structure Interaction

Keggeris et al., 2019



Mayrhofer et al., 2015



Altomare et al., 2014

- Often approached by companies and industrial collaborators with '**Wicked Applications**'
- However, 'Wicked applications' in industry beyond any numerical scheme still require **engineering approximations** (with or without variable resolution)

Challenges Slide

Key Challenges for Exascale particle-based methods in Science & Engineering:

1. Long-range forces can act across the physical domain
2. Vast and rapid changes in scales, e.g. 10^9 for density
3. Enormous range of resolutions required
4. The data structures are vastly different from Mesh-based techniques

TOP SUPERCOMPUTERS IN THE WORLD June 2022

http://www.top500.org

R_{max} and R_{peak} values are in TFlops. For more details about other fields, check the TOP500 description.

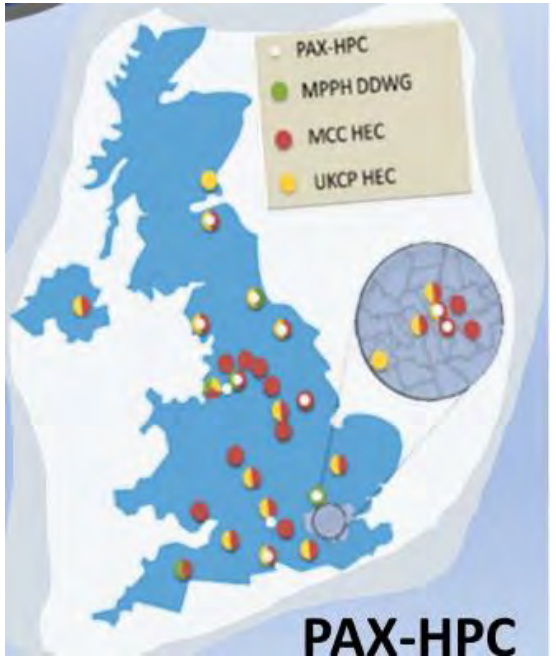
Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	Frontier - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	1,685.65	21,100
2	Supercomputer Fugaku - Supercomputer Fugaku	7,630,848	442.01	537.21	28,899

1º Frontier (USA) 1102 petaflop/s (**consumption: 21,100 kW**)
2º Fugaku (Japan) 415 petaflop/s (**consumption: 28,335 kW**)

Frontier is **heterogeneous**: **CPUs** + **GPUs**
64-core AMD Epyc Radeon Instint MI250X

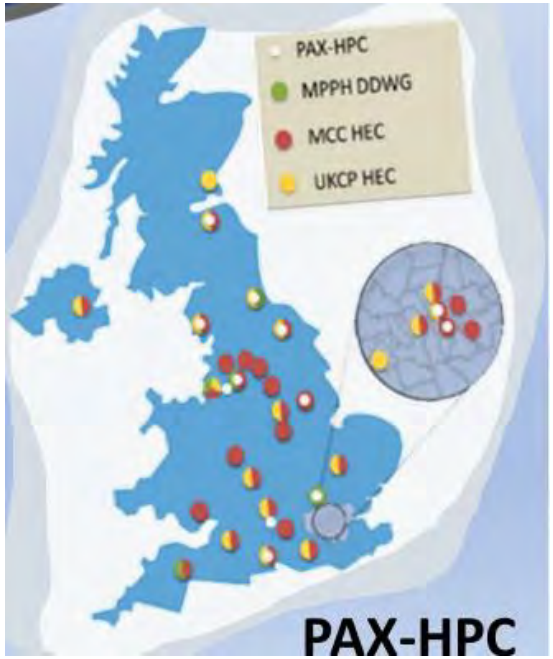
Folding@Home performed 1.22 ExaFLOPS in March 2020 using millions of Home PCs

PAX-HPC Project



- Particles-At-eXascale for High-Performance Computing (PAX-HPC), £3m ExCALIBUR project
- **Aim**: redesign high-priority particle-based codes & algorithms for exascale
- **New Methods of new hardware**
  - Performance portable programming methods (SYCL)
  - Exploit emerging network technologies (DPUs, GPU-aware communications)
- **New parallelisation strategies**
  - High-level additional parallelism (Task-farming, parallel-in-time methods)
  - Low-level task parallelism: Task graphs, task schedulers
- **Complex Workflows**
  - Coupling multiscale, multi-physics modes with FAIR data

PAX-HPC Project Team



PAX-HPC Combines Expertise:

- Massively Parallel Particle Hydrodynamics (MPPH)
- Materials and Molecular Modelling (MMM)
- Materials Chemistry Consortium (MCC)
- UK Car-Parrinello Consortium (UKCP)

- PI: Prof. Scott Woodley (UCL)

Engineering and Physical Sciences Research Council

ExCALIBUR

PAX-HPC - Modelling particles at exascale: from atoms to galaxies

Part 1: Atoms to Meso-scale

Phil Hasnip
University of York

Parallel Eigensolvers

(Marcello Puligheddu and Ian Bush)

$$H\Psi = E\Psi$$

Matrix diagonalisation to calculate eigenvalues & eigenvectors is key to many quantum mechanical simulations of molecules and materials.

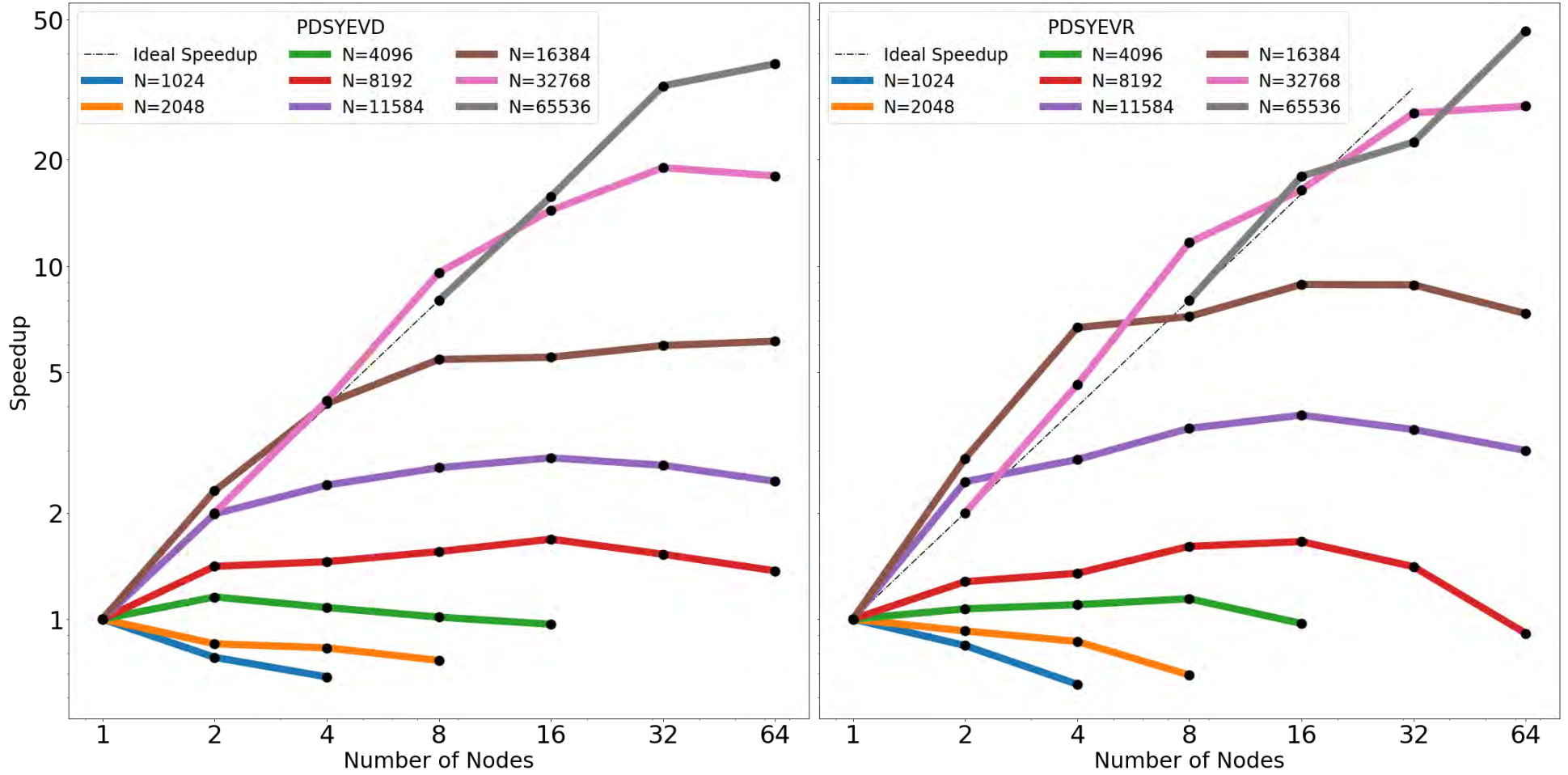
When diagonalising a N x N matrix:

- computational storage scales as N^2
- time scales as N^3.

For local basis set methods, N is the number of basis functions, but the number of electrons may be lower. We may only require 10 – 30% of the eigenvectors.

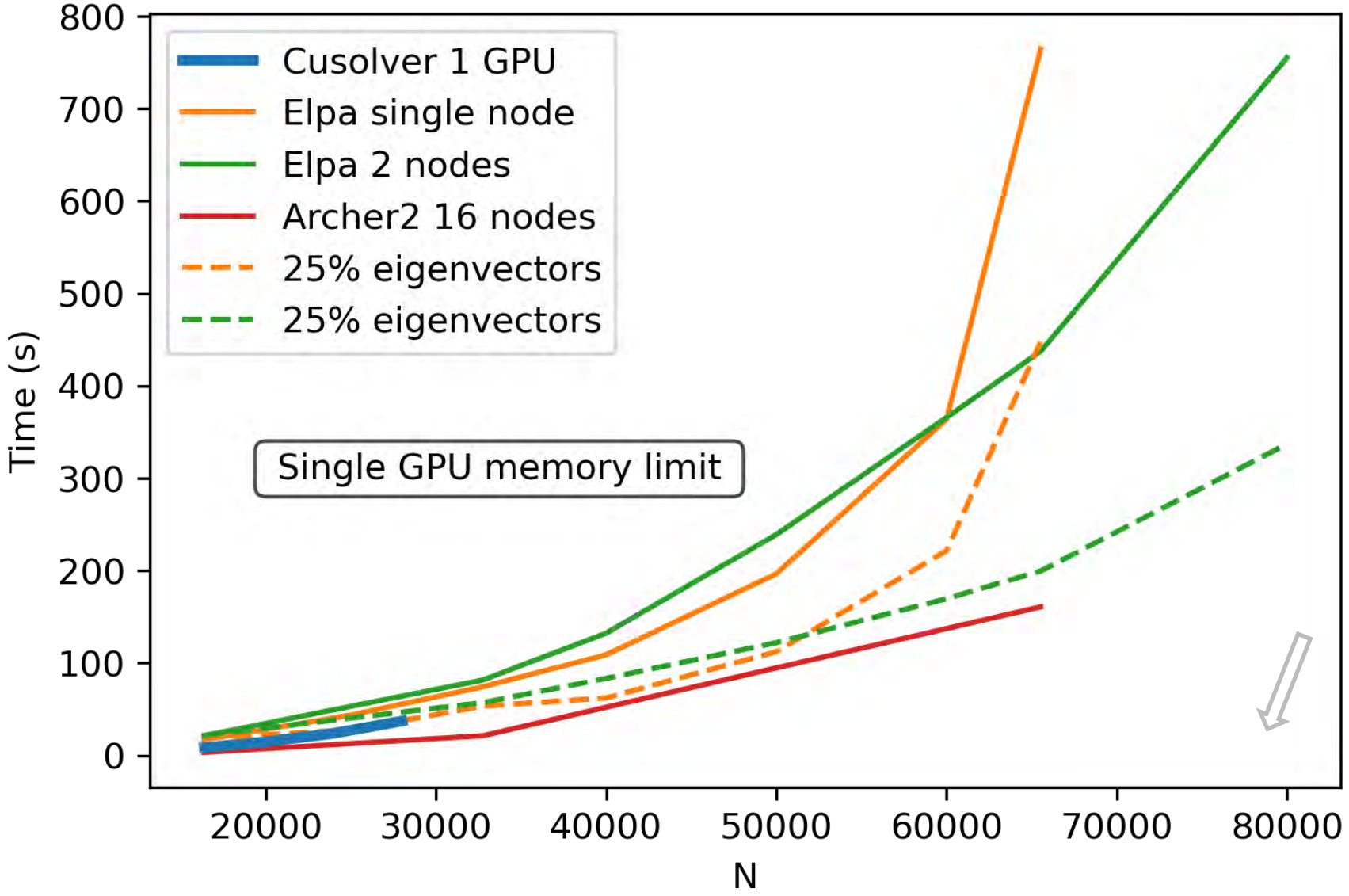
Benchmarking Eigensolvers

ScaLAPACK is a standard interface for parallel eigensolvers.



Archer2 results for hybrid OpenMP-MPI (quickest of up to 8 OpenMP threads per process)

GPU Eigensolvers (ELPA)



ELPA has a more parallelisable algorithm.

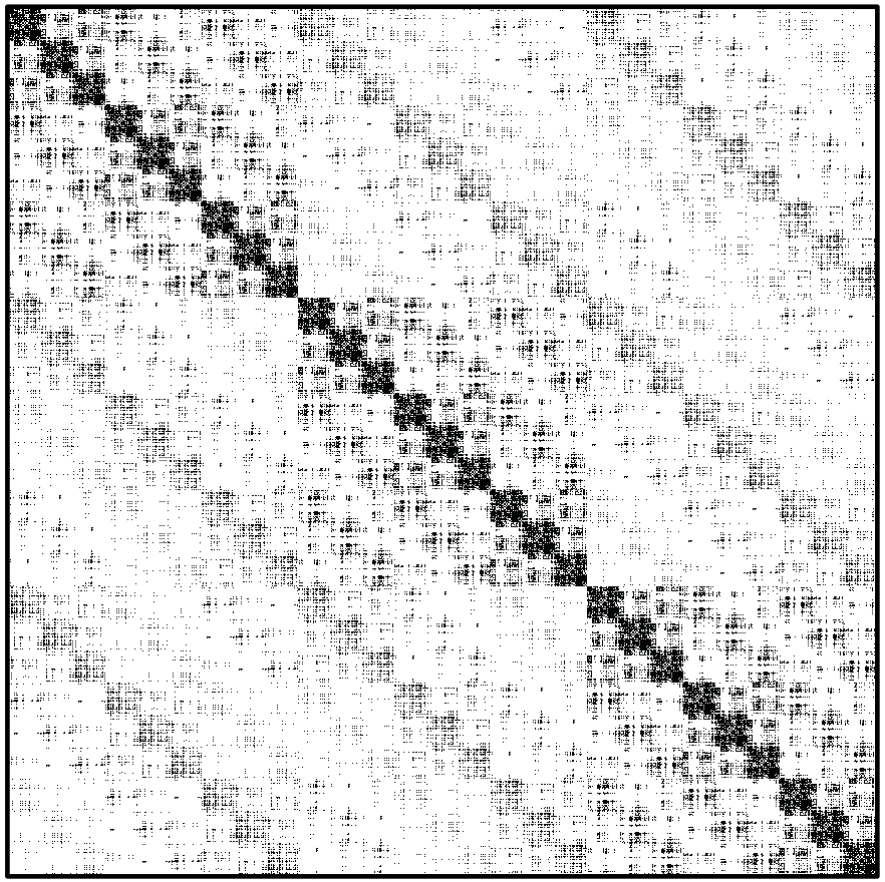
Results shown for 2-stage solver, which can also compute a partial solution.

Scarf results (2x AMD 16 cores + 4 A100 40GB per node) with the 2022.11

GPU Acceleration of the Kohn-Sham Matrix

Creating the Kohn-Sham matrix is complex process in local basis set codes (e.g. CP2K, CRYSTAL), and GPU acceleration is not trivial.

- Matrices are sparse; how can we exploit that?
- First steps:
  - Use the overlap matrix as a simple test-bed
  - Merge into CP2K and CRYSTAL
- Extend to full Kohn-Sham matrix
- 4-centre (2-electron) integrals on GPUs
- Load balancing is challenging!

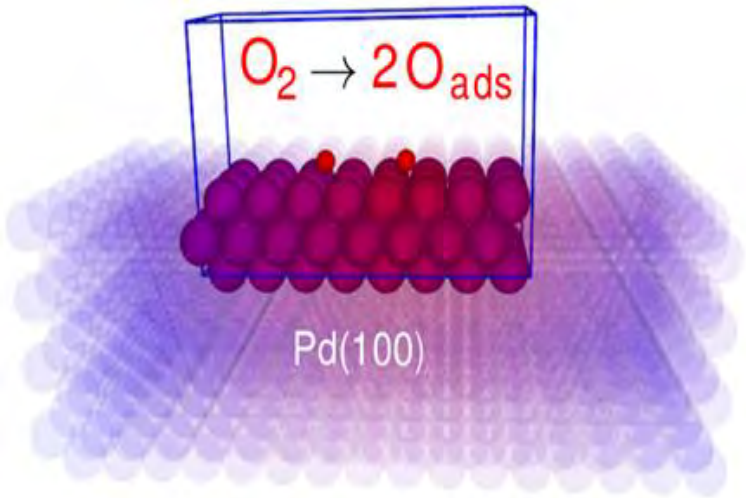


Sparsity pattern of the kinetic energy matrix for a 864 H2O molecule system

Complex materials modelling workflows

Rajany K.V., Alin Elena,
Tom Keal (STFC)

- **Initial focus on *ChemShell* QM/MM package to all PAX-HPC QM codes**
  - New prototype CASTEP interface to ChemShell
  - Work started on closely-coupled CASTEP-ChemShell interface, using a new CASTEP API
  - Setting up benchmarks for CP2K and CASTEP on STFC's SCARF cluster and ARCHER2 for periodic QM/MM of metallic systems, e.g. oxygen adsorption on palladium



- **Extend to new workflows for multiscale modelling**
- **Optimise I/O for exascale HPC (targeting multiple PAX-HPC software projects)**

Engineering and Physical Sciences Research Council

ExCALIBUR

PAX-HPC - Modelling particles at exascale: from atoms to galaxies

Part 2: Earthly fluids to Galaxies

Benedict D. Rogers
University of Manchester

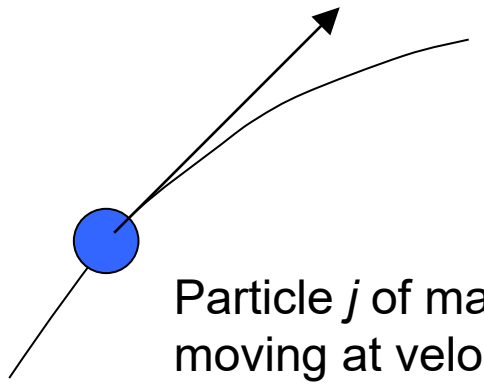
Meshless methods: Basic Idea of SPH

Meshless Our computation points are **particles** that now **move** according to governing dynamics , e.g. Navier-Stokes Equations

Particles move along a trajectory by **integrating** in time their velocity & acceleration

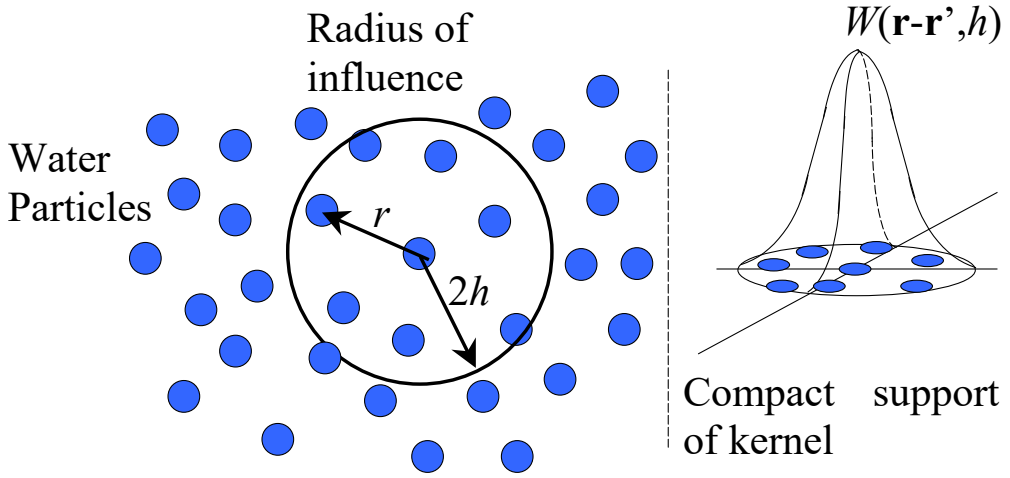
Particles possess **properties** that travel with them, e.g. density, pressure; these can change with time

Local Interpolation (summation) with a **weighting function** (kernel) around each particle to obtain fluid/solid properties



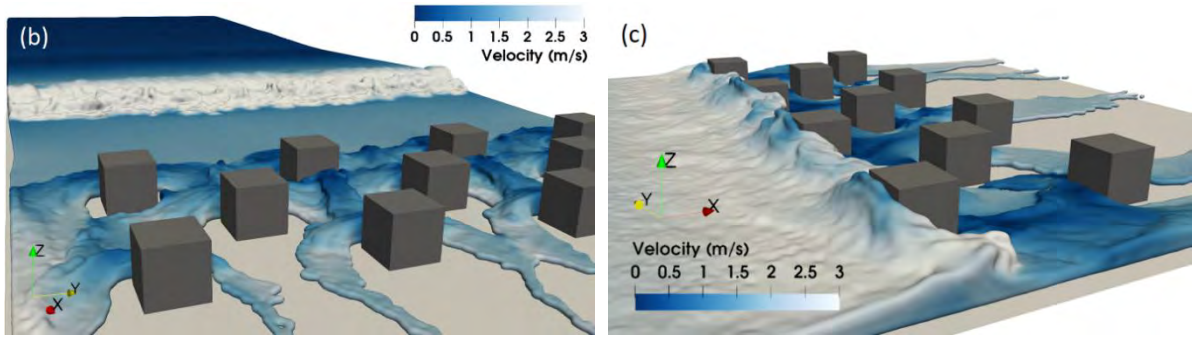
Particle j of mass m_j moving at velocity $\mathbf{v_j}$

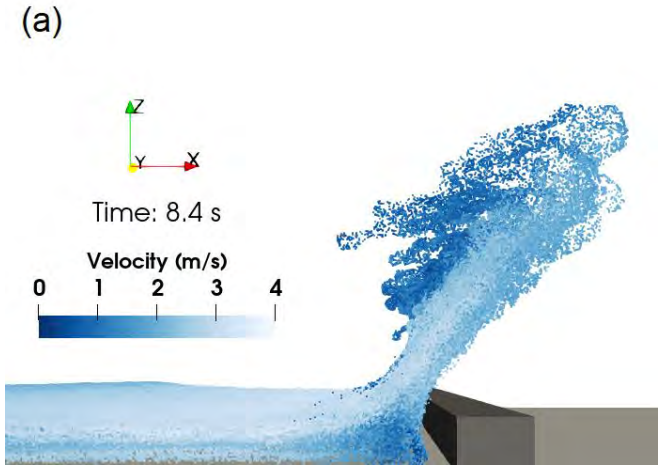
$$\langle f(\mathbf{x})\rangle = \sum_j f_j \, W(\mathbf{x}-\mathbf{x}_j)\frac{m_j}{\rho_j}$$



SPH Application: Tsunami inundation

One of the key applications of SPH is the simulation of violent wave impact on structures: **RESILIENCE + SUSTAINABILITY**





cpu gpu
DualSPHysics

MANCHESTER 1824

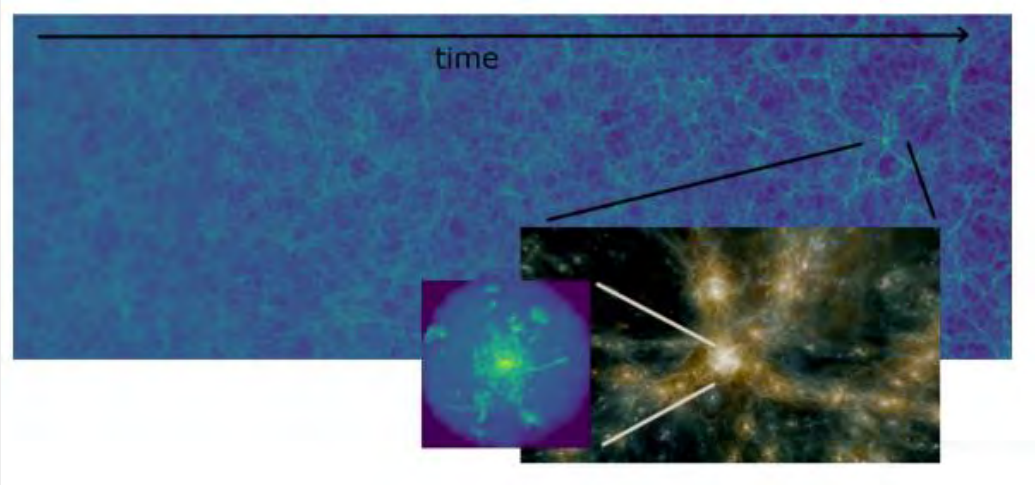
UNIVERSITY OF UDAYANA

Massively Parallel Particle Hydrodynamics (MPPH) Working Group

Astrophysics:

Durham University

Universiteit Leiden
The Netherlands




```
 ______        _________________
/ ___/ |      / /  _/ ___/_  __/
\__ \| | /| / // // /_    / /
___/ /| |/ |/ // // __/   / /
/____/ |__/|__/___/_/     /_/
```

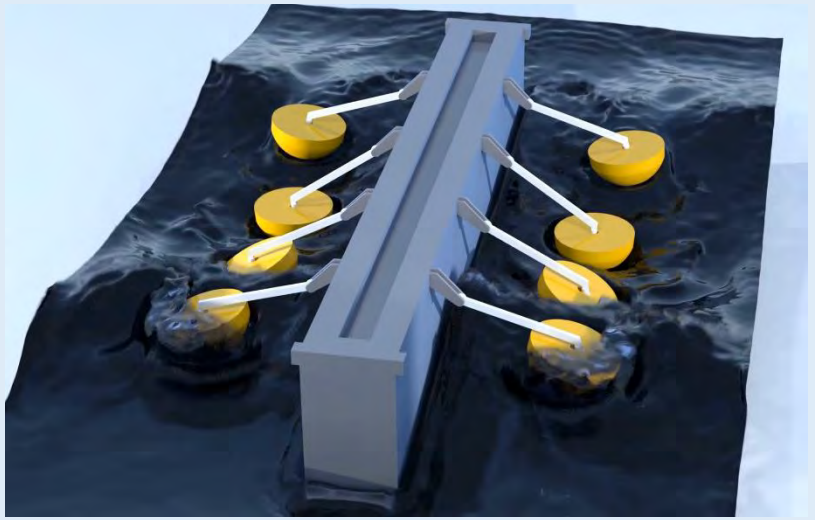
Massively-parallel **CPUs** with Task-based Parallelism

Engineering:

MANCHESTER 1824

University of Hertfordshire UH

UKRI Science and Technology Facilities Council



cpu gpu DualSPHysics

GPU-acceleration

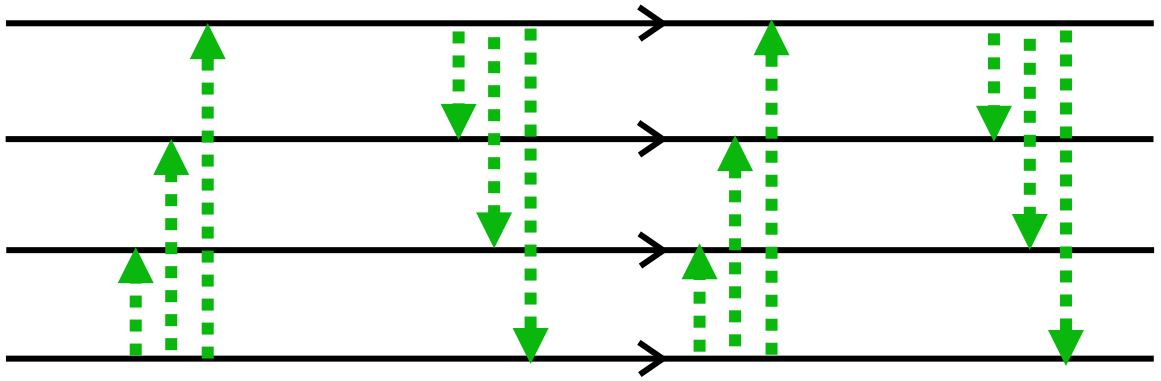
Challenges to develop Exascale-ready SPH

MPPH group has identified 4 key challenges (Bower, Rogers & Schaller, CICESE 2022):

1. Communications

Sending/Receiving data using asynchronous & overlapping communications → unpredictable delays

4 nodes



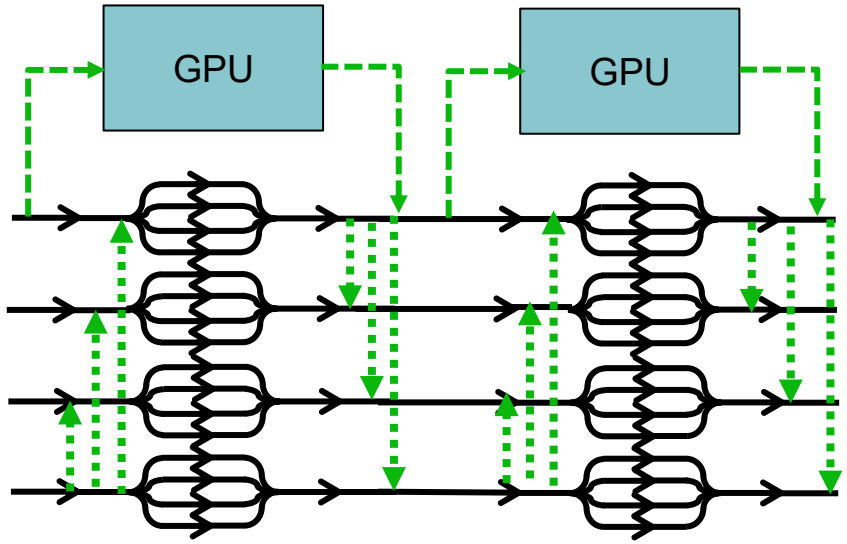
Multiple Processes on Multiple Nodes using MPI (distributed memory)

Challenges to develop Exascale-ready SPH

MPPH group has identified 4 key challenges (Bower, Rogers & Schaller, CICESE 2022):

1. Communications — Sending/Receiving data using asynchronous & overlapping communications → unpredictable delays

2. GPU-enabled code — Exascale machines will be a mix of pure CPUs and accelerators (GPUs)



4 processes x 4 threads

MPI+OpenMP +GPUs

Challenges to develop Exascale-ready SPH

MPPH group has identified 4 key challenges (Bower, Rogers & Schaller, CICESE 2022):

1. Communications — Sending/Receiving data using asynchronous & overlapping communications → unpredictable delays
2. GPU-enabled code — Exascale machines will be a mix of pure CPUs and accelerators (GPUs)
3. New algorithms — How do we exploit time-step adaptivity with variable resolution that scales over 1000k+ cores?

$$\Delta t = CFL \times \min(\Delta t_{force}, \Delta t_{visc}, \Delta t_{diff})$$

Using uniform timestep everywhere?

Really?! Over 500,000+ cores ?

Challenges to develop Exascale-ready SPH

MPPH group has identified 4 key challenges (Bower, Rogers & Schaller, CICESE 2022):

1. Communications	Sending/Receiving data using asynchronous & overlapping communications → unpredictable delays
2. GPU-enabled code	Exascale machines will be a mix of pure CPUs and accelerators (GPUs)
3. New algorithms	How do we exploit time-step adaptivity with variable resolution that scales over 1000k+ cores?
4. Separation of Concerns	Separating SPH from the parallelisation.

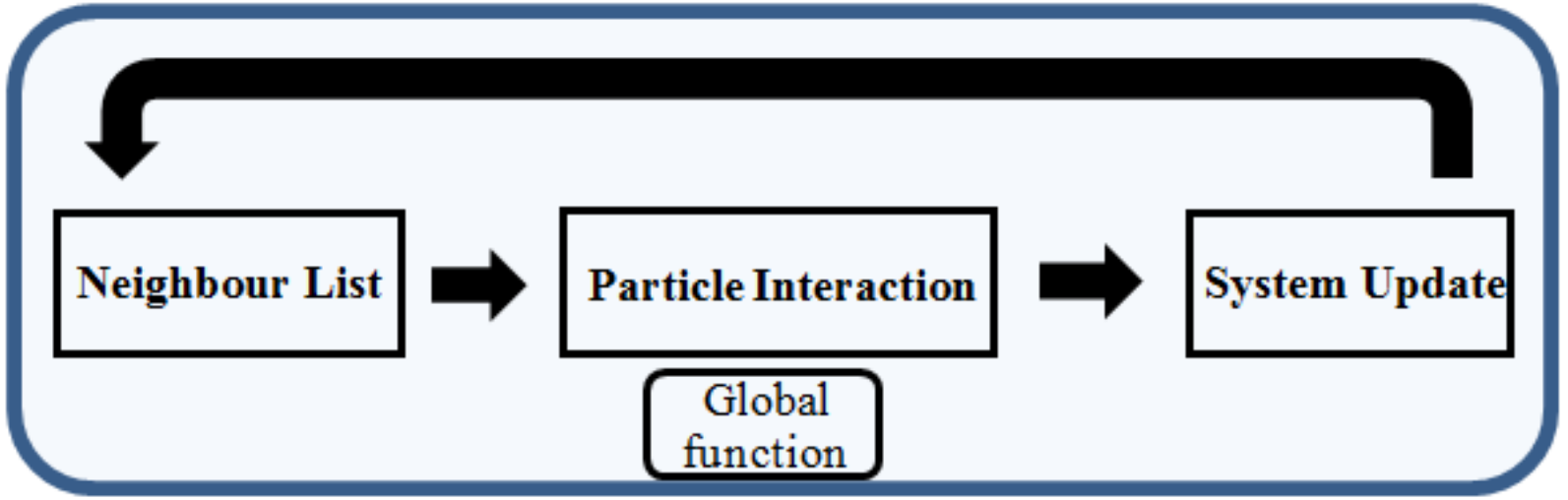
Methodologies for Exascale & Initial Results

1. Task-based parallelism
2. Asynchronous Communication
3. Task-execution on GPUs
4. Separation of Concerns

1. Task-based parallelism

Conventional SPH: **We have 3 main steps:**

1. Neighbour list (NL)
2. Particle Interaction (PI)
3. System Update (SU)



Crespo et al., 2011

SWIFT: Task-based parallelism

Tasks:

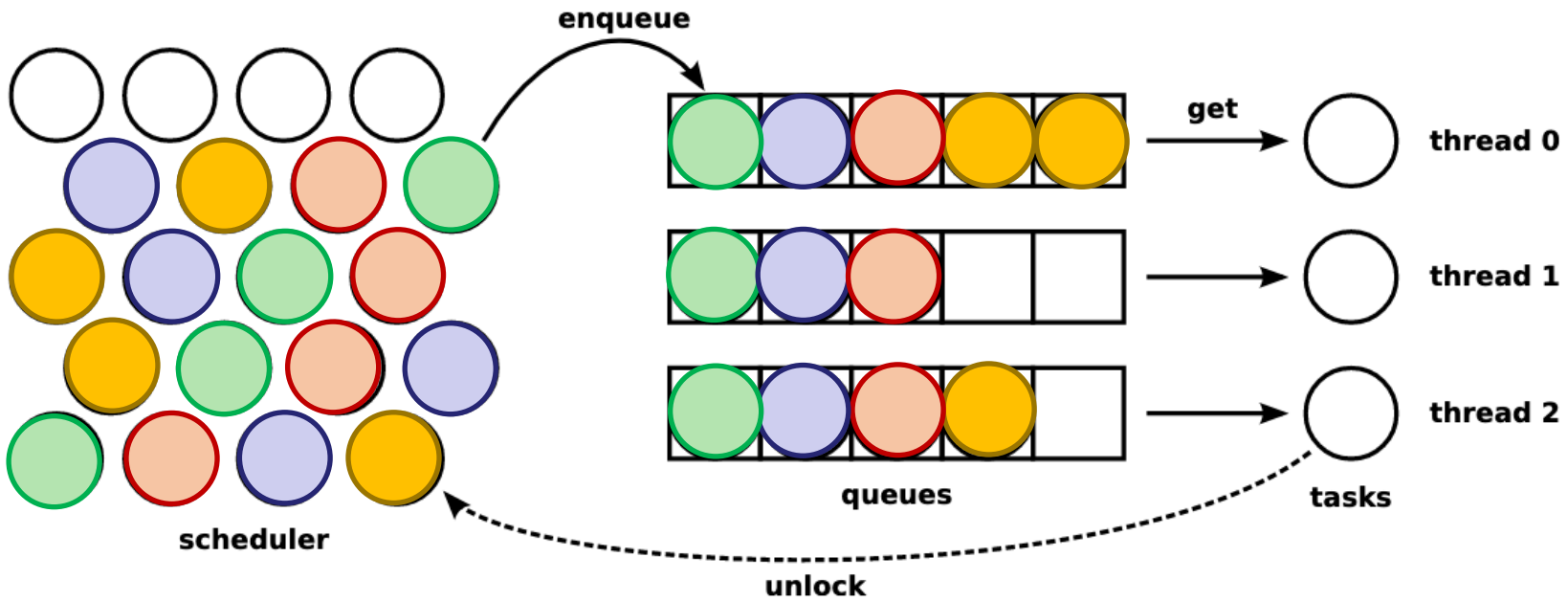
$$h_a = h_0 \left(\frac{\rho_0}{\rho_a}\right)^{\frac{1}{N_D}}$$

$$\rho_a = \sum_a m_b W(r_{ab}, h_a)$$

SEND Data

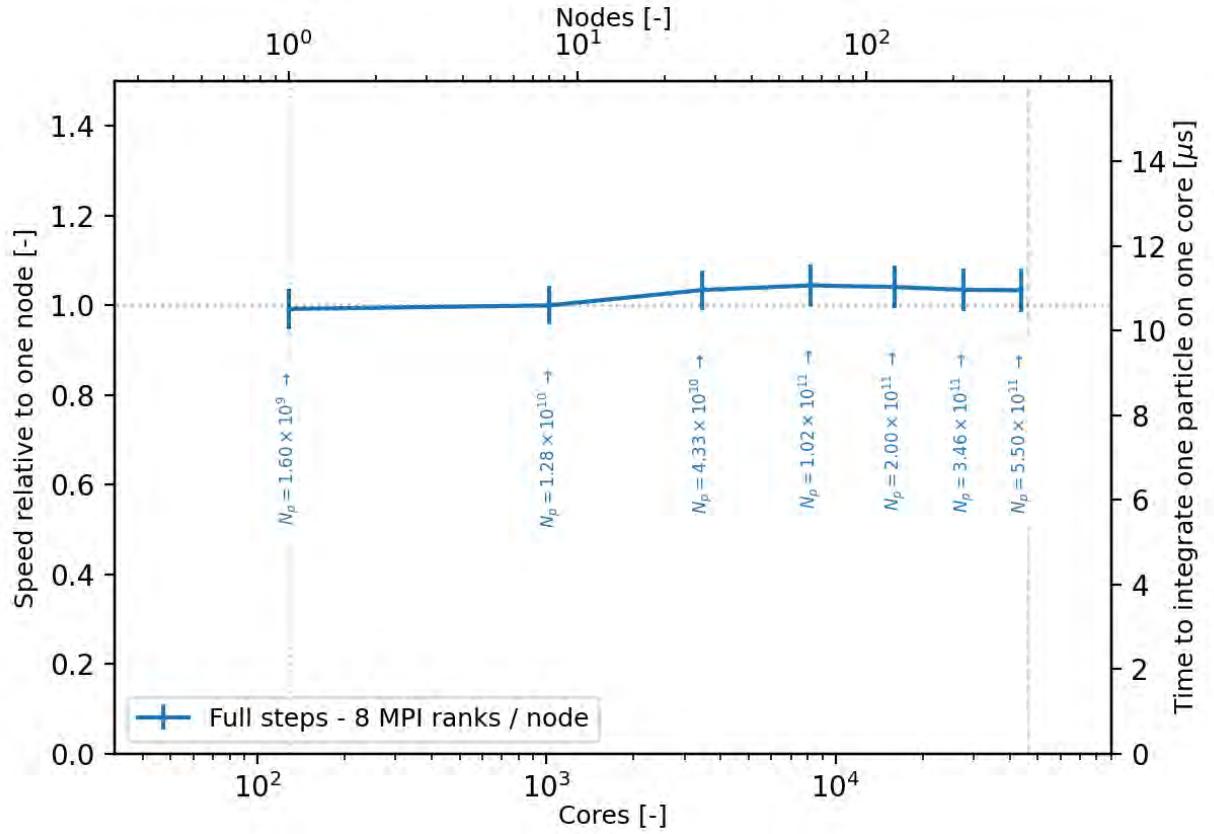
$$\frac{d\mathbf{v}_a}{dt} = -\sum_a m_b \left(\frac{P_a + P_b}{\rho_a \rho_b}\right) \nabla_a W(r_{ab}, h_a) + \mathbf{F}_v + \mathbf{F}_B$$

QUICKSCHED



Detailed full-scale testing of SWIFT

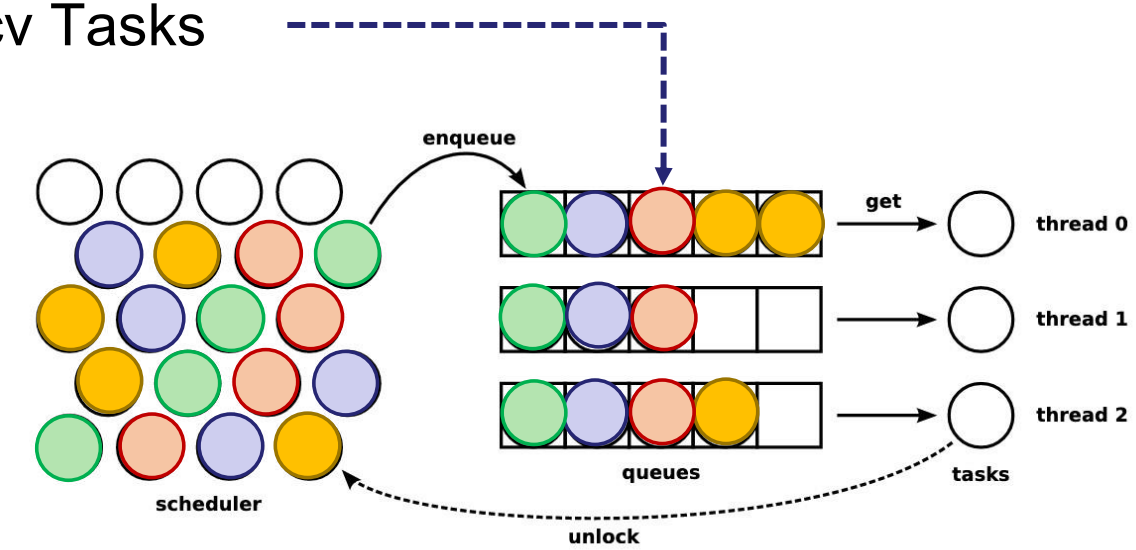
- COSMA-8 computer run by DiRAC (www.dirac.ac.uk University of Durham)
- 43,904 compute cores with 1.9 PetaFLOP peak performance
- Up to **5.5 x 10^{11} Particles** = 12.5 million particles / compute core
- Weak-scaling of 3-D periodic Kelvin-Helmholtz instability



2. Asynchronous Communication

Message Passing Interface (MPI) enables use of

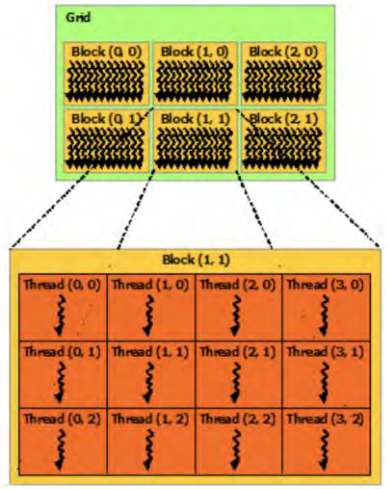
- Non-blocking data transfer → concurrent computation + data transfer
- Scheduler QUICKSCHED identifies Send-Recv Tasks



- BUT, for large simulations:
  - (i) Number of MPI communications grows
  - (ii) MPI Communications can run out of Buffer space

U-Durham using Remote Data Memory Access (RDMA) to investigate this issue.

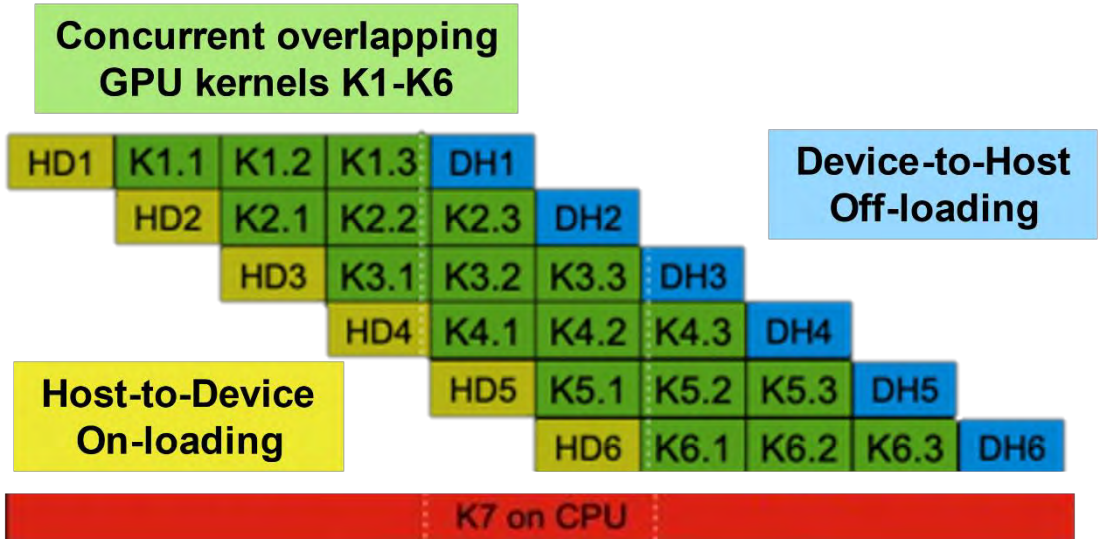
3. Task-execution on GPUs



- With blocks of streaming multi-processors, GPUs are ideal for processing **large blocks of particle interactions** (e.g. DualSPHysics, GPU-SPH)

- How to use GPUs to increase concurrency of SWIFT ?

- Our solution is to overlap CPU-GPU communication and compute CUDA kernels

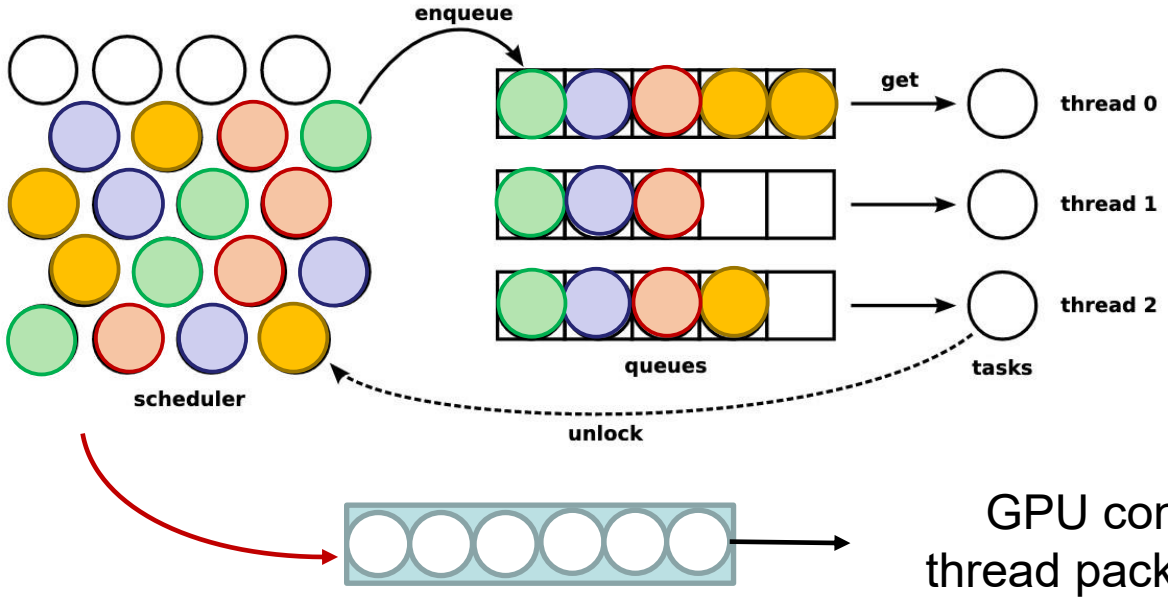


BUT

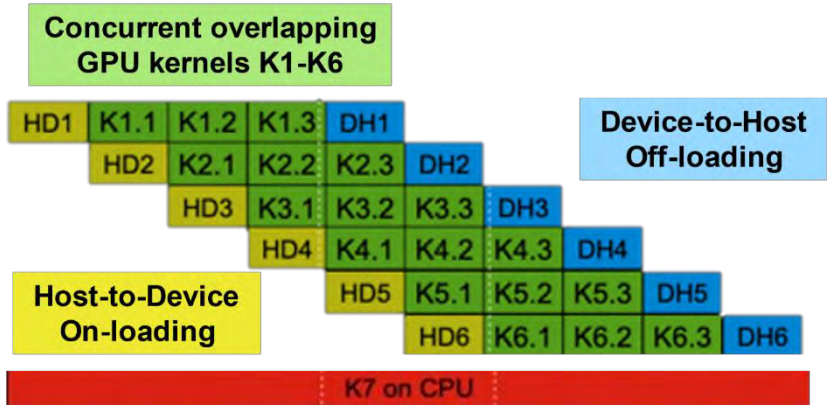
- The data structures in SWIFT (AoS) is not well suited to GPU architecture.
- How many Tasks per GPU stream do we allocate and role of pointers?

Increase concurrency: GPU on-loading/off-loading

Abouzied Nasar: Packing CPU data for GPU execution



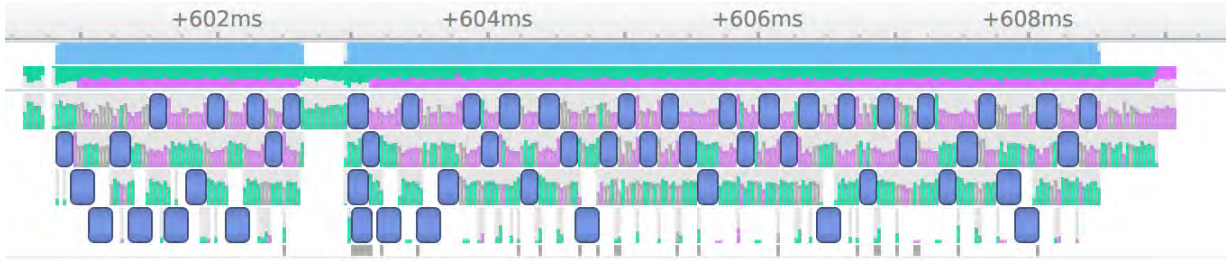
GPU control thread packs data



Optimal method still being developed.

Bundle of 8 tasks per stream: Low kernel concurrency but high memcpy/kernel concurrency

→ CPU/GPU transfer hidden

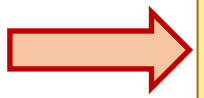


4. Separation of Concerns

SPH coder does not want to know about the parallelisation, just formulation / physics:

$$h_a = h_0 \left(\frac{\rho_0}{\rho_a}\right)^{\frac{1}{N_D}}$$

$$\rho_a = \sum_a m_b W(r_{ab}, h_a)$$



$$\frac{d\mathbf{v}_a}{dt} = -\sum_a m_b \left(\frac{P_a + P_b}{\rho_a \rho_b}\right) \nabla_a W(r_{ab}, h_a) + \mathbf{F}_v + \mathbf{F}_B$$

How to achieve this in such a massively parallel code?

SWIFT uses **Task Templates**:

- User-defined Task Graph: identifying dependencies, e.g. compute smoothing length h_i **before** computing smoothing kernel $W(r_{ij}, h_i)$
- **Challenge**: integrate with GPUs

Challenges

- Massive parallelism
  - Need new algorithms for many common kernels
  - New parallel decompositions & exploit parallelism at all levels
  - Task-based parallelism
  - How can we use network accelerators efficiently?
- GPUs
  - Complex data structures are inefficient on GPUs
  - How can multiple MPI processes share GPU data portably?
  - How can we use multiple GPUs efficiently?
- We don’t have a UK (pre-)exascale machine…

Perspectives

- Exascale computing is coming!
- Exascale machines will be heterogeneous: CPU + Accelerators (GPU) + Interconnnect
- Presents 4 Key Challenges for SPH: Communications, GPU-enabled code, New algorithms, Separation of Concerns
- Methodologies and Results: Interesting / Promising so far …

Martyn Guest (ARCCA, Cardiff University)

Performance of Community Codes on Multi-core Processors. An Analysis of Computational Chemistry and Ocean Modelling Applications

Abstract: This session will overview the parallel benchmark performance of a variety of popular community codes on a number of HPC systems, with our analysis based on both computational chemistry and ocean modelling applications. The former feature codes from Molecular Dynamics (AMBER, LAMMPS and NAMD) and Materials Science (VASP, CASTEP), while representative codes from the ocean modelling community include NEMO and FVCOM.

The variety of systems considered focus on both the Intel Ice Lake and AMD EPYC Milan family of processors. Using the Intel Skylake Gold 6148 and AMD EPYC Rome 7502 as the baselines, an assessment is made across a variety of Ice Lake (8358, 8352Y, 8368Q, 8360Y and 8380) and Cascade Lake SKUs (e.g., the 9242-AP and 6248), with system interconnects from both NVIDIA Networks and Cornelis Networks. Attention is also focused on systems featuring the 64-core Milan and Rome AMD processor (the Rome 7702 and Milan 7713, 7763 & 7773X) and the corresponding 32-core processors (the Rome 7502 and Milan 7543 & 7573X).

The benefit of the Intel® oneAPI Toolkit is demonstrated throughout this analysis. To best capture a 'like for like' comparison amidst the extensive array of core densities, our analysis remains based on both a “node-by-node” and the more traditional “core-by-core” consideration.

Bio: 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 HPC Chemistry Group Leader at PNNL, before returning to the UK as Associate Director of Daresbury’s Computational Science and Engineering Department. Martyn joined Cardiff University in April 2007 as their Director of Advanced Research Computing. He is also Technical Director of the Supercomputing Wales programme and is co-I on the Isambard 2 system at the GW4 Tier-2 HPC regional centre.

Martyn’s 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 260 journal articles.

COMPUTING INSIGHT UK 2022

1 - 2 DECEMBER 2022
Manchester Central, UK
www.stfc.ac.uk/ciuk

UKRI Science and Technology Facilities Council

Performance of Community Codes on Multi-core Processors

An Analysis of Computational Chemistry and Ocean Modelling Applications.

Martyn Guest, Jose Munoz & Thomas Green

Advanced Research Computing @ Cardiff (ARCCA) &

Supercomputing Wales

Introduction and Overview

- Presentation part of our ongoing assessment of the performance of community codes on multi-core processors.
- Focus here on systems featuring **processors from AMD** (EPYC Milan SKUs) and **Intel** (Ice Lake SKUs) with IB and Cornelis Networks interconnects.
  - Baseline cluster: the Skylake (SKL) **Gold 6148/2.4 GHz** and **AMD EPYC Rome 7502 2.5Gz** cluster – "Hawk" – at Cardiff University.
  - **Five** Intel Xeon Ice Lake clusters, the 32-core Platinum **8358** (2.6 GHz) and **8352Y** (2.2 GHz), the 40-core **8380** (2.3 GHz), 38-core **8368Q** (2.6 GHz), 36-core **8360Y** (2.4GHz) plus other Cascade Lake & Cascade Lake-AP systems.
  - **Four** AMD EPYC Milan clusters featuring the 64-core **7713** (2.0 GHz) and **7773X** (2.2 GHz) and the 32-core **7543** (2.8 GHz) and **7573X** (2.8 GHz).
  - Consider performance of both synthetic and **end-user applications**. Latter include molecular simulation (**DL_POLY, AMBER**), materials modelling (**CASTEP, VASP**), & electronic structure (**GAMESS-UK**), plus representative ocean modelling codes including **NEMO** and **FVCOM**.
  - Scalability analysis by **processing elements (cores)** and by **nodes** (ARM Performance Reports). Baselined against **P100** & **V100** NVIDIA GPUs.

Methodology and Approach

1. Provide guidance based on evaluating performance that a **standard user** would experience on the systems
2. Target performance regime – **mid-range clusters**. No real effort invested in optimising the applications having used standard implementations when available
3. All benchmarks run on systems in general production i.e. not dedicated to this exercise – used standard Slurm job schedulers
4. CIUK'22 preparation again **challenging**. Target to evidence performance across a variety on MPI versions using variants of Intel Parallel Studio XE e.g. 2018/4, 2019/5, 2019/12 and 2020/4 and OneAPI proved over ambitious.
   - Number of problems encountered, particularly on **AMD Milan** systems
   - As noted before, working code using 2019/5 on AMD Rome systems failed on Milan, as did attempts to use earlier compilers / MPI libraries. 2019/12 worked on occasion but still led to failures with codes hanging at arbitrary core counts
   - **Intel oneapi resolved many of these** issues. But issues remain with performance compared to earlier variants of Intel Parallel Studio XE. e.g., a major decline in both VASP and CASTEP performance on AMD Rome when moving from "mpi/intel/2018/2" to "mpi/intel/2020/2"

AMD EPYC Milan multi-chip package

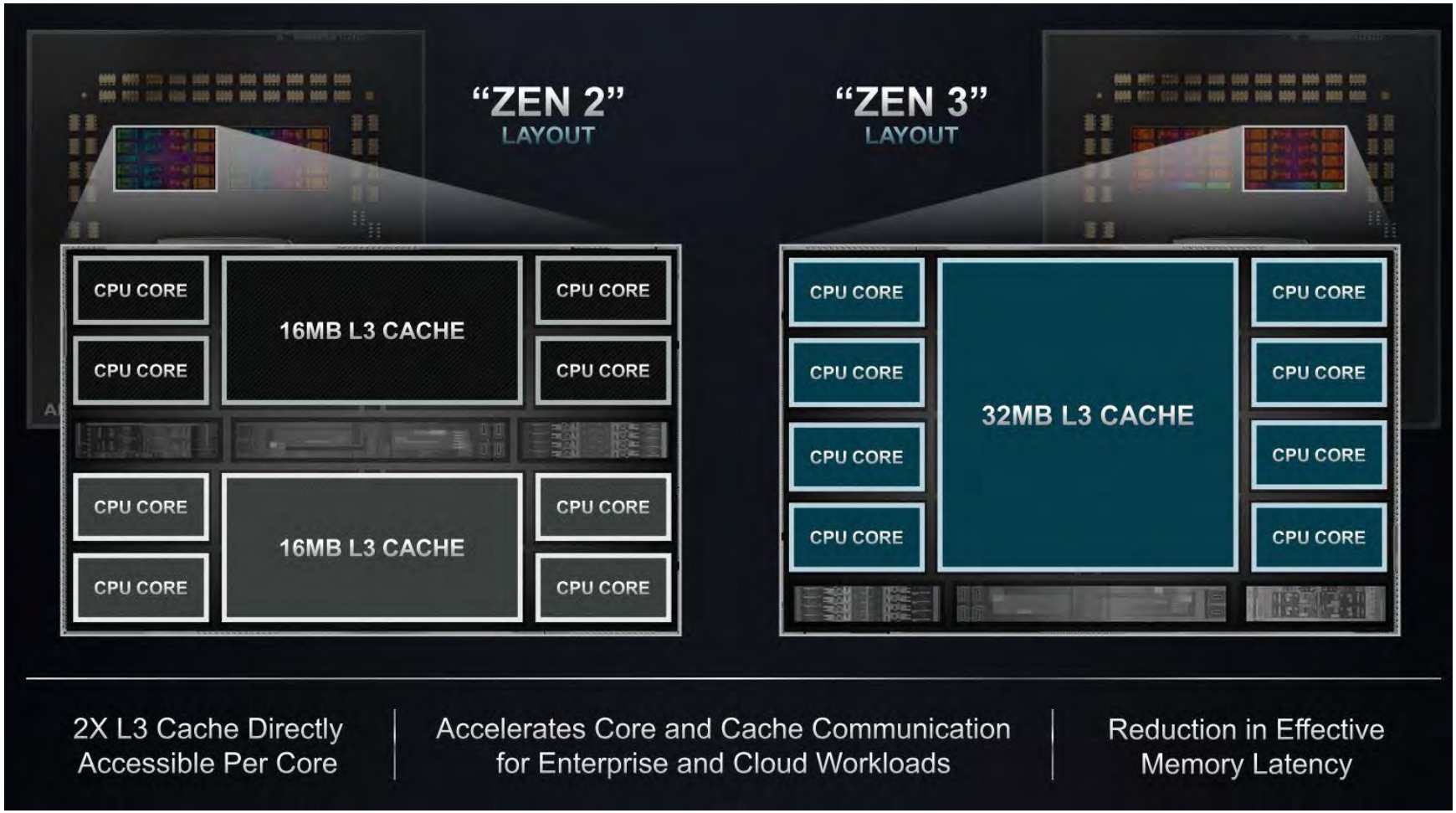


Figure. AMD EPYC Milan multi-chip package.

AMD EPYC Milan-X: Stacked L3 cache

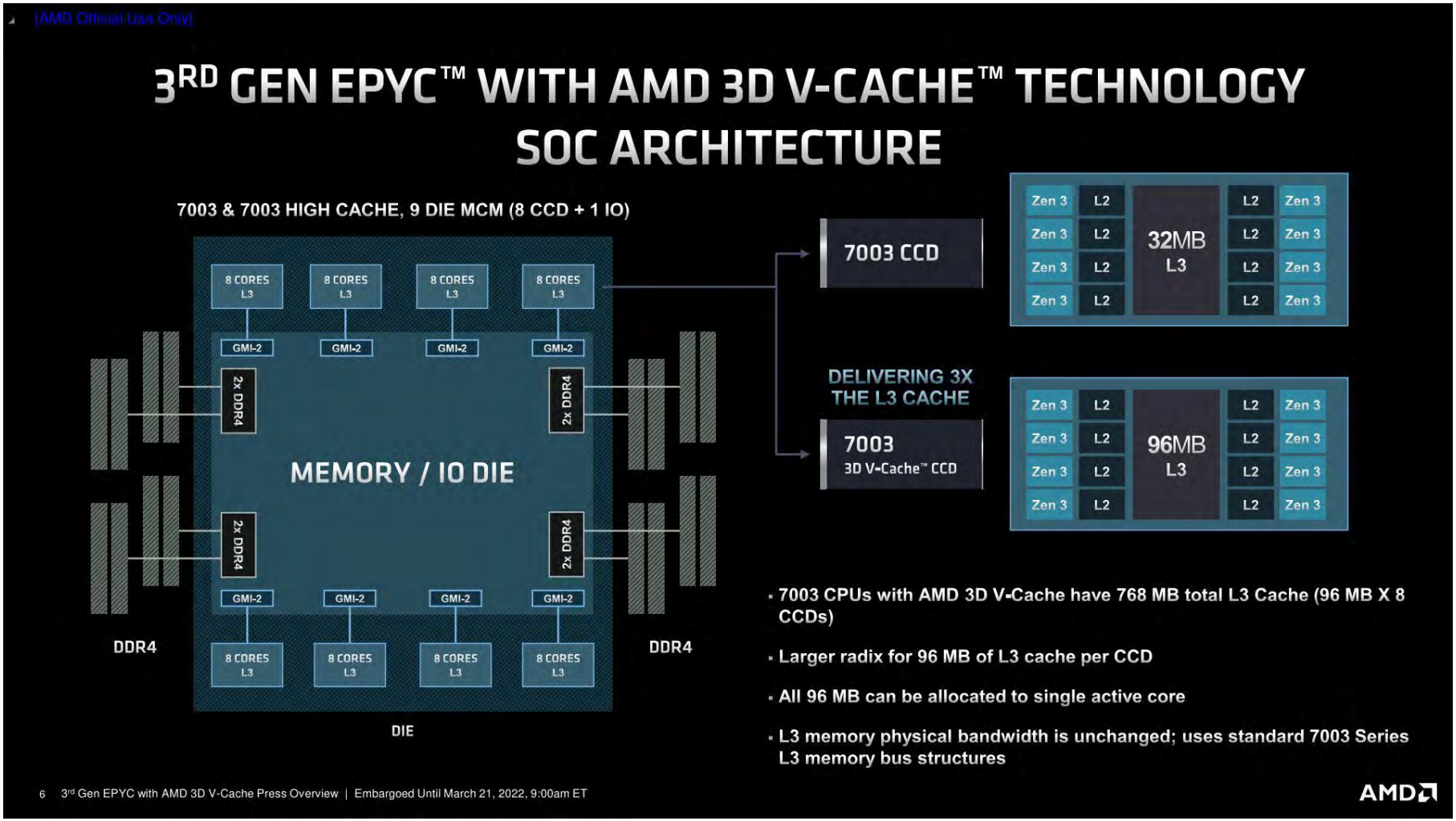
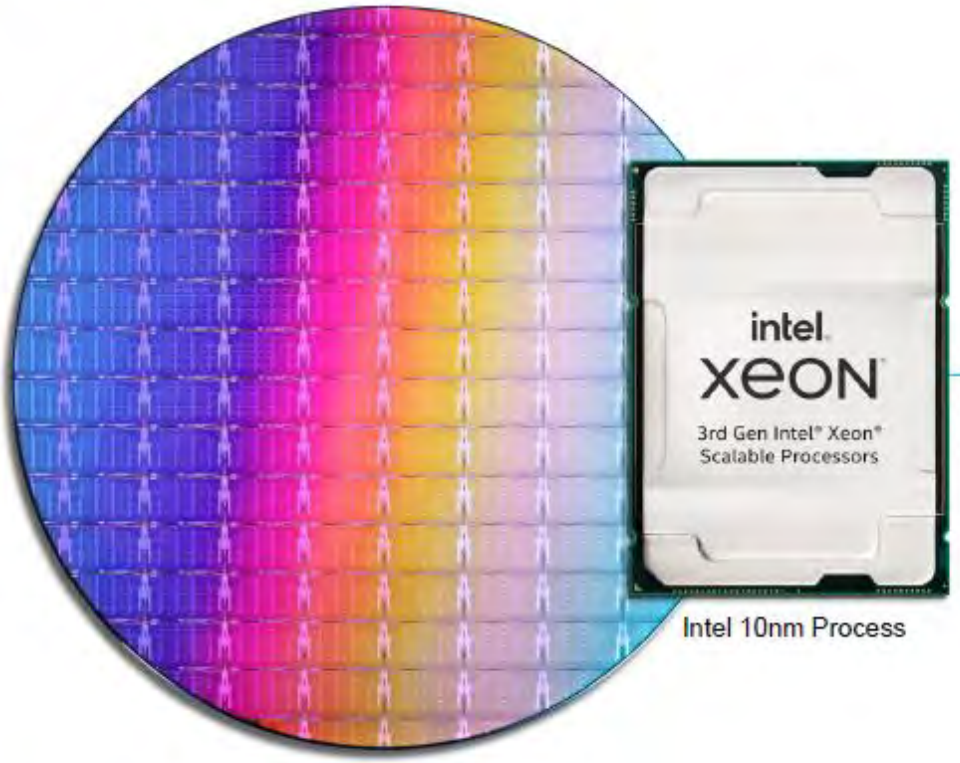


Figure. Milan-X: Upgraded version of Milan using the stacked L3 cache packaging technology. Will use added L3 cache on each chiplet, creating a processor with a total 768 MB of L3 cache e.g., AMD Milan 32c 7573X/2.8 GHz.

Intel 3rd Generation Xeon Scalable Processors

3rd Gen Intel® Xeon® Scalable processors

Performance made flexible



Up to 40 cores
per processor

20% IPC improvement
28 core, ISO Freq, ISO compiler

1.46x average performance increase
Geomean of Integer, Floating Point, Stream Triad, LINPACK
8380 vs. 8280

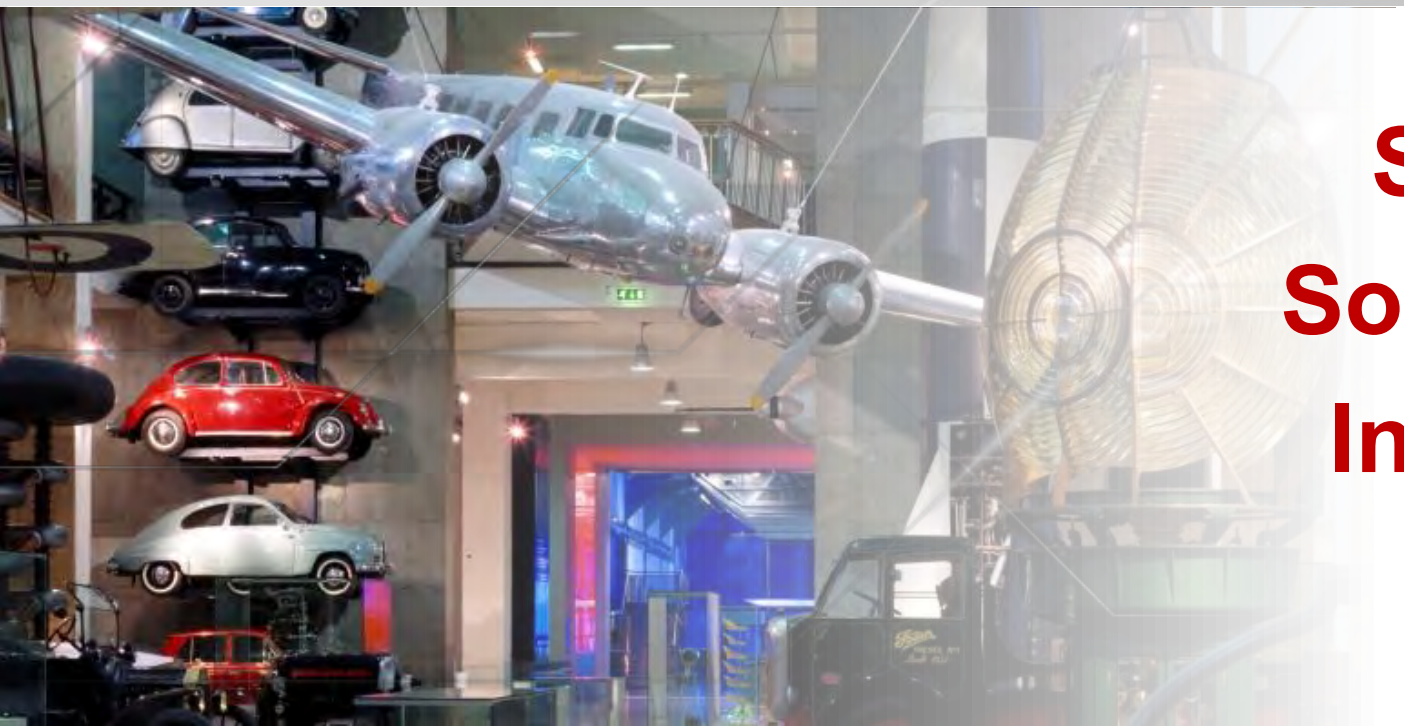
1.74x AI inference increase
8380 vs. 8280 BERT

2.65x average performance increase
vs. 5-year-old system
8380 vs. E5-2699v4

Intel Xeon Cascade Lake and Ice Lake

	Cascade Lake (per core)	Ice Lake (per core)
Out-of-order Window	224	352
In-flight Loads + Stores	72 + 56	128 + 72
Scheduler Entries	97	160
Register Files – Integer + FP	180 + 168	280 +224
Allocation Queue	64/thread	70/thread; 140/1 thread
L1D Cache (KB)	32	48
L2 Unified TLB (STLB)	1.5K	2K
STLB-IG Page support	**16**	**1024 (shared w/4K)**
STLB-IG Page support	**16**	**1024 (shared w/4K)**
Mid-level Cache (MB)	1	1.25

Performance of Computational Chemistry and Ocean Modelling Codes



Systems, Software and Installation

Baseline Cluster System

Supercomputing Wales "Hawk" Cluster Configuration

"Phase-1" - Intel Skylake Partition	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.
"Phase-2" AMD Rome Partition	64 nodes, totalling 4,096 cores, 32 TB total memory. • CPU: 2 x AMD EPYC Rome 7502 CPU @ 2.50GHz with 32 cores each; RAM: 512 GB, and GPU nodes; GPU: 30 x nVidia V100 GPUs with 16GB of RAM on 15 nodes
Researcher Funded Partitions	• 4,616 cores – Intel Skylake dedicated researcher expansion • 2,064 cores – Intel Broadwell and Haswell Raven migrated sub-system nodes

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

Intel Xeon Cascade Lake Clusters

Cluster / Configuration
Dell Zenith cluster at the Dell Technologies HPC & AI Innovation Lab – Intel Xeon sub-systems with Mellanox HDR interconnect fabric running Slurm. Aging systems that were subject to withdrawal from service, impacting on # nodes available.
• Intel **Xeon Gold 6248 Processor / 2.50 GHz**; # of CPU Cores: **20**; # of Threads: 40; Max Turbo Frequency: 3.90 GHz Base Clock: **2.50 GHz;** Cache 27.5 MB**;** Default TDP / TDP: 150W; Mellanox HDR **200Gb/s**
• Intel **Xeon Platinum 8280 Processor / 2.70 GHz**; # of CPU Cores: **28**; # of Threads: 56; Max Turbo Frequency: 4.00 GHz Base Clock: **2.70 GHz;** Cache 38.5 MB**;** Default TDP / TDP: 205W; Mellanox HDR **200Gb/s**
Cascade Lake-AP cluster at Intel HPC Laboratory with Cornelis OPE fabric running Bright release 8.1, optane filesystem.
• 48 **CLX-AP nodes** (Cascade Lake Advanced Performance**) 9242 processors / 2.3 GHz**; # of CPU Cores: **48**; # of Threads: 96; Max Turbo Frequency: 3.80 GHz Base Clock: **2.30 GHz;** Cache 71.5 MB**;** Default TDP / TDP: 350W

Intel Xeon Ice Lake Clusters

Cluster / Configuration
Dell Zenith cluster at the Dell Technologies HPC & AI Innovation Lab – Intel Xeon sub-systems with Mellanox HDR interconnect fabric running Slurm
• 50 nodes × Intel **Xeon Platinum 8358 Processor / 2.60 GHz**; # of CPU Cores: **32**; # of Threads: 64; Max Turbo Frequency: 3.40 GHz Base Clock: **2.60 GHz;** Cache 48 MB**;** Default TDP / TDP: 250W; **Mellanox HDR 200Gb/s**
• 70 nodes × Intel **Xeon Platinum 8352Y Processor / 2.20 GHz**; # of CPU Cores: **32**; # of Threads: 64; Max Turbo Frequency: 3.40 GHz Base Clock: **2.20 GHz;** Cache 48 MB**;** Default TDP / TDP: 205W; **Mellanox HDR 200Gb/s**
Ice Lake clusters at Intel's OpenHPC Laboratory with Cornelis OPE fabric running Bright release 8.1 and optane filesystem.
• 4 nodes × Intel **Xeon Platinum 8368Q Processor / 2.60 GHz**; # of CPU Cores: **38**; # of Threads: 76; Max Turbo Frequency: 3.70 GHz Base Clock: **2.60 GHz;** Cache 57 MB**;** Default TDP / TDP: 270W; **Cornelis OPE**
• 4 nodes × Intel **Xeon Platinum 8360Y Processor / 2.40 GHz**; # of CPU Cores: **36**; # of Threads: 72; Max Turbo Frequency: 3.50 GHz Base Clock: **2.40 GHz;** Cache 54 MB**;** Default TDP / TDP: 270W; **Cornelis OPE**
Intel's Endeavour cluster with Cornelis OPE fabric running Slurm
• 8 nodes × Intel **Xeon Platinum 8380 Processor / 2.30 GHz**; # of CPU Cores: **40**; # of Threads: 80;
• 10 nodes × Intel **Xeon Platinum 8360Y Processor / 2.40 GHz**; # of CPU Cores: **36**; # of Threads: 72

AMD EPYC Milan Clusters

Cluster / Configuration
Dell Minerva cluster at the Dell Technologies HPC & AI Innovation Lab – AMD EPYC Rome and Milan sub-systems with Mellanox HDR interconnect fabric running Slurm
• **4 nodes × AMD EPYC Milan 7543 / 2.80 GHz**; # of CPU Cores: 32; # of Threads: 64; Max Boost Clock: 3.7 GHz Base Clock: **2.80 GHz; L3** Cache 256 MB**;** Default TDP / TDP: 225W; Mellanox HDR-100 **200Gb/s**
• **6 nodes × AMD EPYC Milan 7573X / 2.80 GHz**; # of CPU Cores: 32; # of Threads: 64; Max Boost Clock: 3.6 GHz Base Clock: **2.80 GHz; L3 Cache 768 MB;** Default TDP / TDP: 280W; Mellanox HDR-100 **200Gb/s**
• **170 nodes × AMD EPYC Milan 7713 / 2.00 GHz**; # of CPU Cores: 64; # of Threads: 128; Max Boost Clock: 3.675 GHz Base Clock: **2.00 GHz; L3** Cache 256 MB**;** Default TDP / TDP: 225W; Mellanox HDR-100 **200Gb/s**
• **4 nodes × AMD EPYC Milan 7763 / 2.45 GHz**; # of CPU Cores: 64; # of Threads: 128; Max Boost Clock: 3.5 GHz Base Clock: **2.45 GHz; L3** Cache 256 MB**;** Default TDP / TDP: 280W; Mellanox HDR-100 **200Gb/s**
SPARTAN cluster at the Atos HPC, AI & QLM Benchmarking Centre – AMD EPYC Rome system with Mellanox ConnectX-6 HDR100 interconnect fabric
• **240 × AMD EPYC Rome 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 100 InfiniBand:** Memory: 256GB DDR4 2677MHz RDIMMs per node: **DDN lustre 7990 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 and AMBER (molecular dynamics)

VASP and CASTEP (ab initio Materials properties)

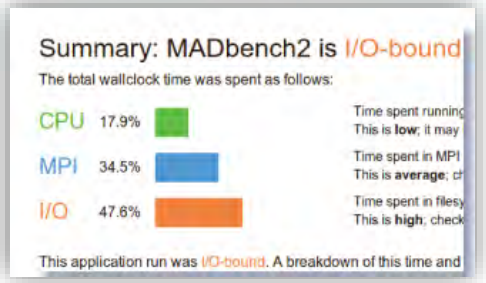
GAMESS-UK (molecular electronic structure)

FVCOM and NEMO (ocean modelling codes)

- 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.

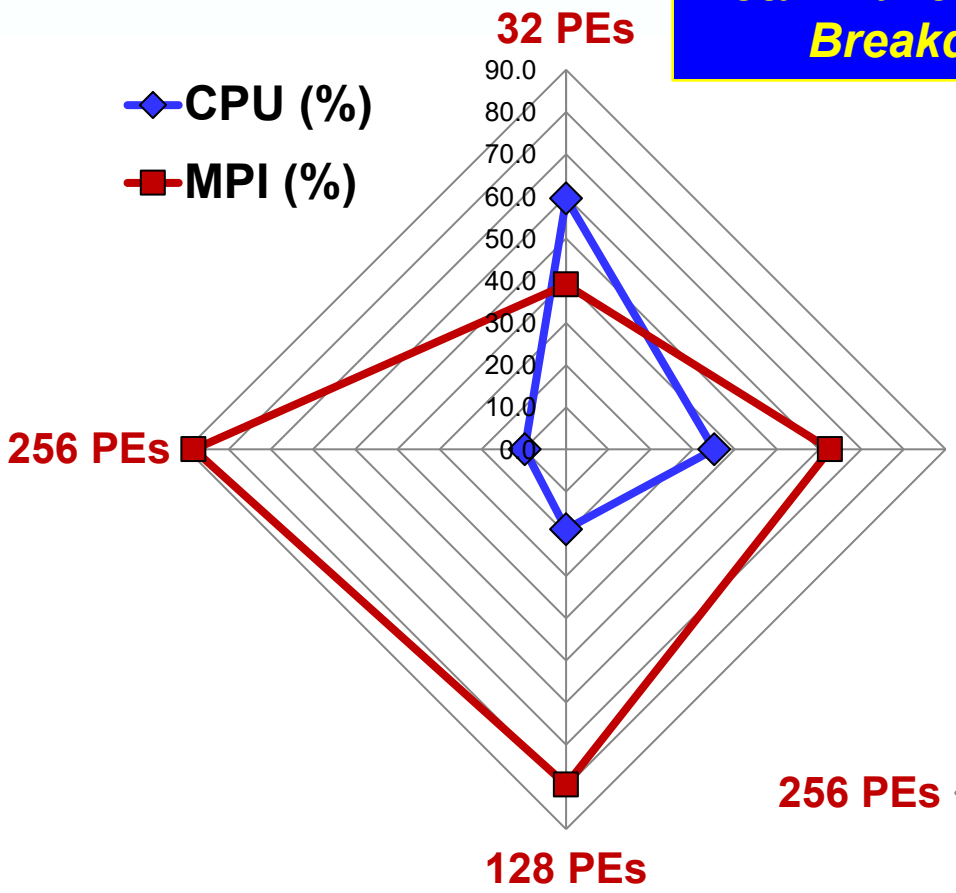
  perf-report mpirun $code

- ***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)**

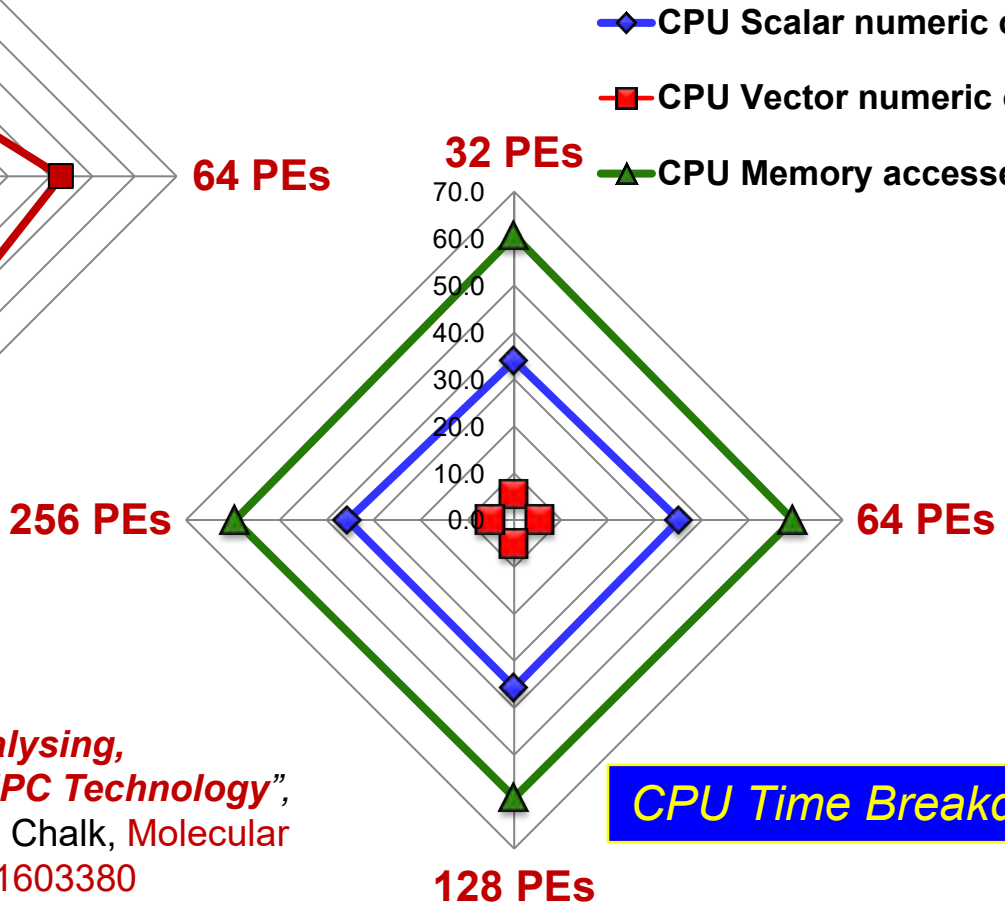
DLPOLY4 – Performance Report

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

STREAM (AMD Minerva 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
```

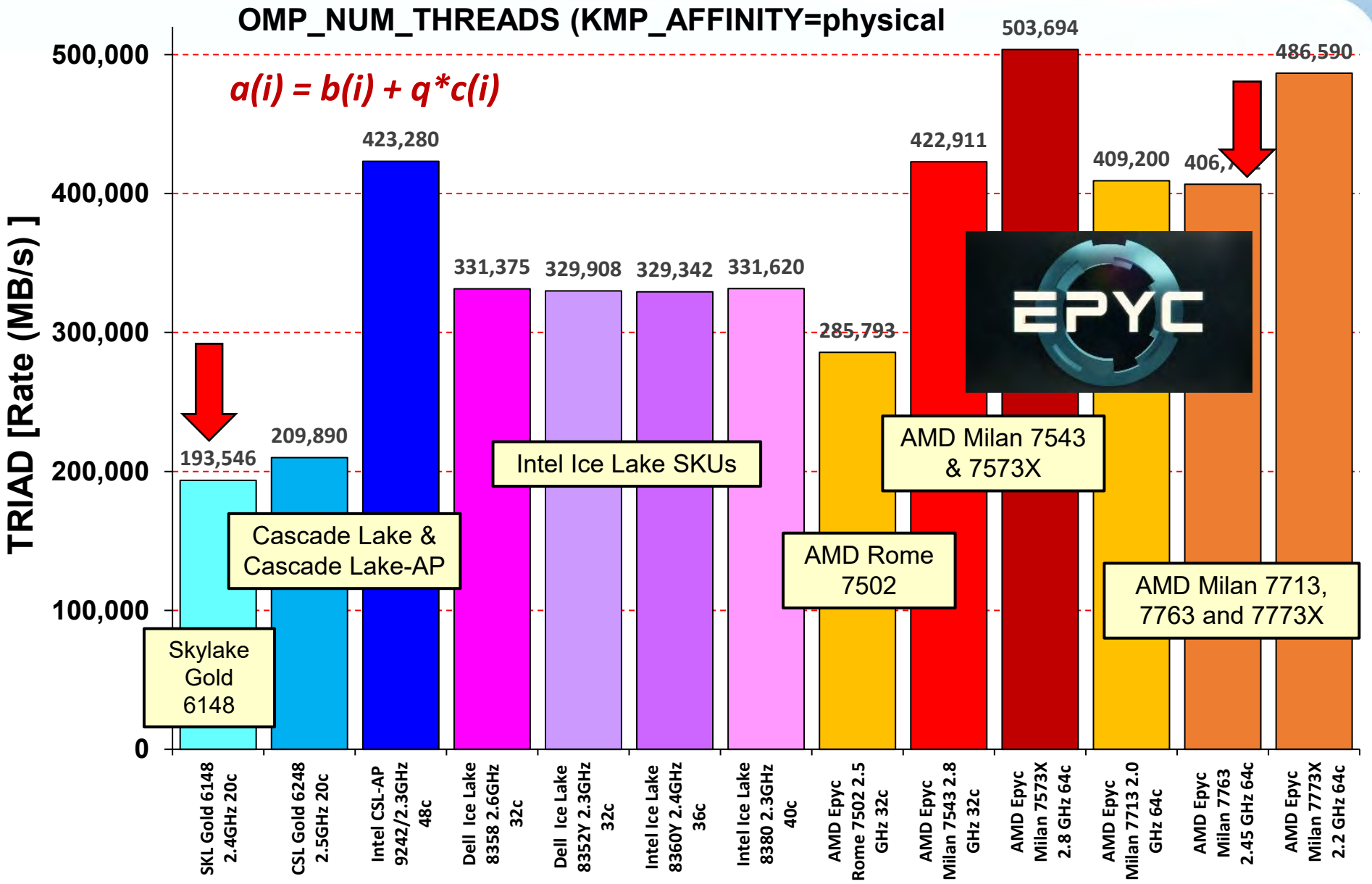
```
# Version of Intel compiler to use and way to source it
source /opt/intel/compilers_and_libraries_2020.2.254/linux/bin/compilervars.sh -
ofi_internal=1 intel64

# Increasing use of oneAPI: e.g., source /opt/intel/oneapi/setvars.sh

# Use of specific version of Intel MKL, further versions do not allow the setting
of AVX2 on non-Intel processors.
source /opt/intel/compilers_and_libraries_2019.6.324/linux/mkl/bin/mklvars.sh
intel64

# When using IntelMPI on AMD Rome/Milan
export I_MPI_FABRICS=shm:ofi
export I_MPI_SHM=clx_avx2
export FI_PROVIDER=mlx

# On AMD Rome/Milan when using Intel MKL
export MKL_DEBUG_CPU_TYPE=5
```

Compilation:

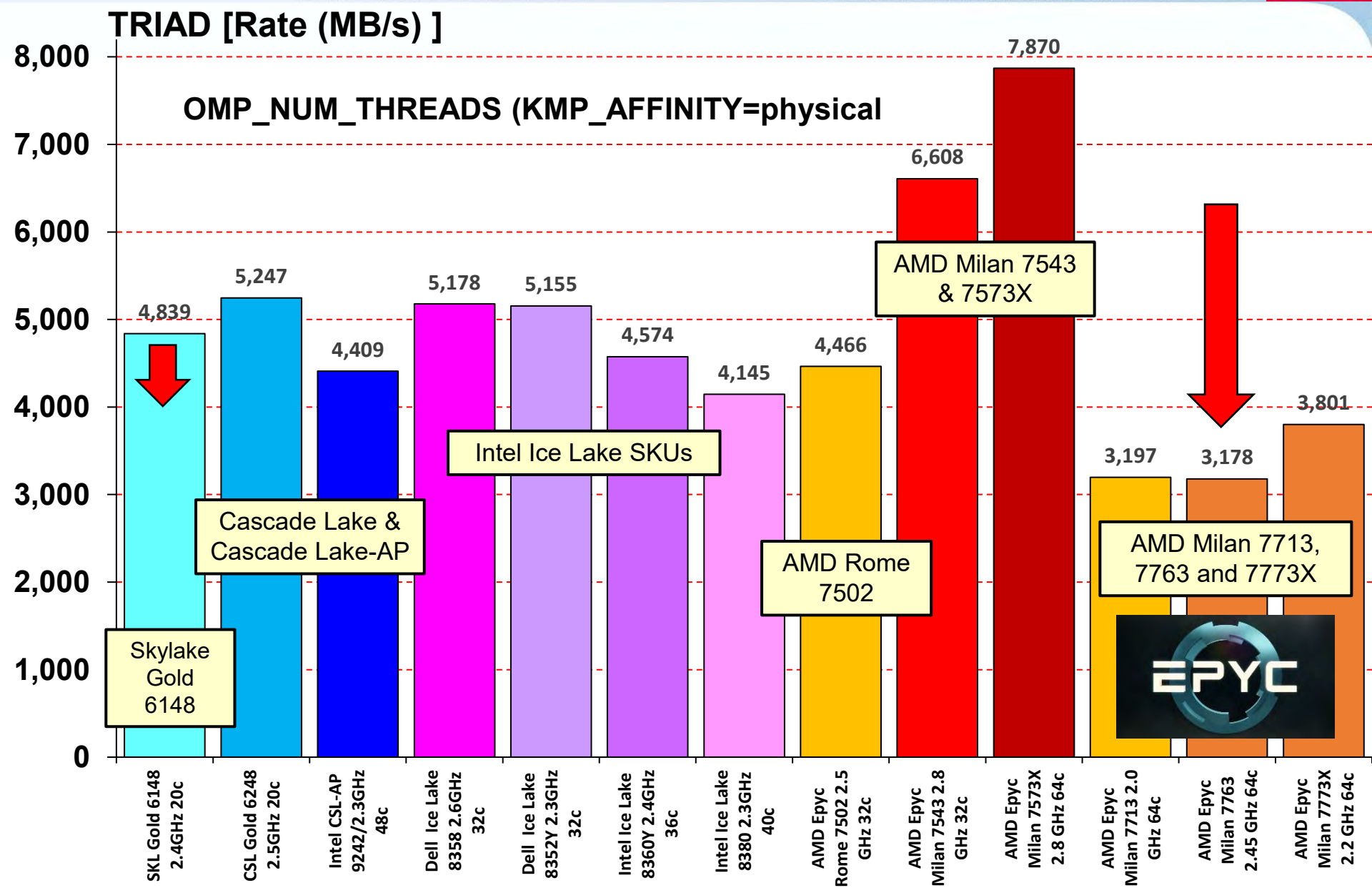
INTEL SKL: –O3 –xCORE-AVX512

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

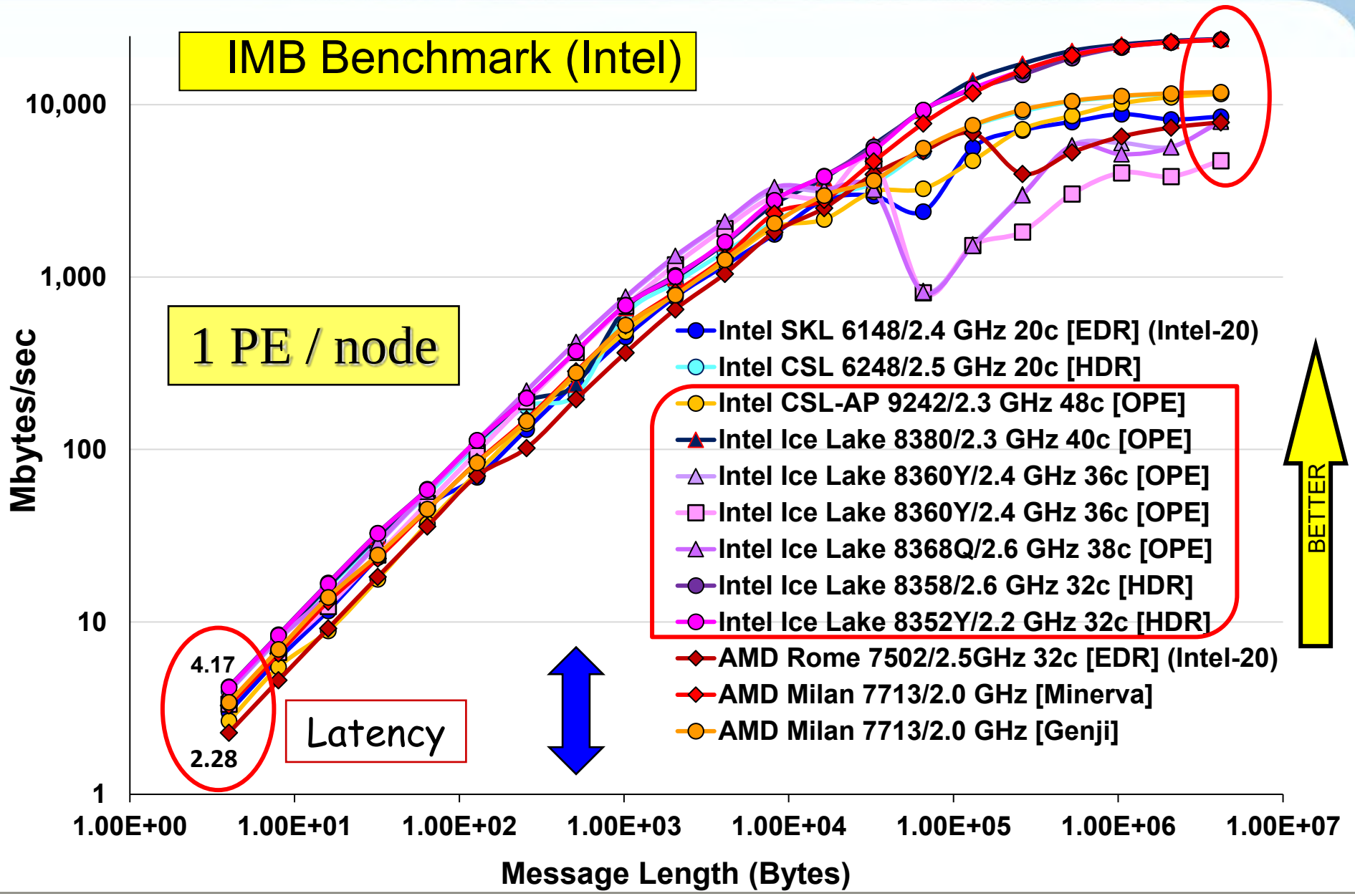
Memory B/W – STREAM performance



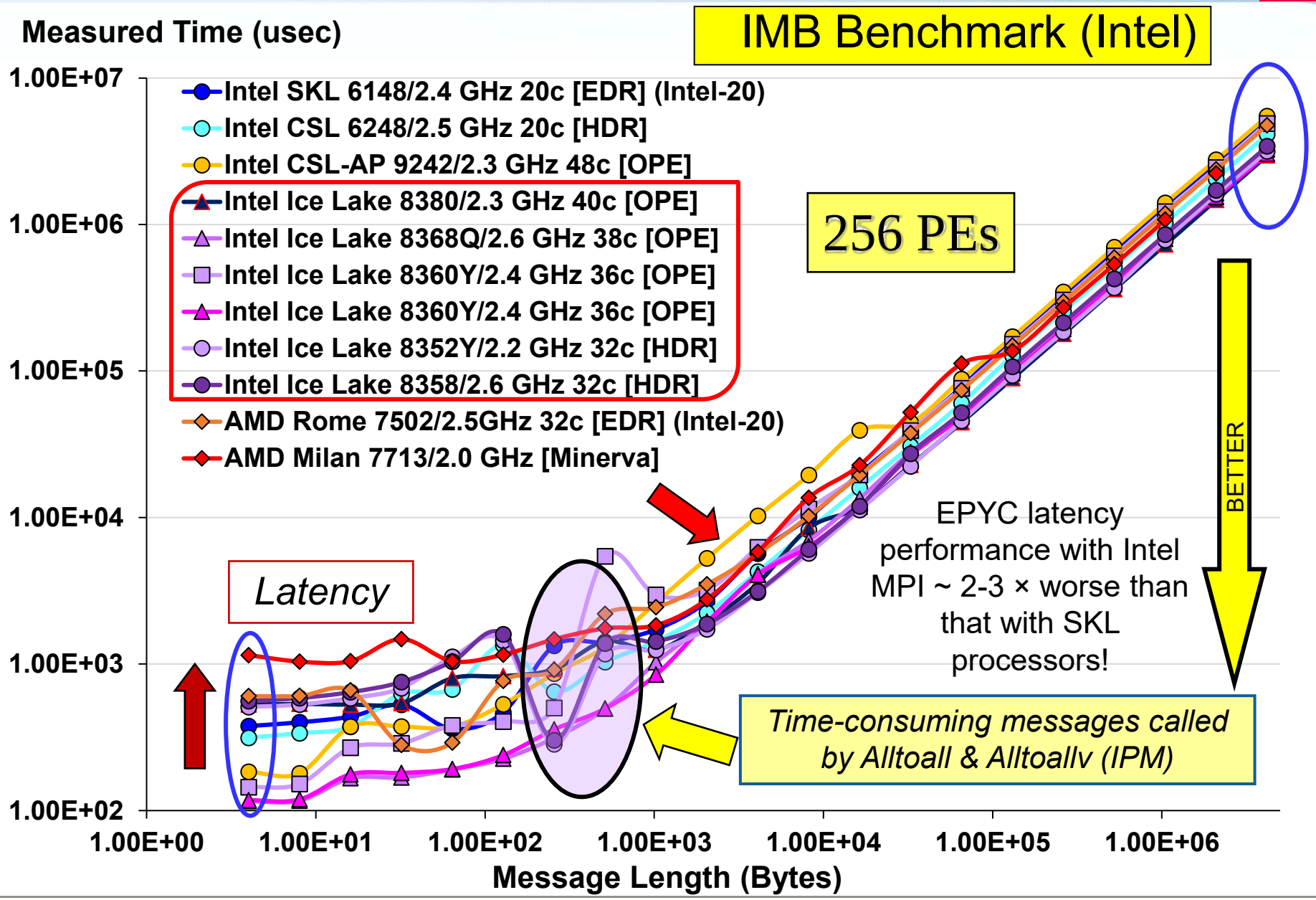
Memory B/W – STREAM / core performance



MPI Performance – PingPong



MPI Collectives – Alltoallv (256 PEs)



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 a variety of "**node to node**" and "**core-to-core**" comparisons e.g., :

1	***Hawk - Dell |EMC Skylake Gold 6148 2.4GHz (T) EDR with 40 cores / node***	***AMD EPYC Milan 7713 nodes with128 cores per node. [1-7 nodes]***
2	***Hawk - Dell |EMC Skylake Gold 6148 2.4GHz (T) EDR with 40 cores / node***	***Intel Xeon Platinum Ice Lake 8358 nodes with 64 cores per node. [1-7 nodes]***

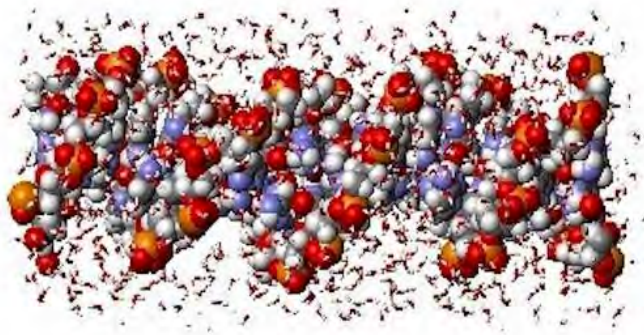
Performance of Computational Chemistry and Ocean Modelling Codes

Molecular Simulation;

1. DL_POLY

Molecular Simulation I. DL_POLY

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



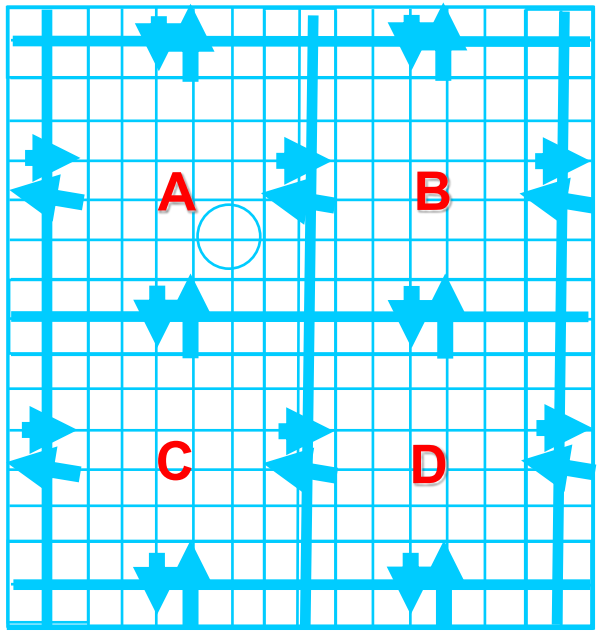
DL_POLY

- 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.

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

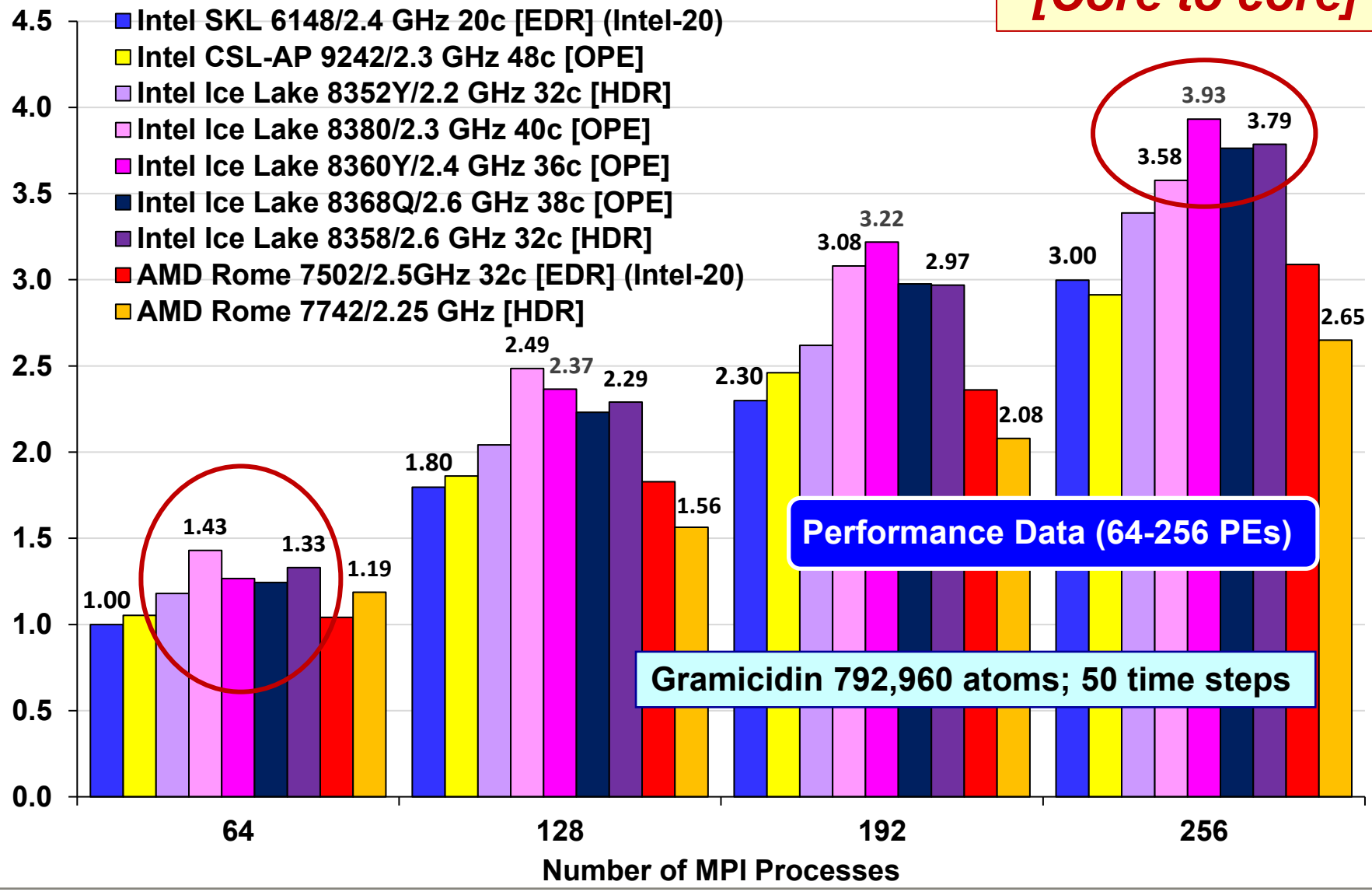
http://www.scd.stfc.ac.uk//research/app/ccg/software/DL_POLY/44516.aspx

DL_POLY 4 – Gramicidin Simulation

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

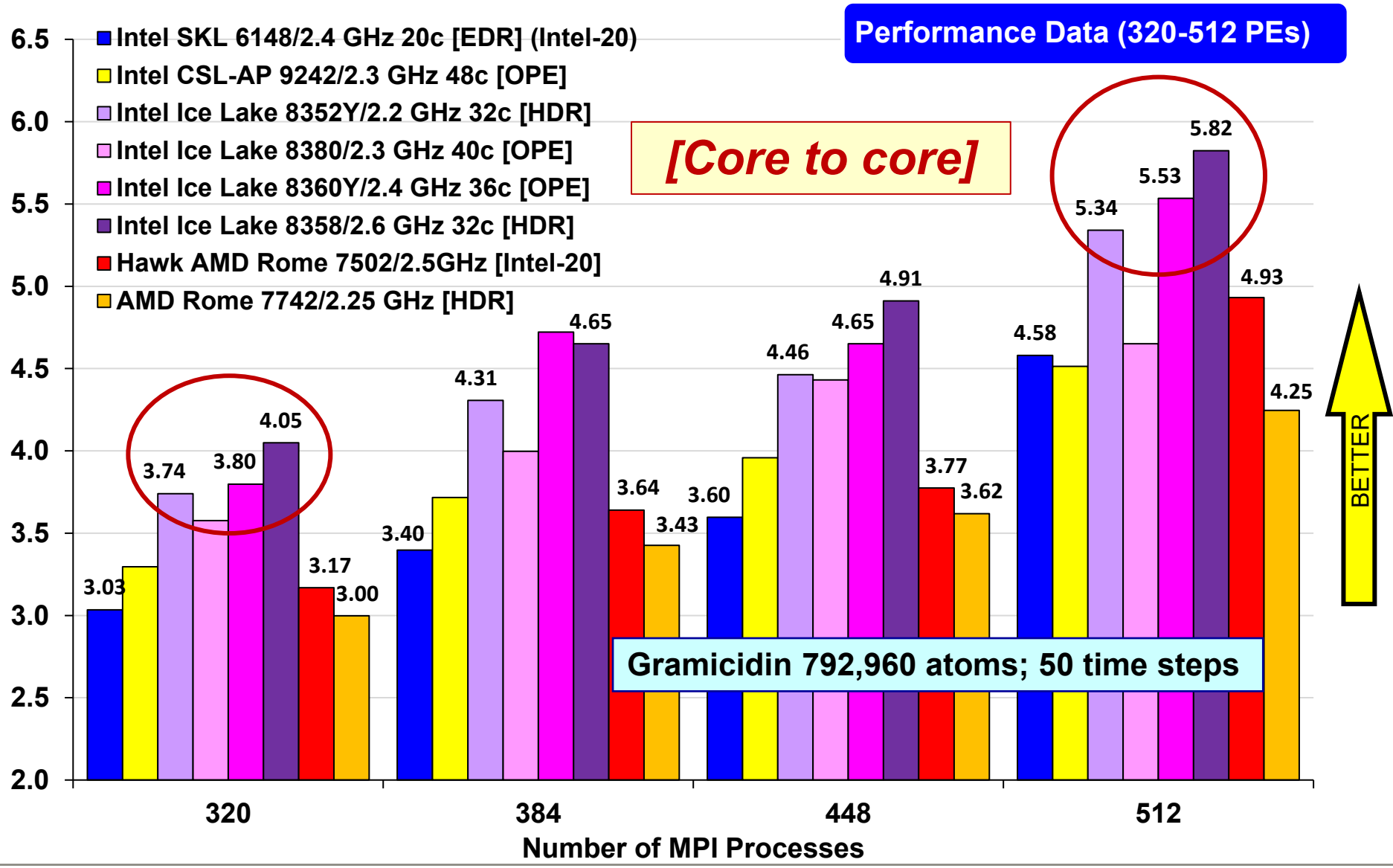
Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*

[Core to core]

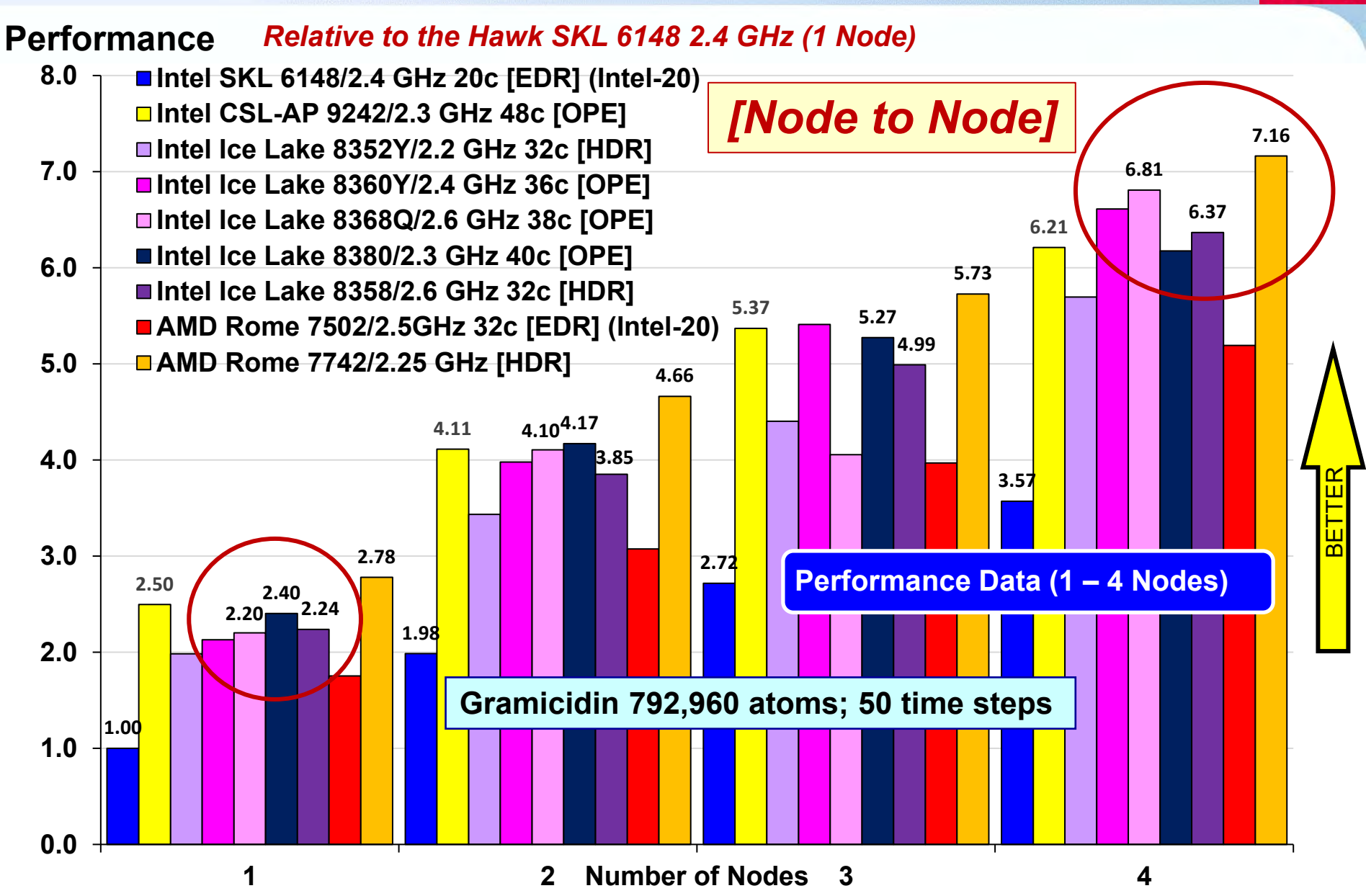


DL_POLY 4 – Gramicidin Simulation

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

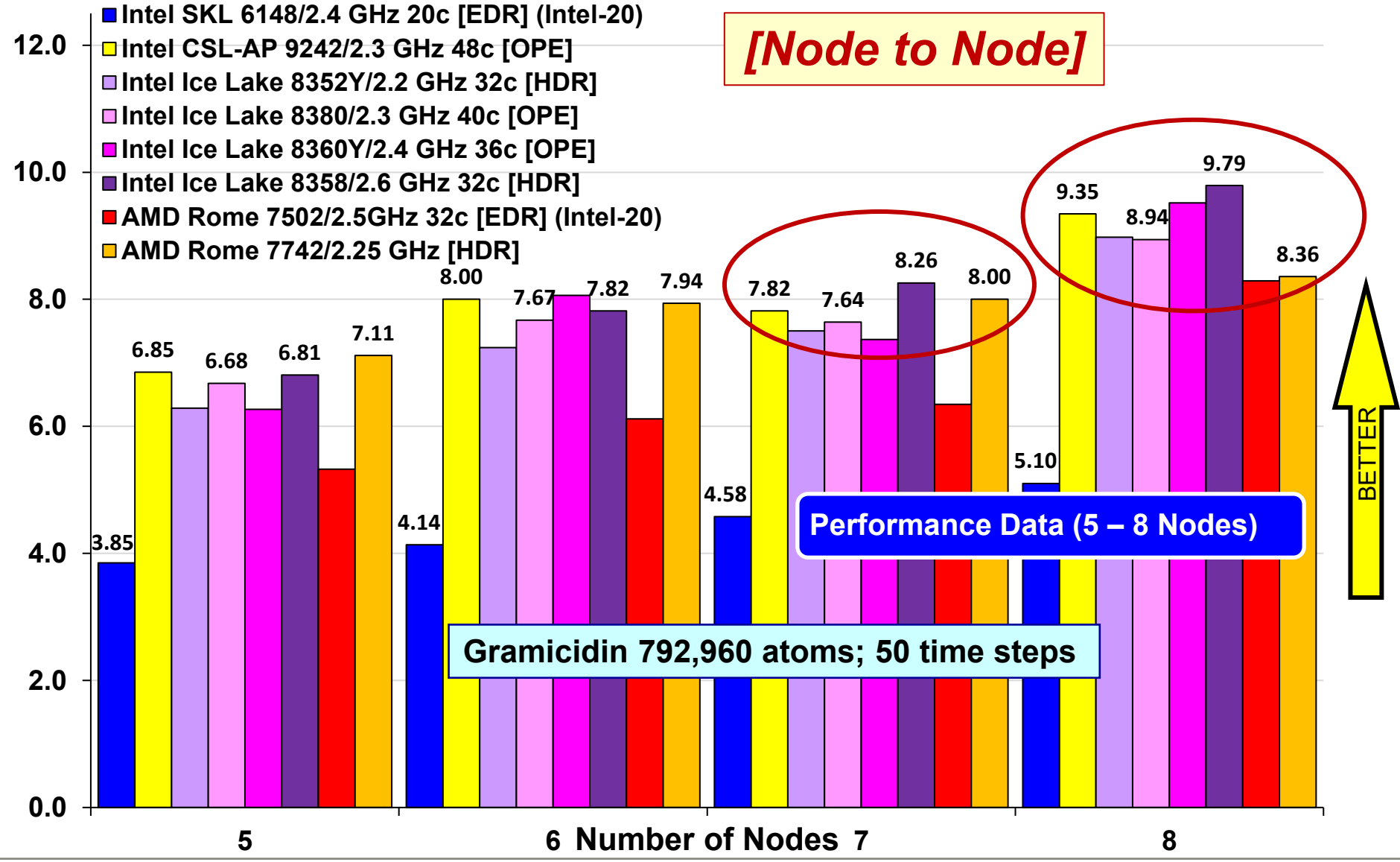


DL_POLY 4 – Gramicidin Simulation

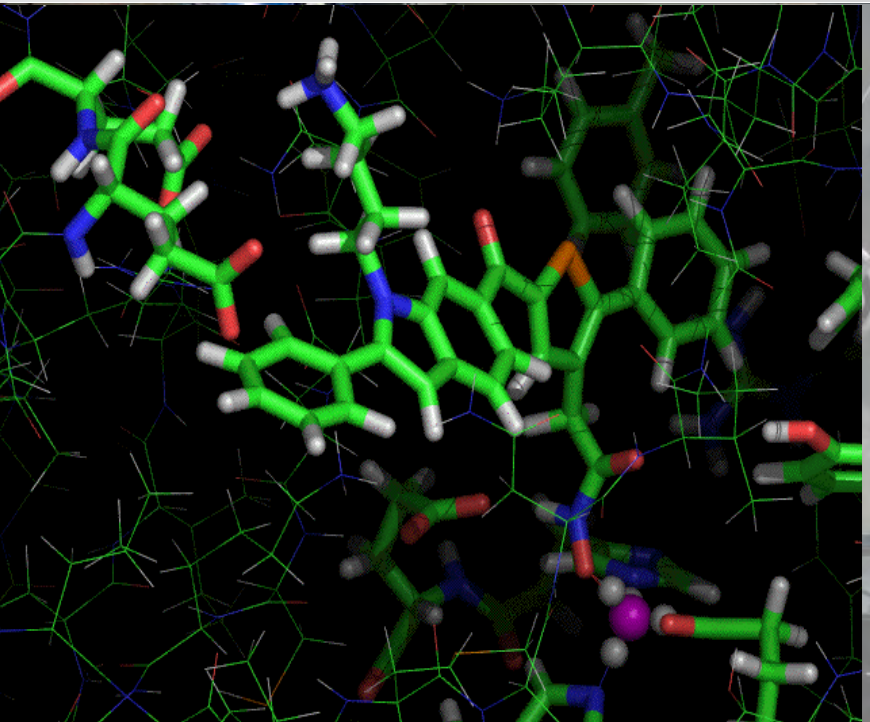


DL_POLY 4 – Gramicidin Simulation

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

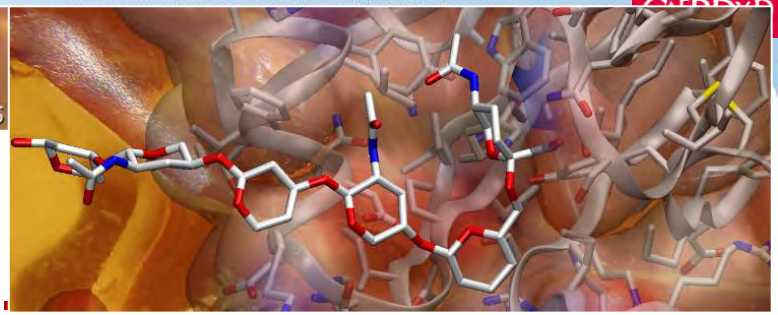


Performance of Computational Chemistry and Ocean Modelling Codes



Molecular Simulation: 2. AMBER

Molecular Simulation II. AMBER



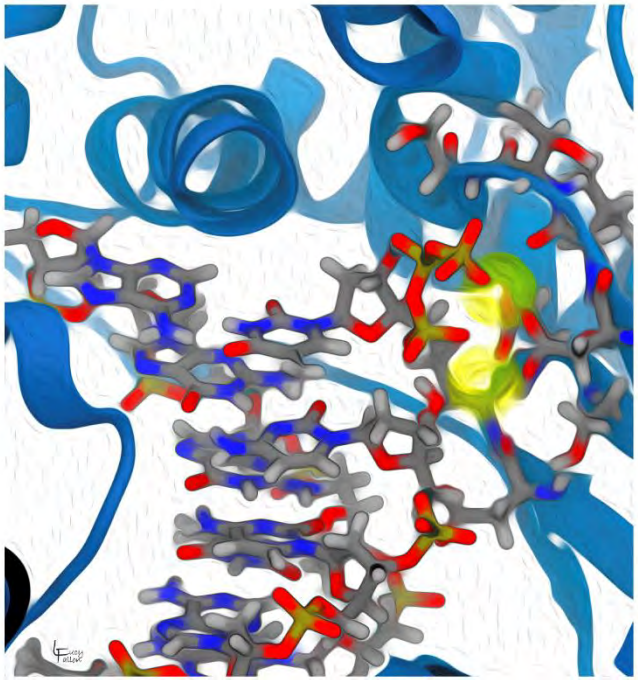
- AMBER18 and AMBER22 used:
  PMEMD & GPU accelerated PMEMD.
- **M01 Benchmark**
  - Major Urinary Protein (MUP) + IBM ligand (21,736 atoms)
- **M06 Benchmark**
  - Cluster of six MUPs (134,013 atoms)
- **M27 Benchmark**
  - **Cluster of 27 MUPs (657,585 atoms)**
- **M45 Benchmark**
  - **Cluster of 45 MUPs (932,751 atoms)**

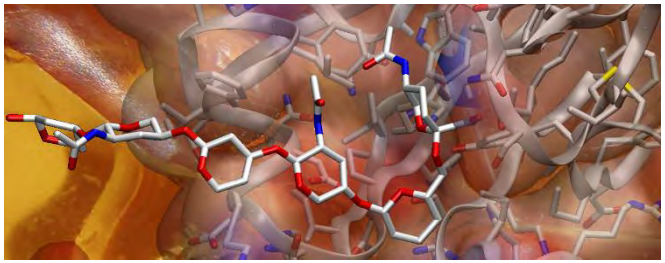
*All test cases run 30,000 steps * 2fs = 60ps simulation time. Periodic boundary conditions, constant pressure, T=300K. Position data written every 500 steps.*

R. Salomon-Ferrer, D.A. Case, R.C. Walker. An overview of the Amber biomolecular simulation package. WIREs Comput. Mol. Sci. 3, 198-210 (2013).

D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. Woods. The Amber biomolecular simulation programs. J. Computat. Chem. 26, 1668-1688 (2005).

Molecular Simulation II. AMBER

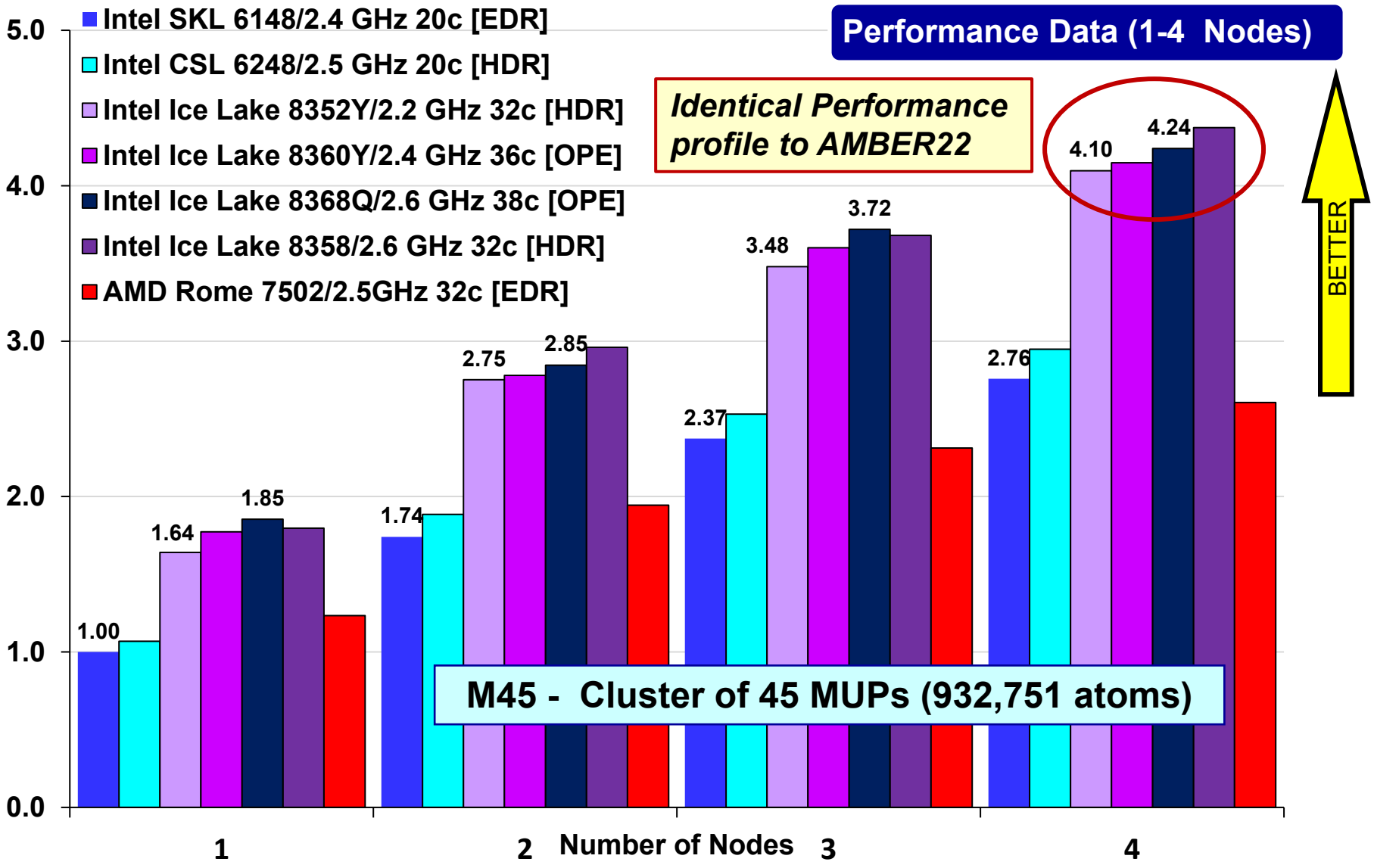




- **AMBER22** released (on April 27, 2022).
- The Amber22 package builds on AmberTools22 by adding the pmemd program, which resembles the sander (MD) code in AmberTools, but provides better performance on multiple CPUs, and dramatic speed improvements on GPUs.
- **AMBER18** (released in 2018) also used in this study. In practice we find **also identical performance of the two code releases** when running the M01, M06, M27 and M45 performance test cases.
- Presentation limited to the **M27 and M45** test cases for M01 and M06 are now too small for meaningful analysis

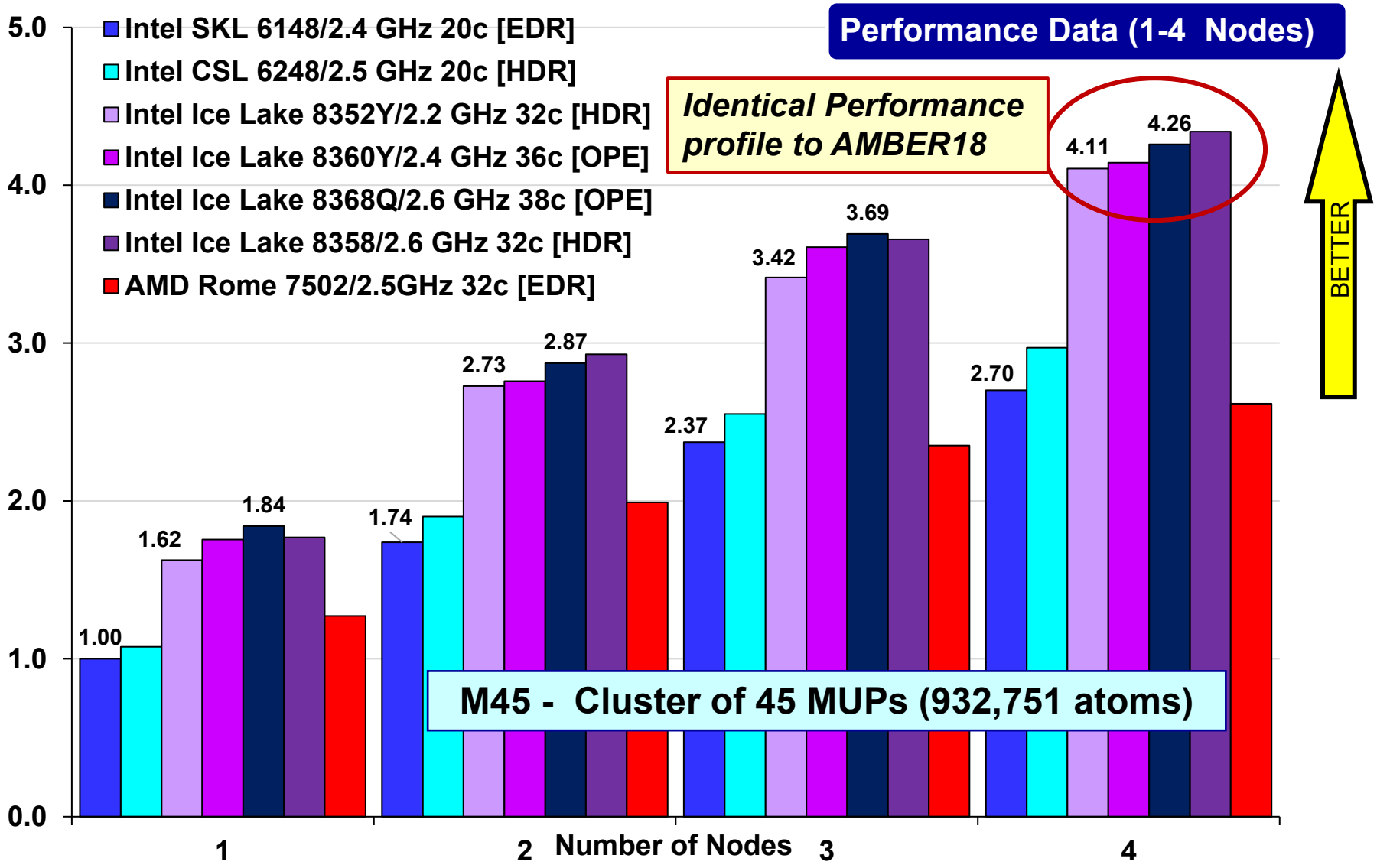
AMBER18 - M45 Performance results

Performance *Relative to the Hawk SKL 6148 2.4 GHz (40 PEs)*



AMBER22 - M45 Performance results

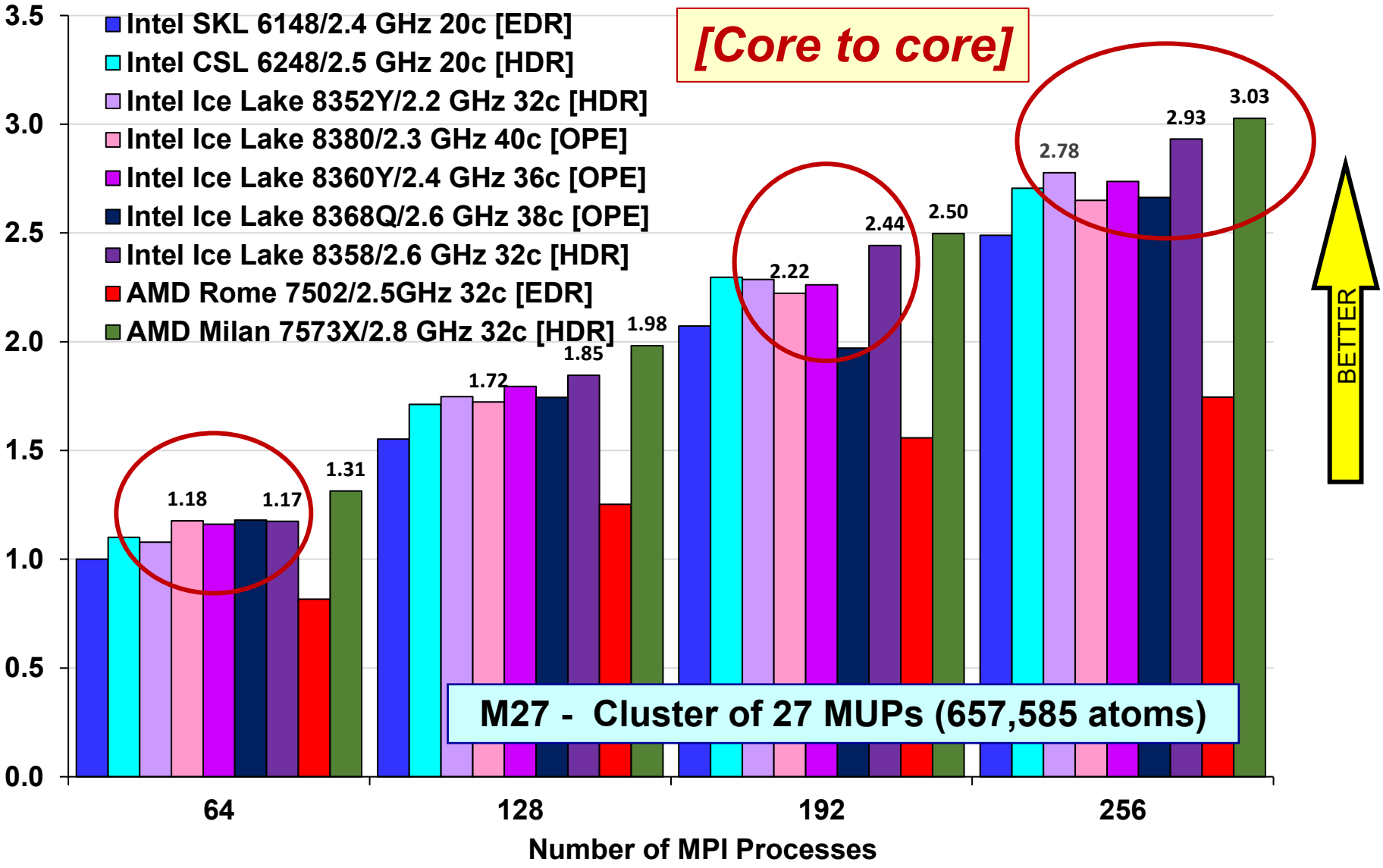
Performance *Relative to the Hawk SKL 6148 2.4 GHz (40 PEs)*



AMBER18 - M27 Performance Analysis

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

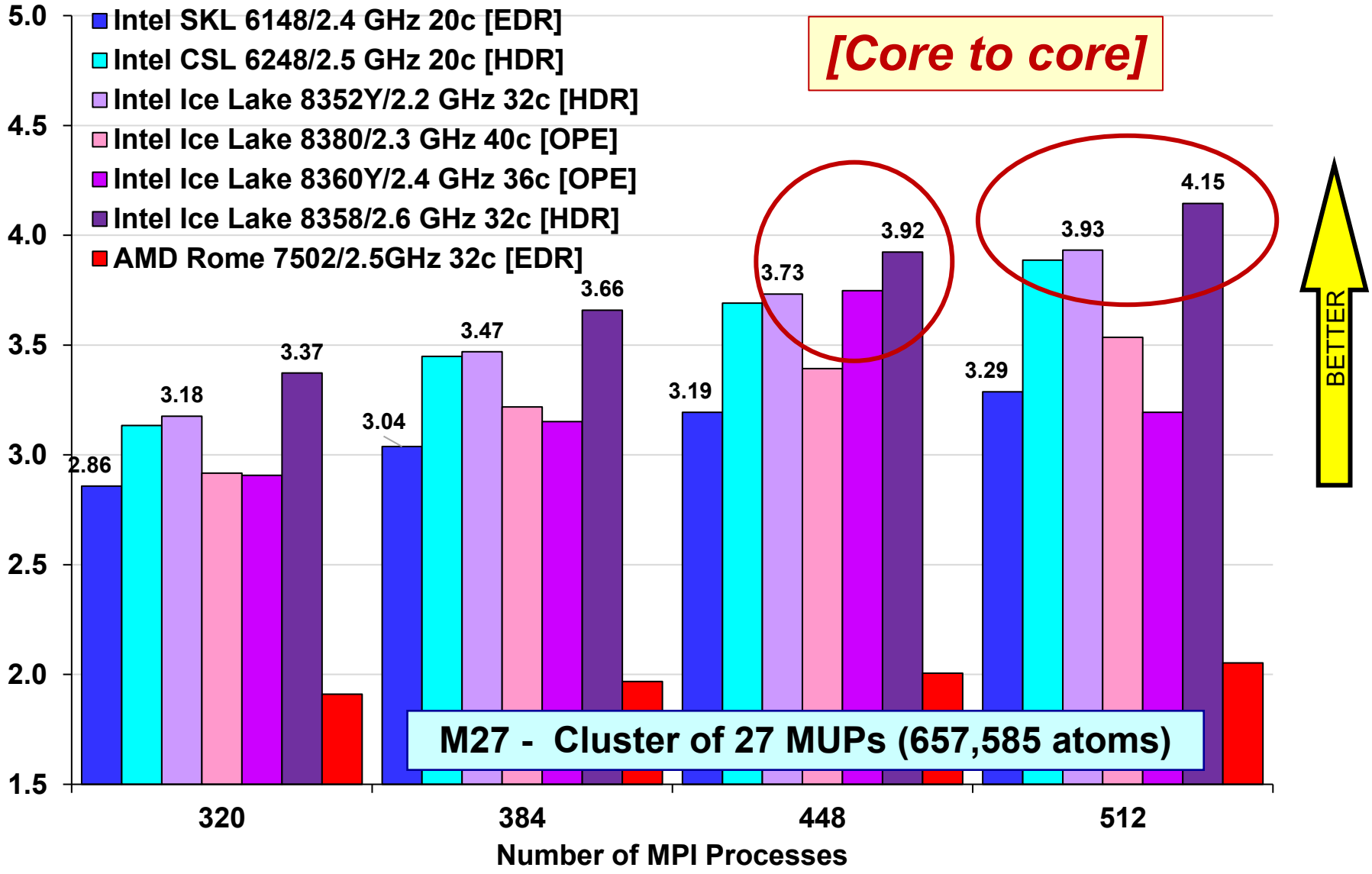
Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*



AMBER18 - M27 Performance Analysis

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

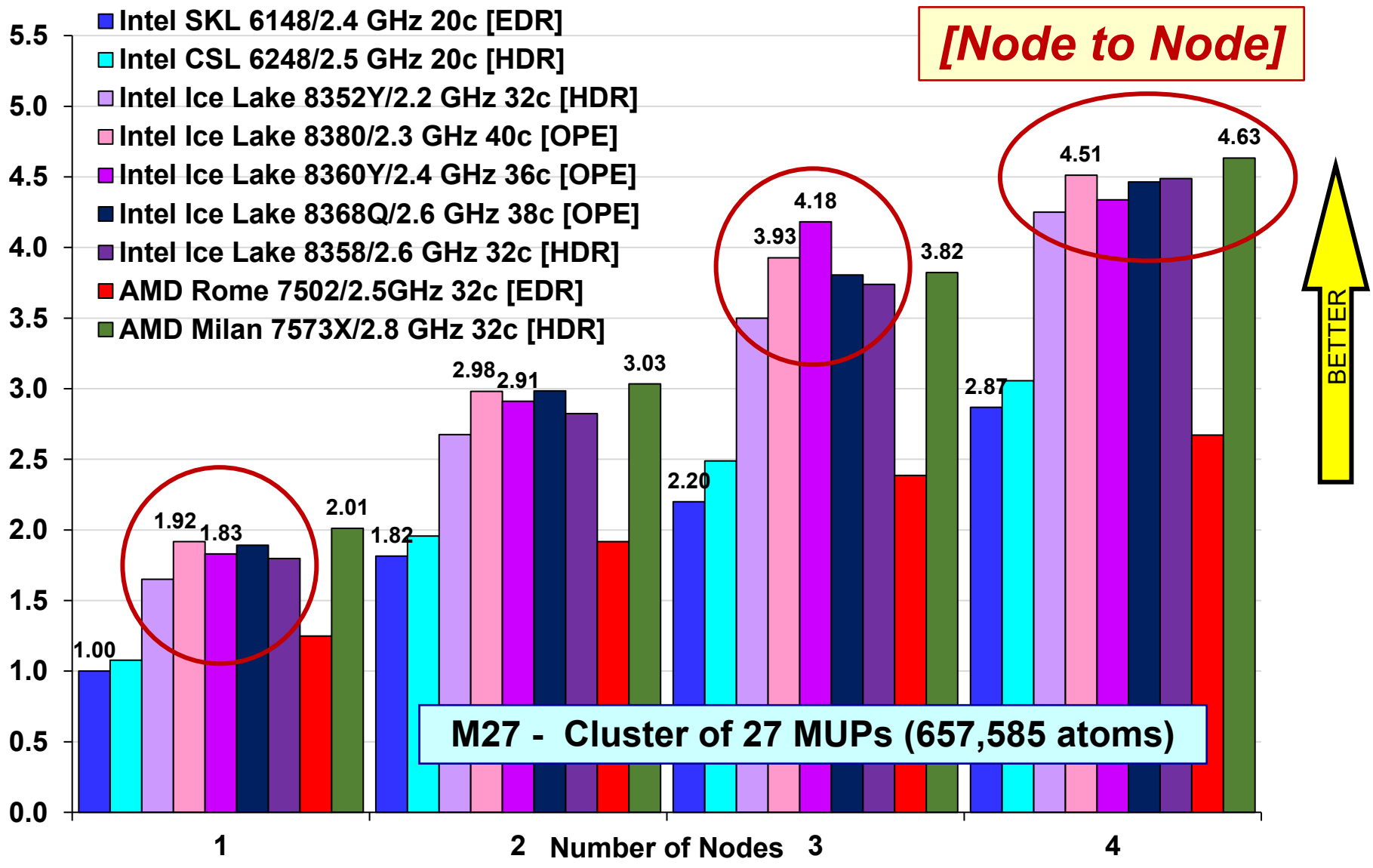
Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*



AMBER18 - M27 Performance Analysis

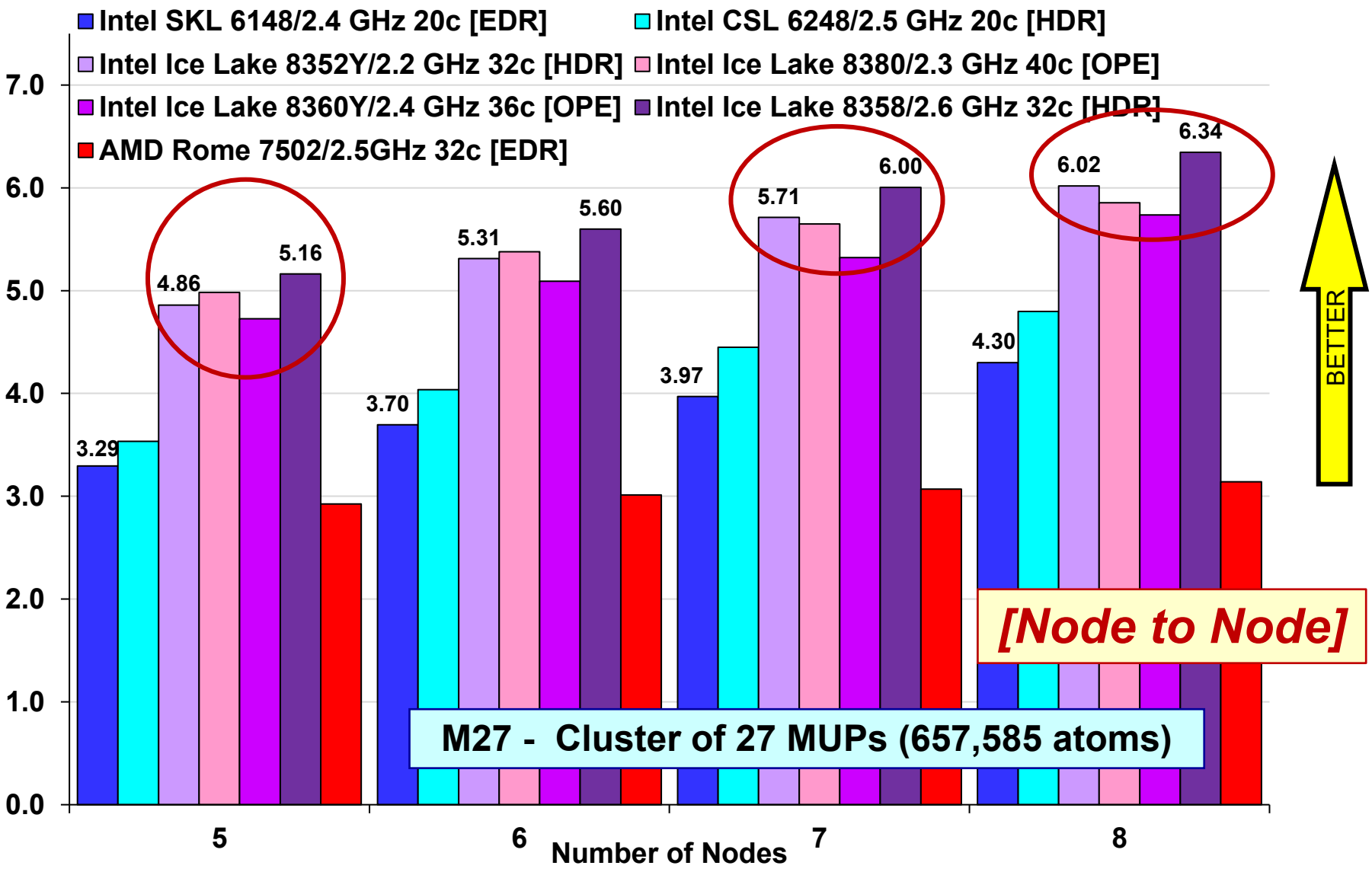
CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

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

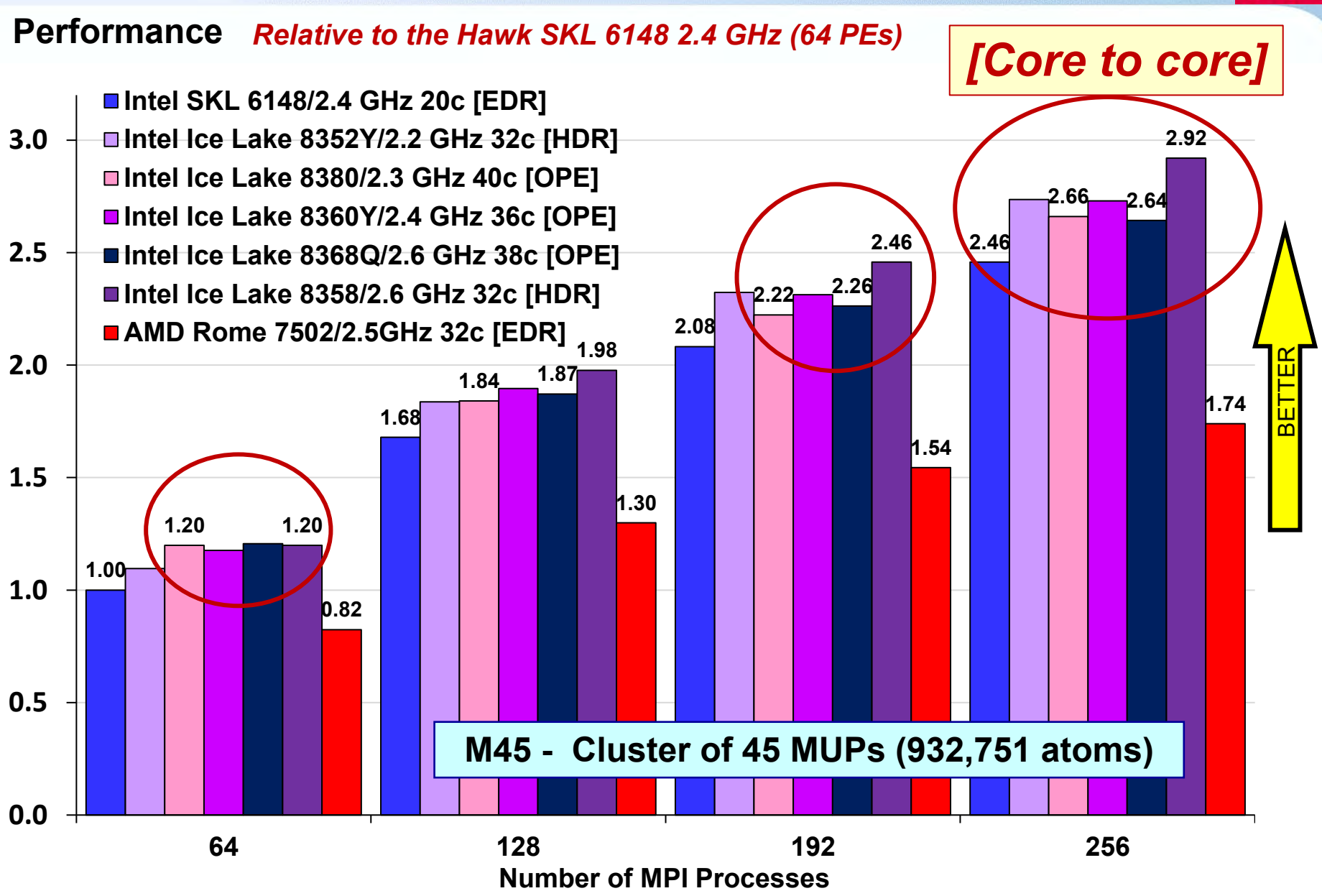


AMBER18 - M27 Performance Analysis

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

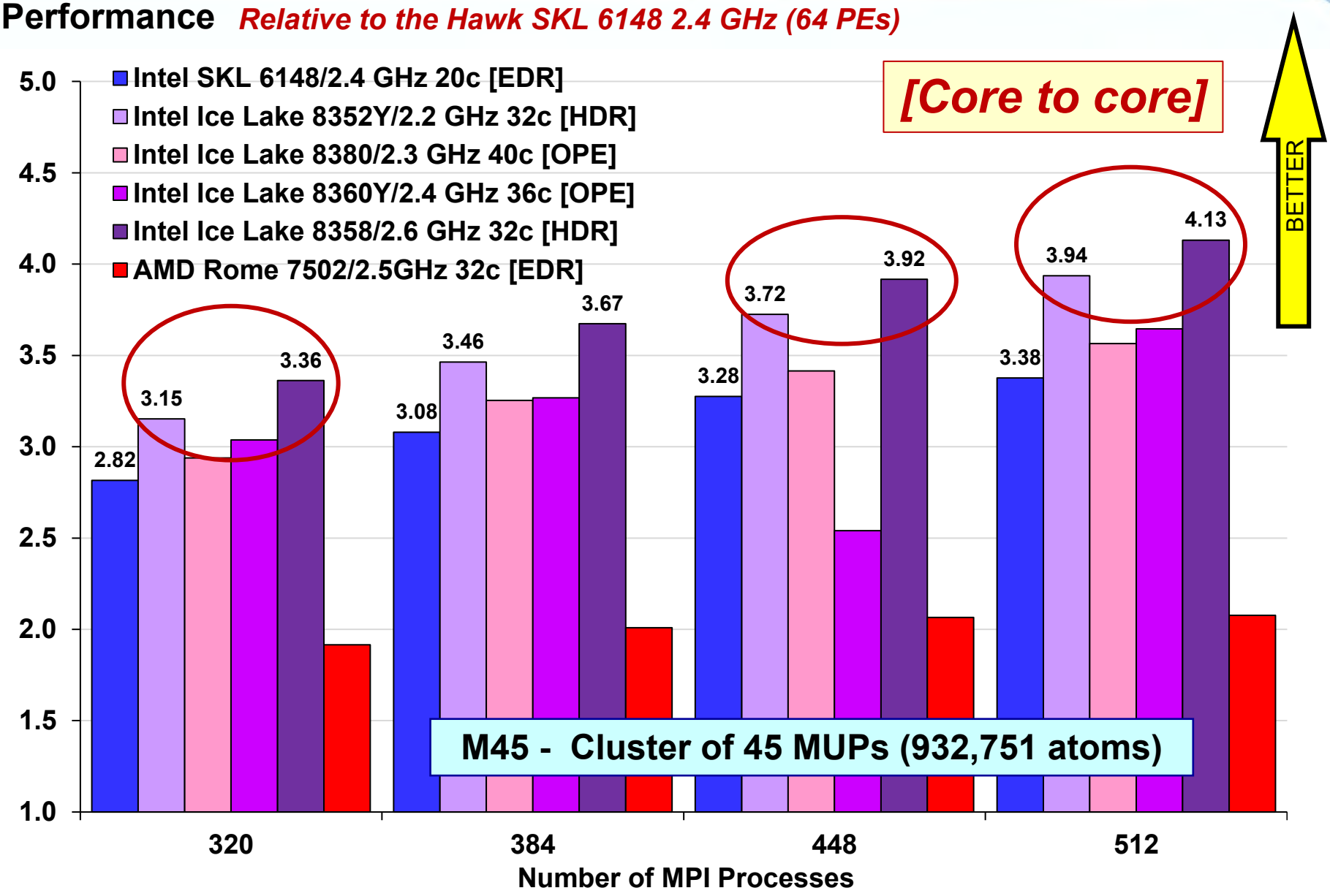


AMBER18 - M45 Performance Analysis

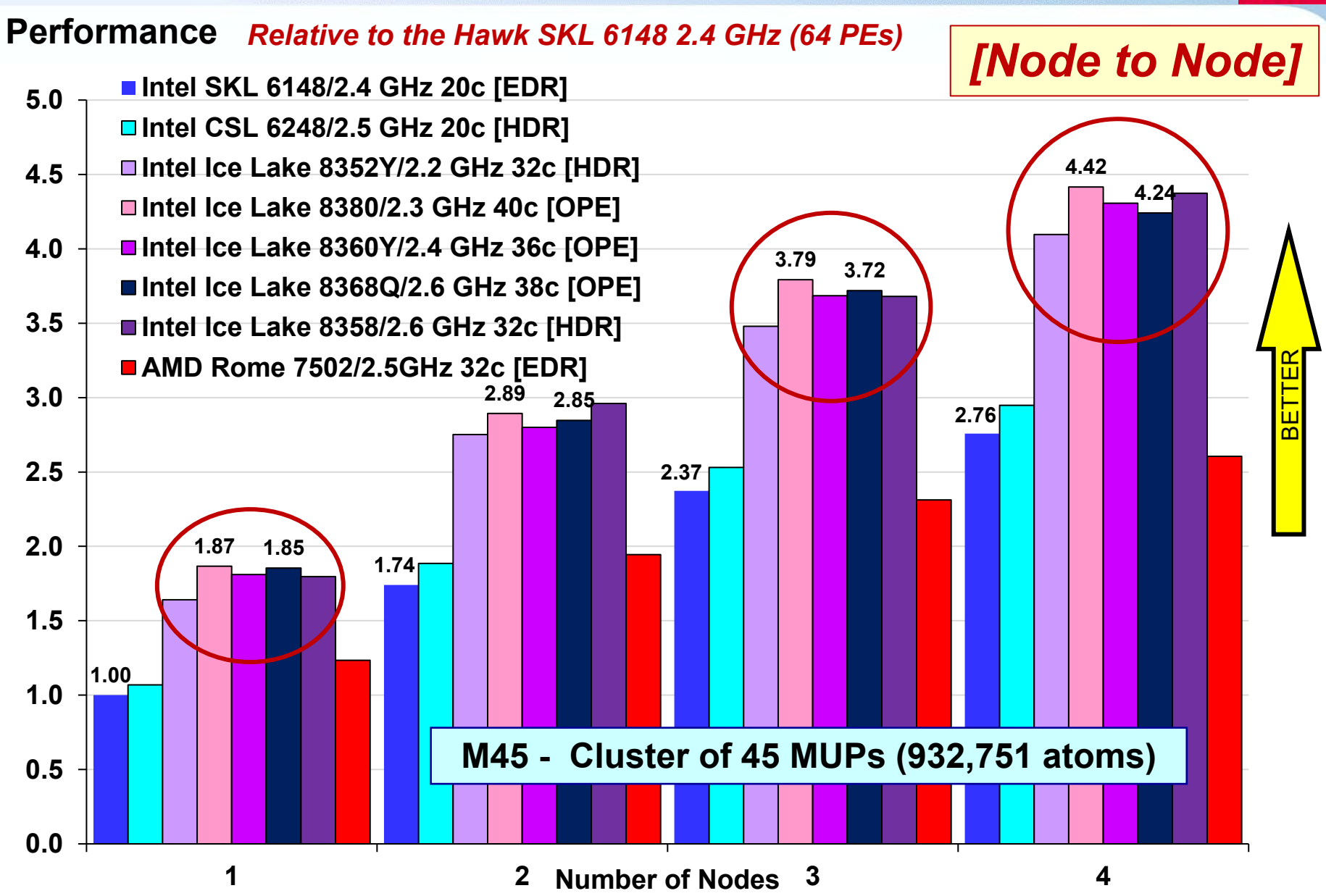


AMBER18 - M45 Performance Analysis

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



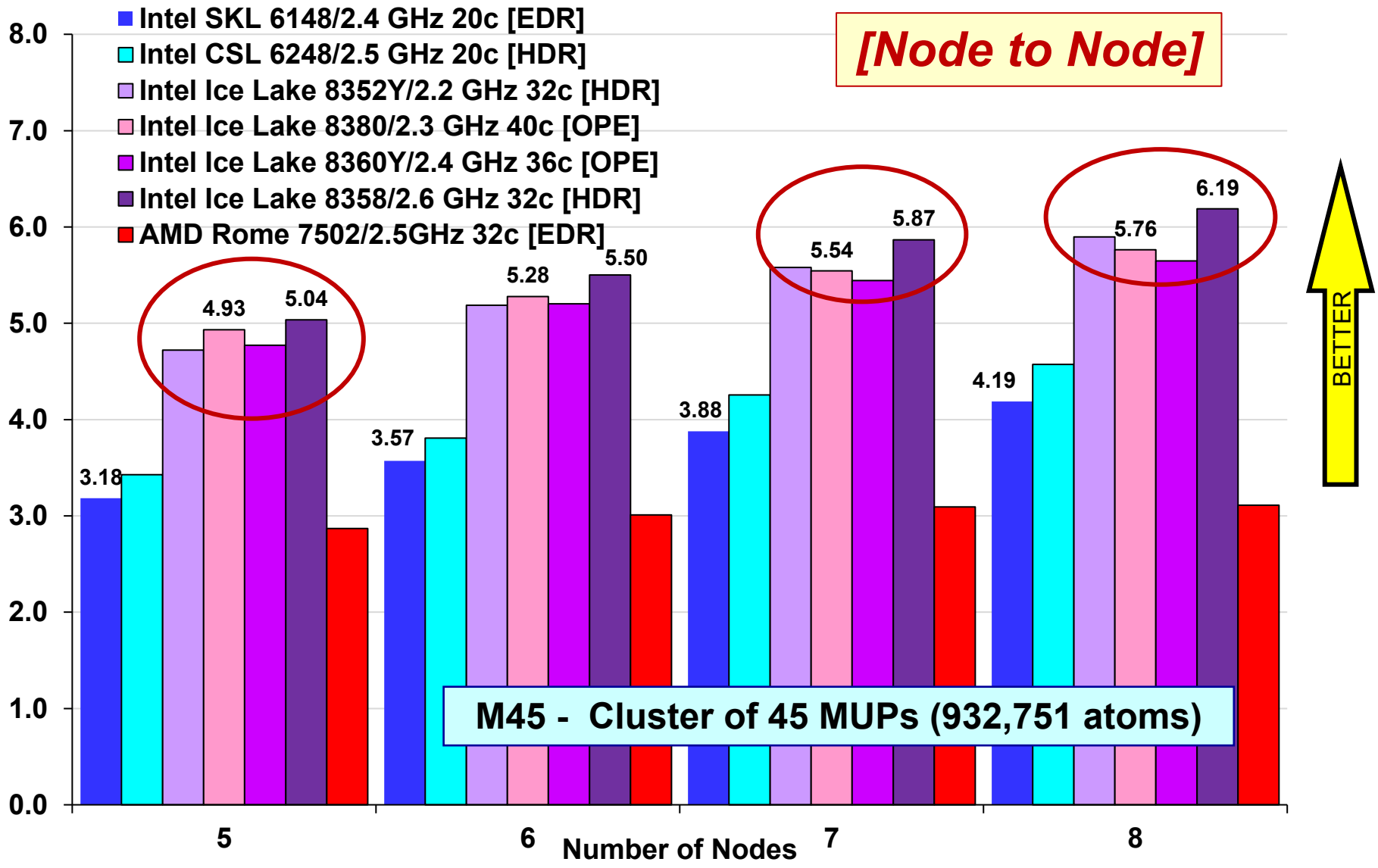
AMBER18 - M45 Performance Analysis



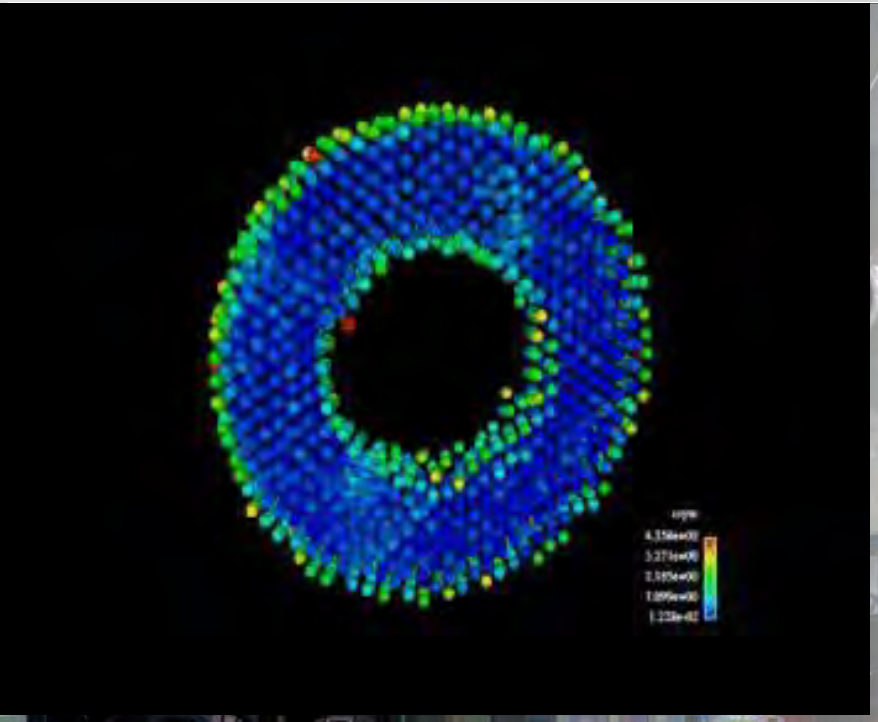
AMBER18 - M45 Performance Analysis

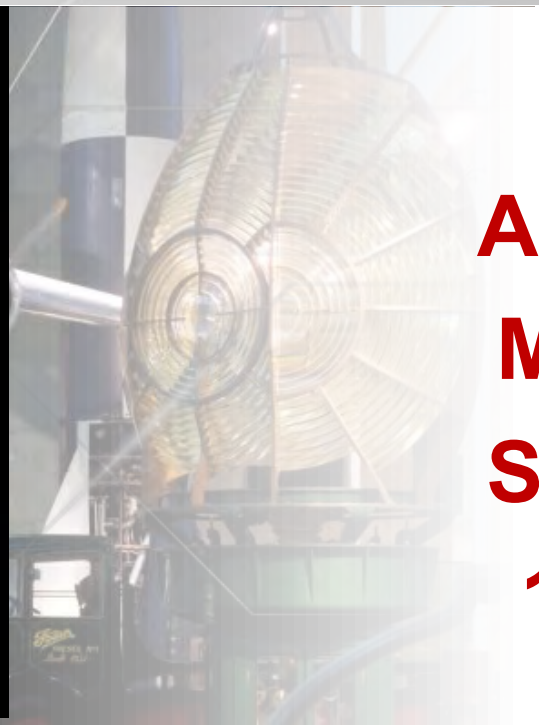
CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

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



Performance of Computational Chemistry and Ocean Modelling Codes





Advanced Materials Software:

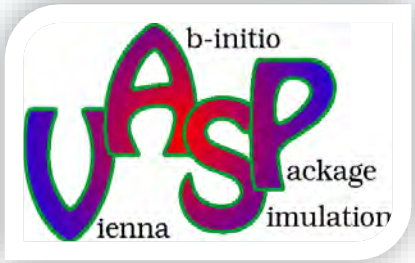
1. VASP

Advanced Materials Software

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**
- **CASTEP** – a full-featured materials modelling code based on a first-principles QM description of electrons and nuclei. Uses robust methods of a **plane-wave basis set and pseudopotentials**.
- **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



VASP (**6.3**) 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

Archer Rank: 1

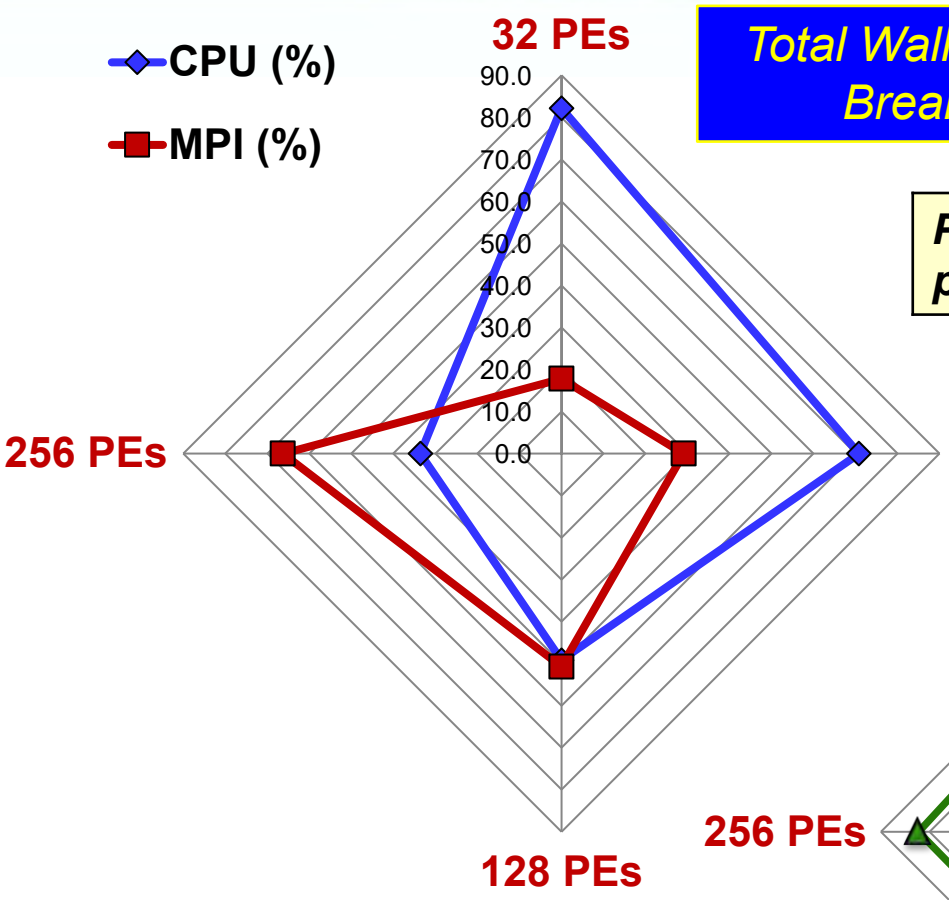
Pd-O Benchmark

- 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

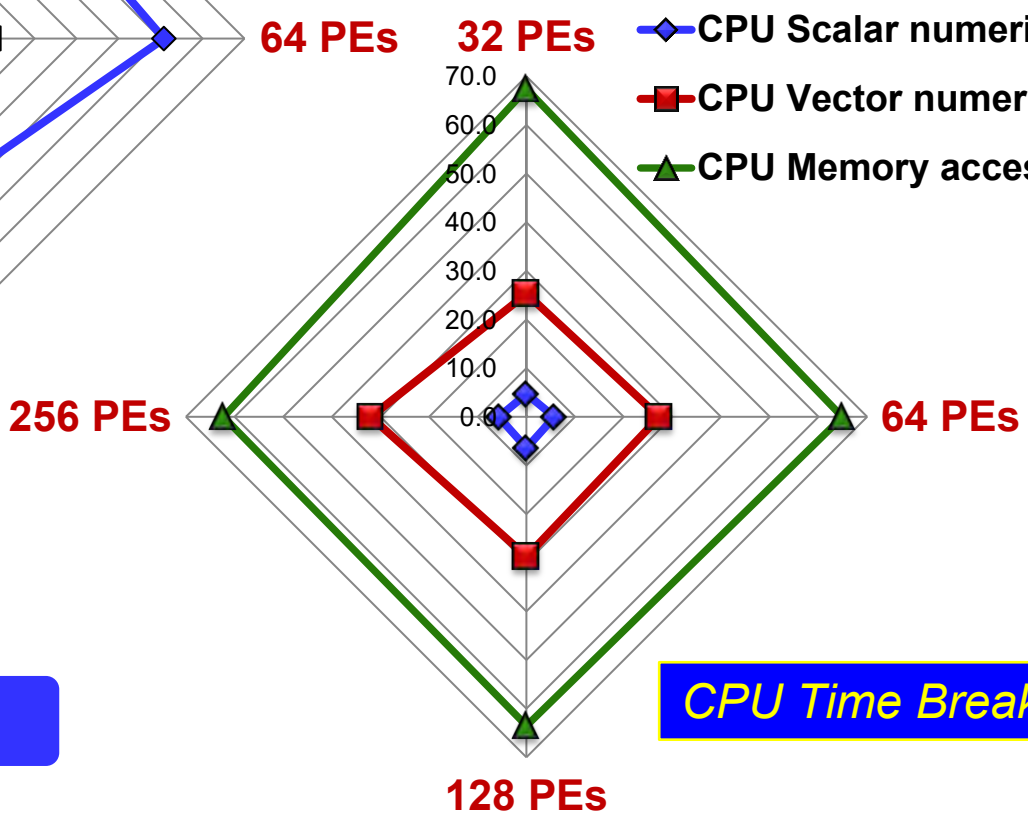
- Zeolite with the MFI structure unit cell running a single point calculation and a planewave cut off of 400eV using the PBE functional
- 2 k-points; maximum number of plane-waves: 96,834
- FFT grid; NGX=65, NGY=65, NGZ=43, giving a total of 181,675 points

VASP – Pd-O Benchmark Performance Report



Total Wallclock Time Breakdown

Palladium-Oxygen complex ($Pd_{75}O_{12}$), 10 k-points, FFT grid: (31, 49, 45), 68,355 points



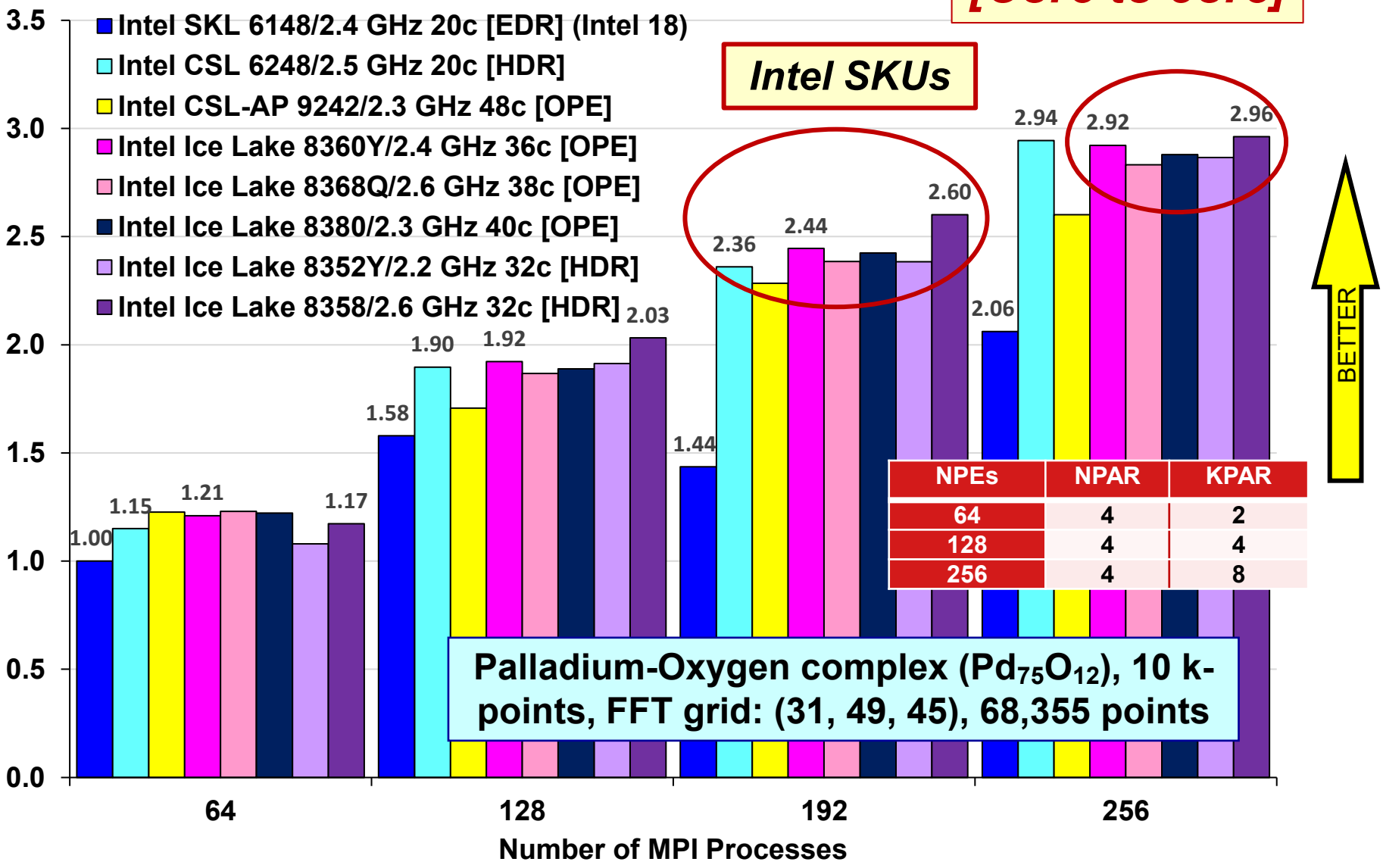
CPU Time Breakdown

Performance Data (32-256 PEs)

VASP 6.3 – Pd-O Benchmark - Parallelisation on k-points

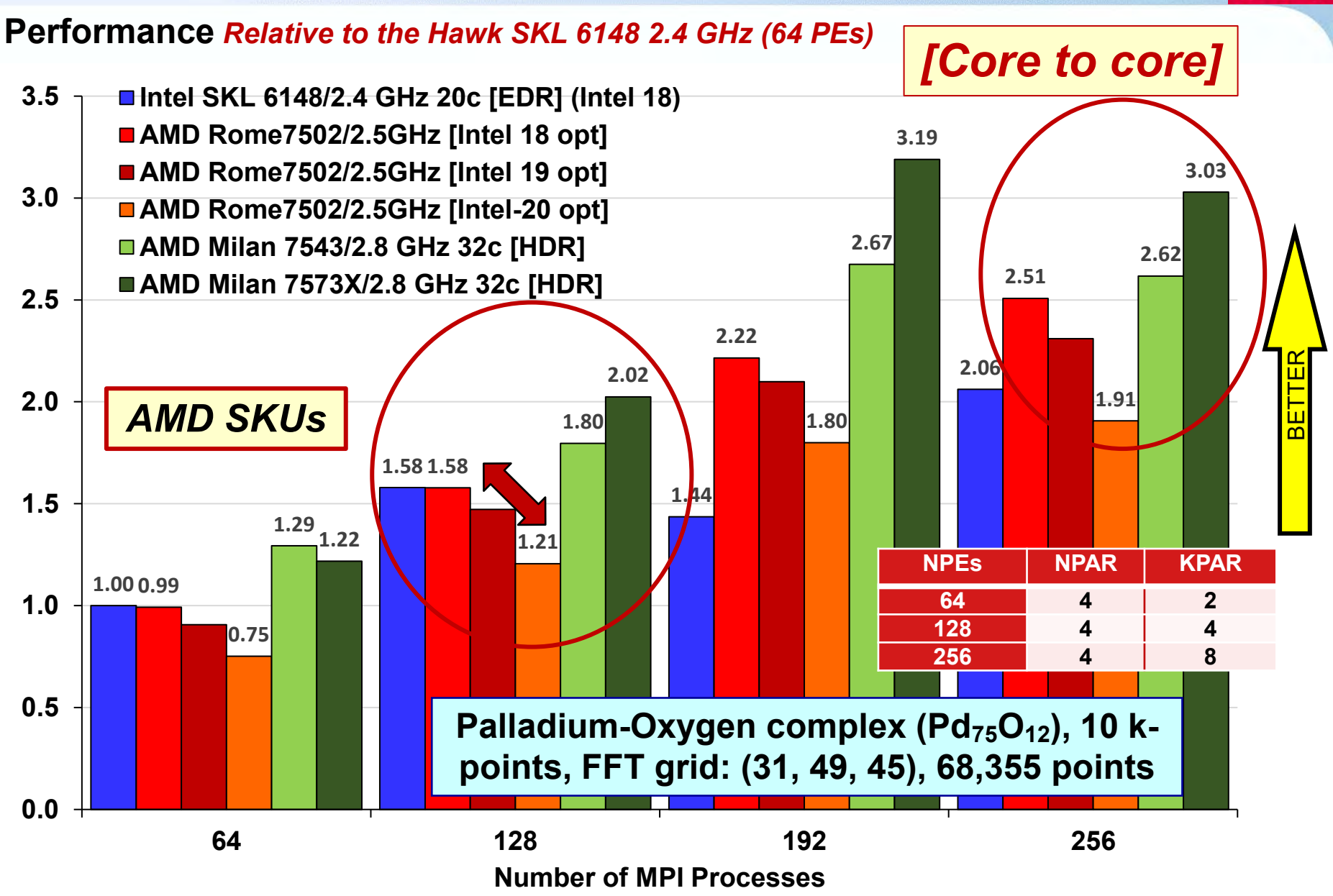
Performance ***Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)***

[Core to core]



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

VASP 6.3 – Pd-O Benchmark - Parallelisation on k-points

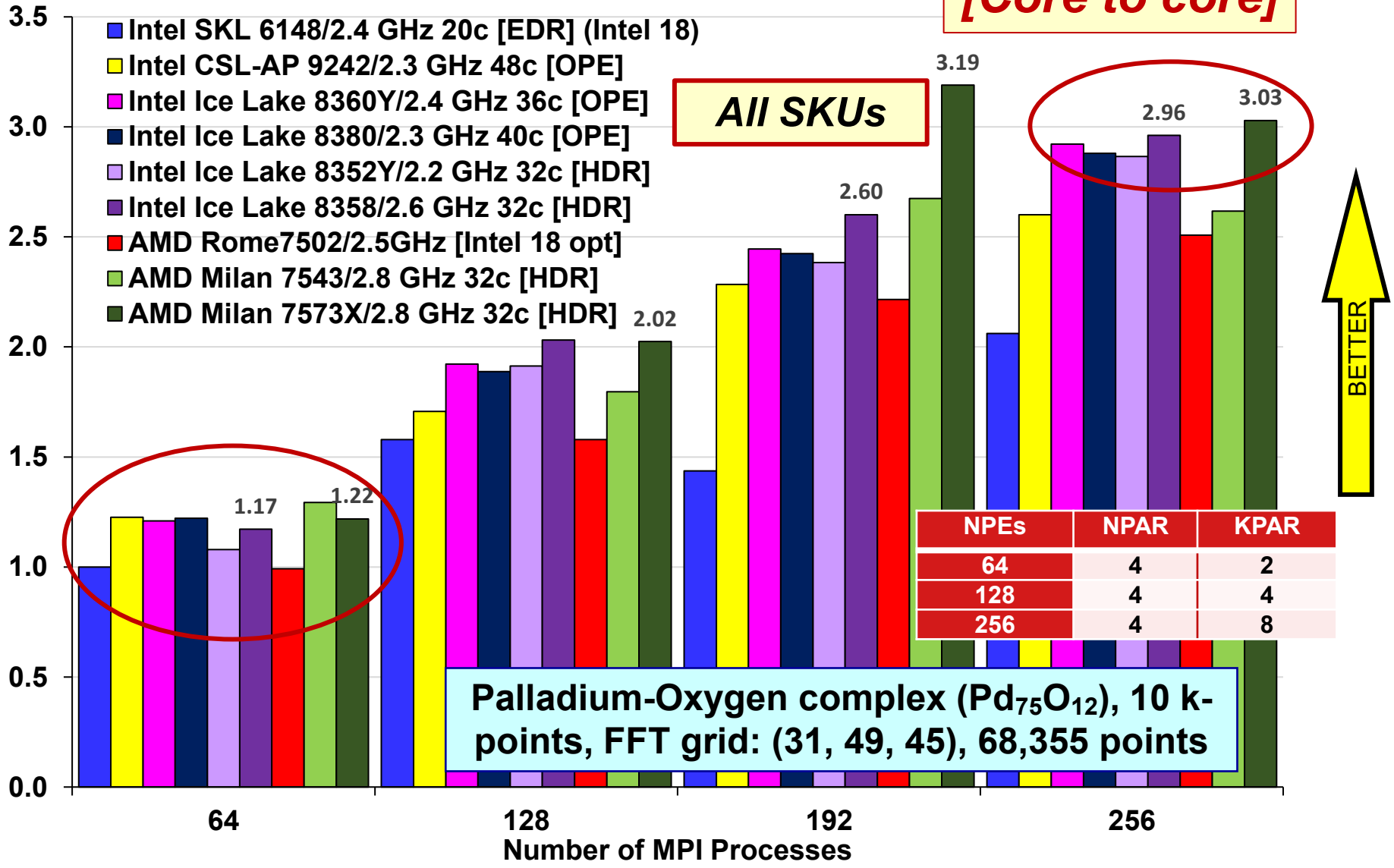


VASP 6.3 – Pd-O Benchmark - Parallelisation on k-points

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

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

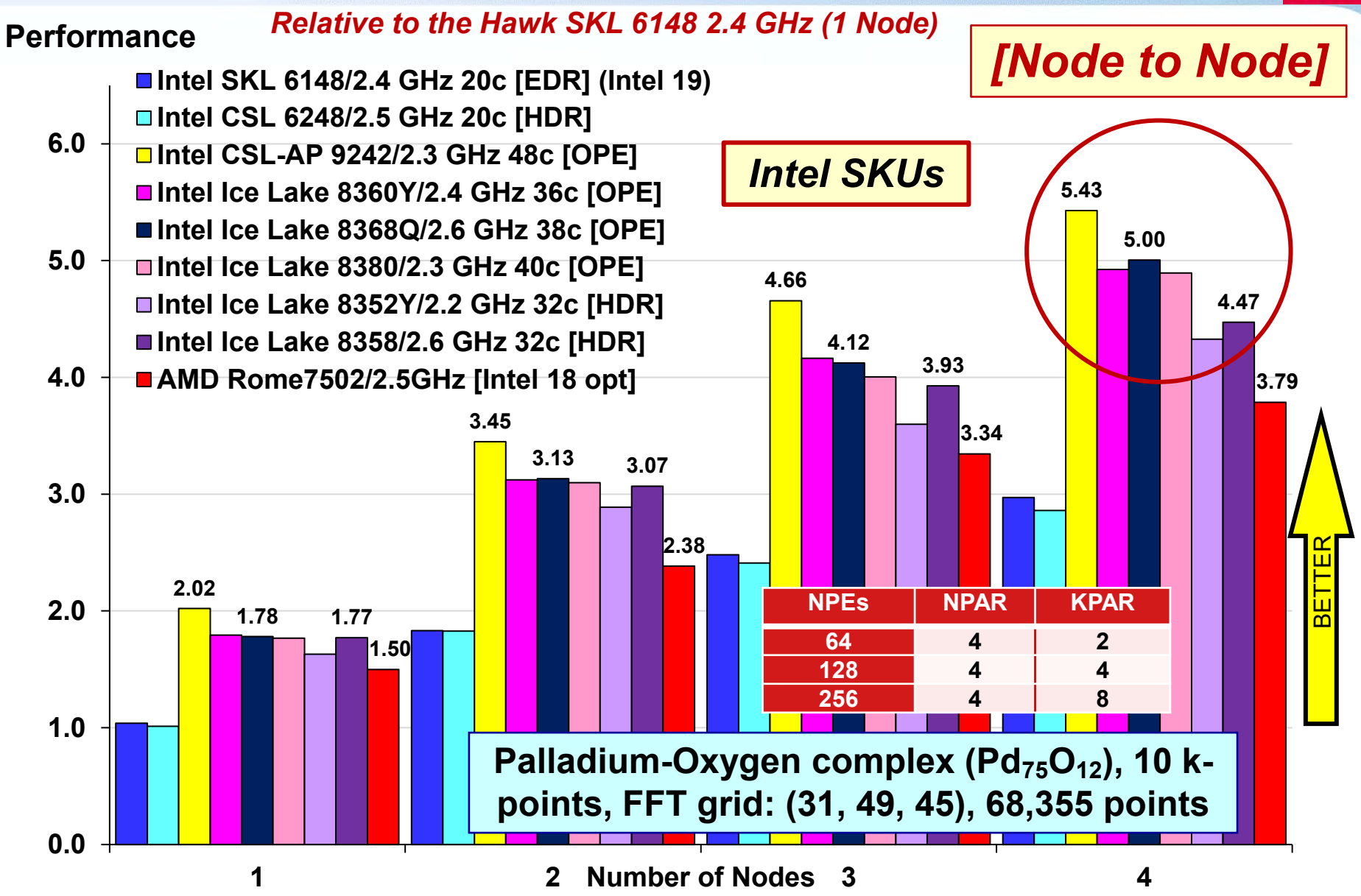
[Core to core]



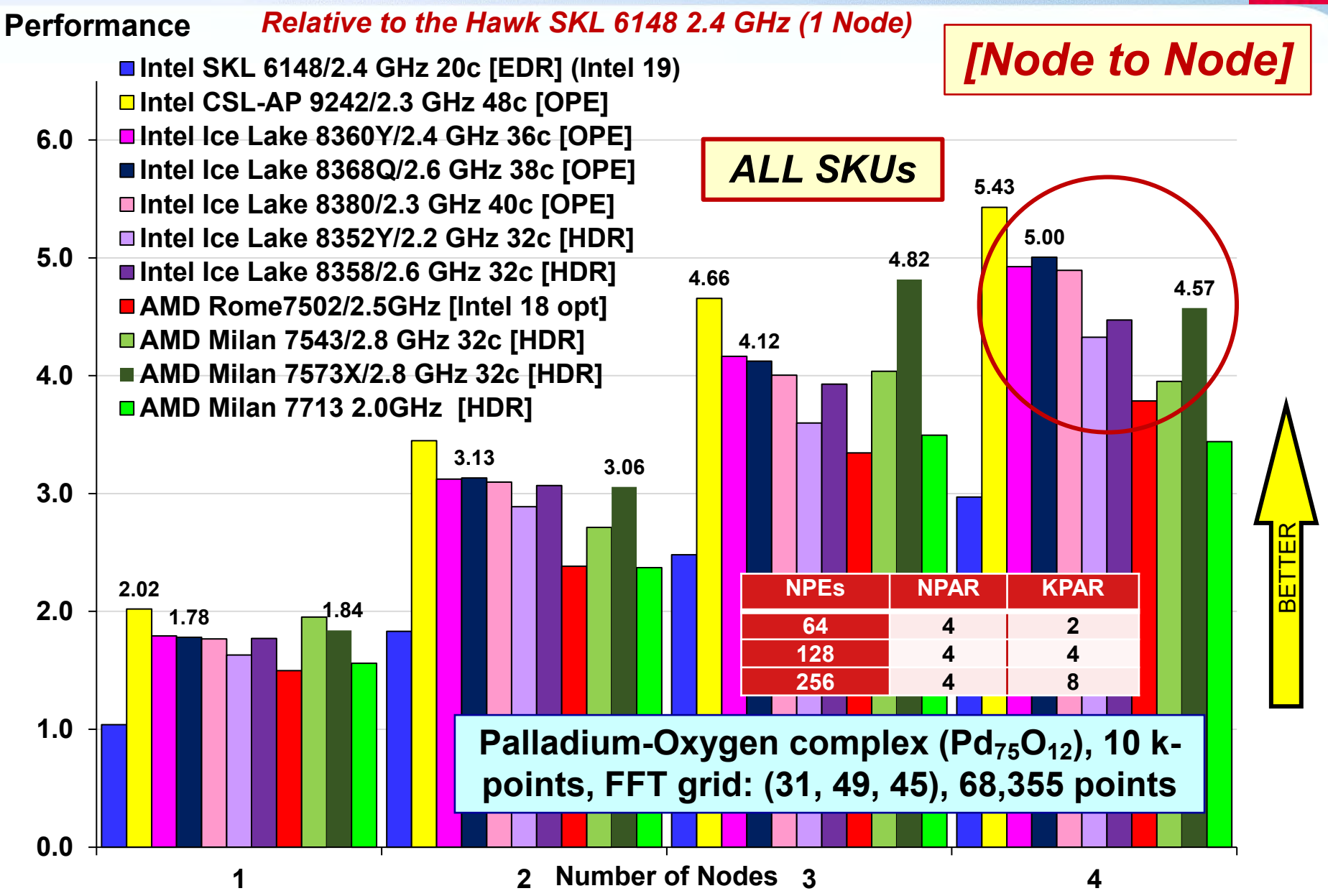
NPEs	NPAR	KPAR
64	4	2
128	4	4
256	4	8

Palladium-Oxygen complex ($Pd_{75}O_{12}$), 10 k-points, FFT grid: (31, 49, 45), 68,355 points

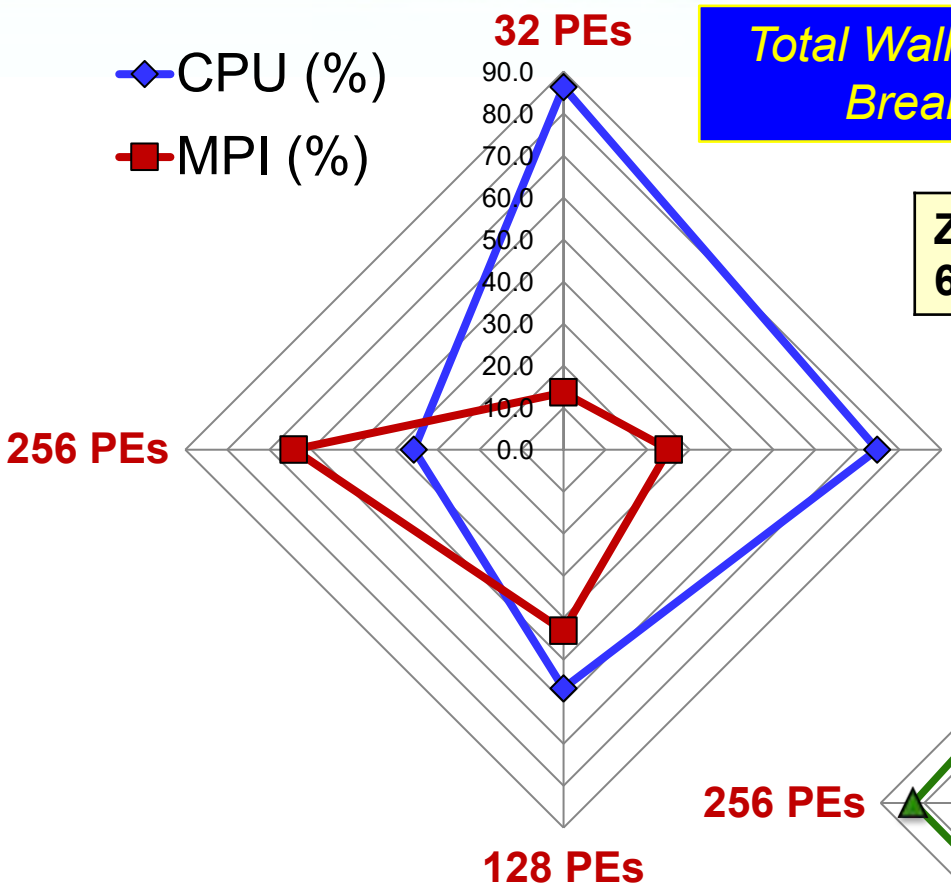
VASP 6.3 – Pd-O Benchmark - Parallelisation on k-points



VASP 6.3 – Pd-O Benchmark - Parallelisation on k-points

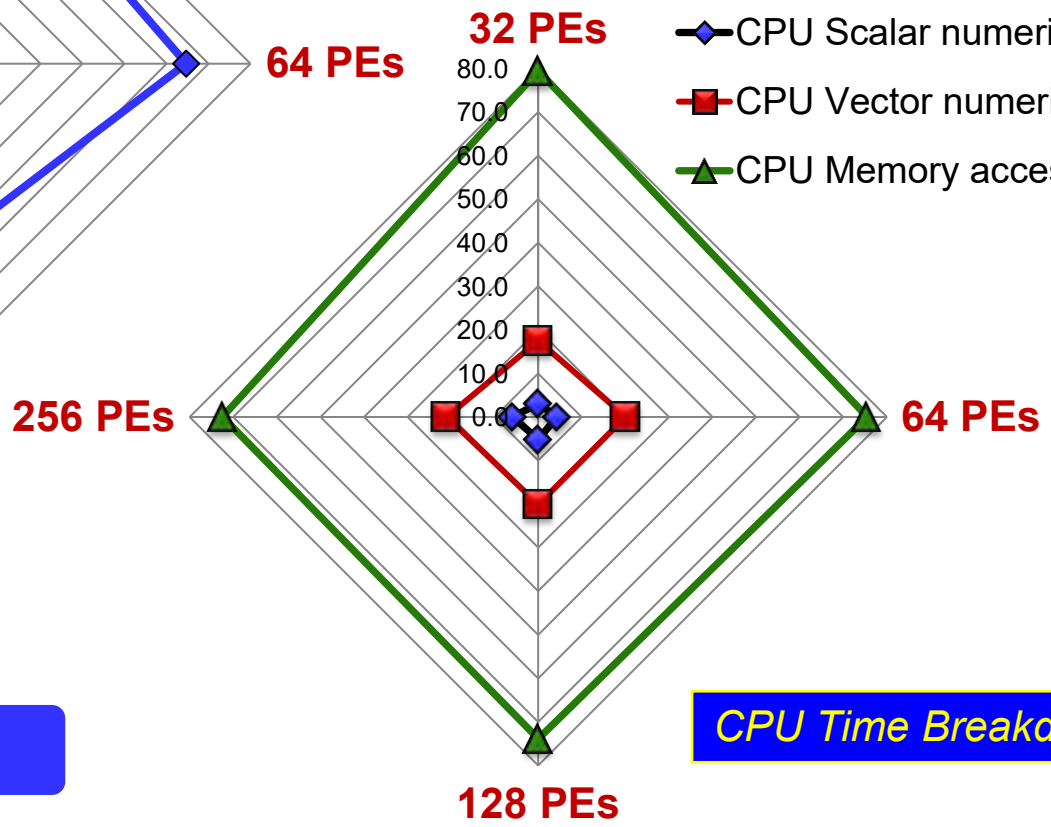


VASP – Zeolite Cluster Performance Report



Total Wallclock Time Breakdown

Zeolite ($Si_{96}O_{192}$), 2 k-points, FFT grid: (65, 65, 43); 181,675 points

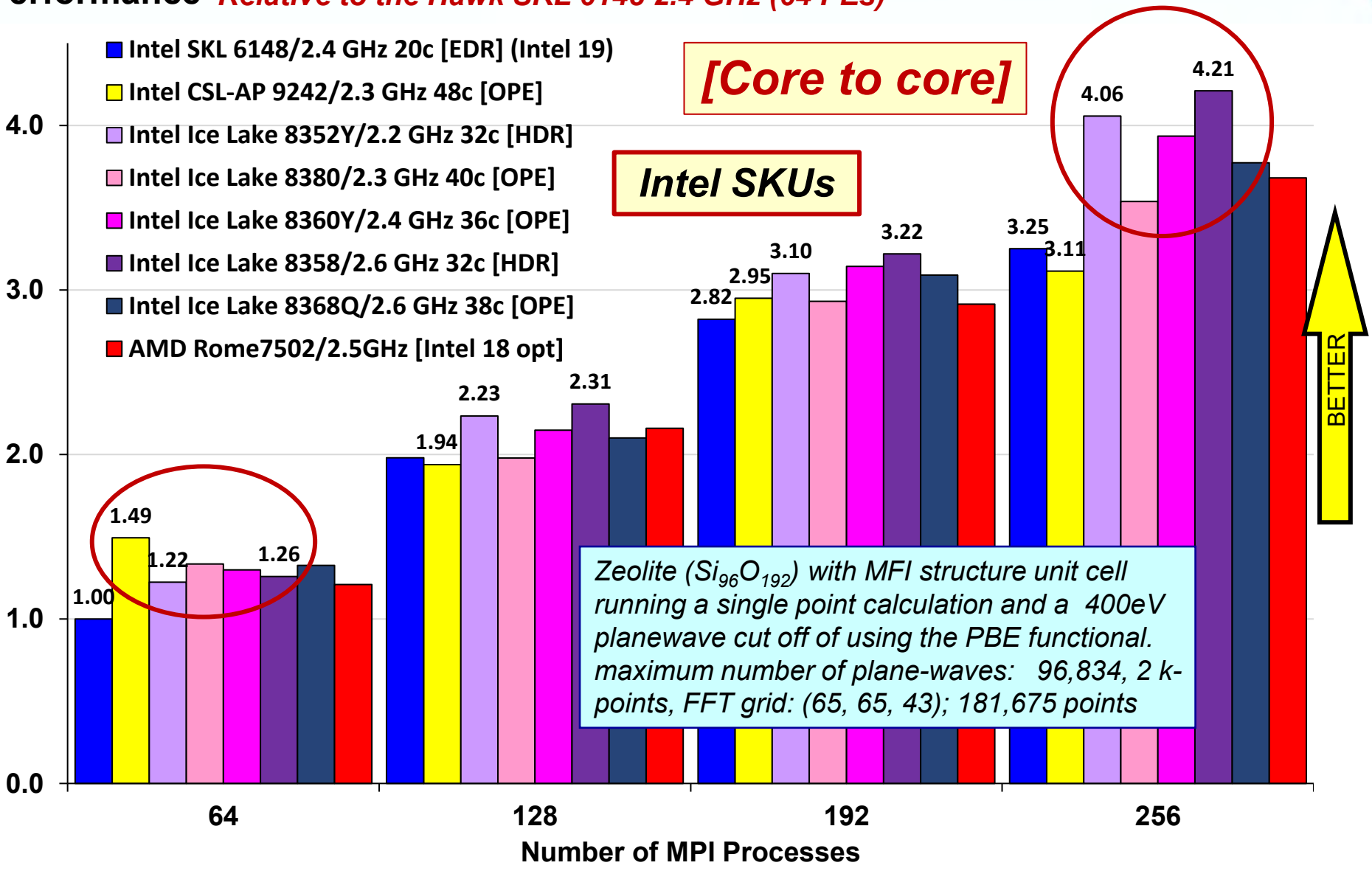


CPU Time Breakdown

Performance Data (32-256 PEs)

VASP 6.3 – Zeolite Benchmark - Parallelisation on k-points

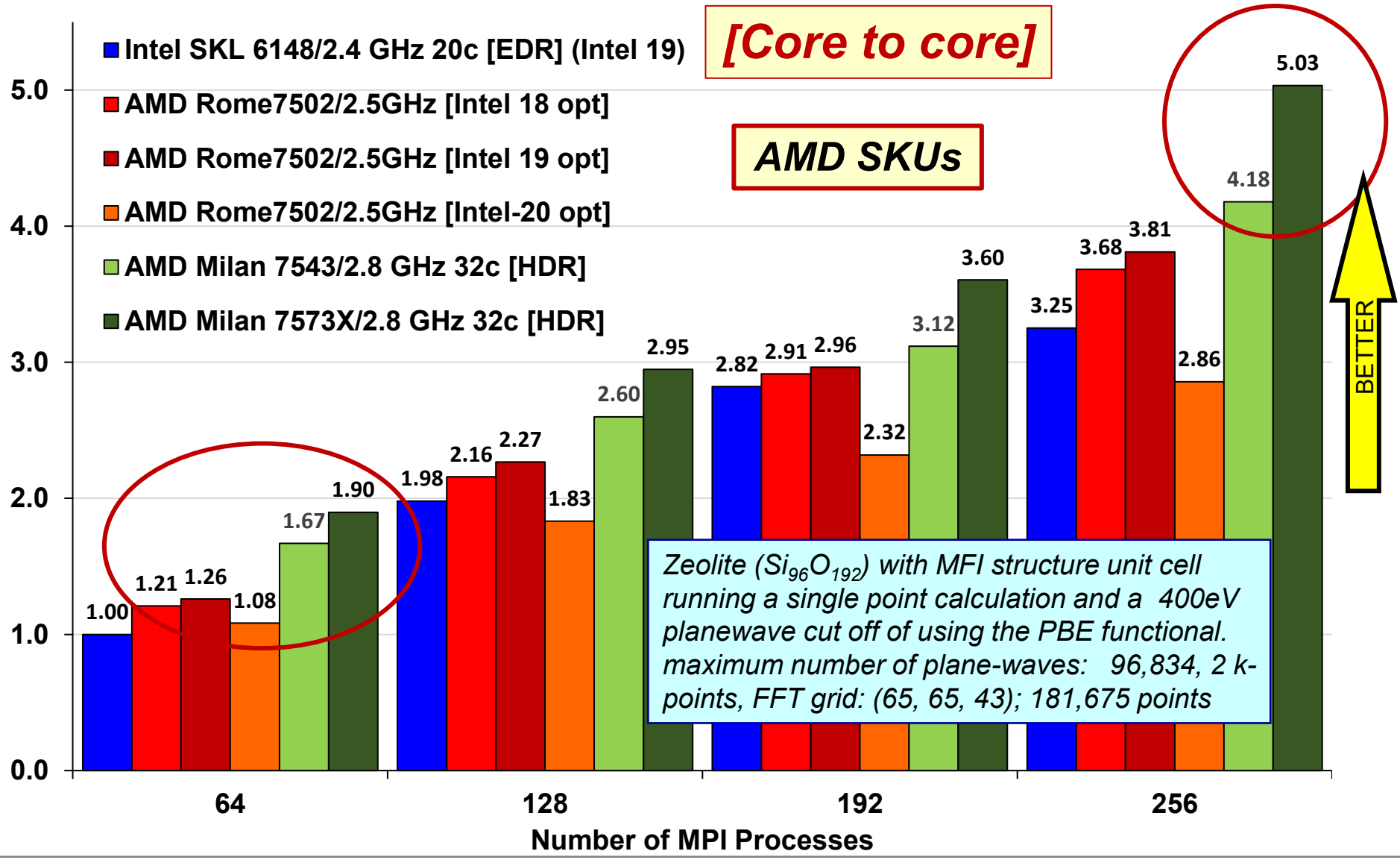
Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*



VASP 6.3 – Zeolite Benchmark - Parallelisation on k-points

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

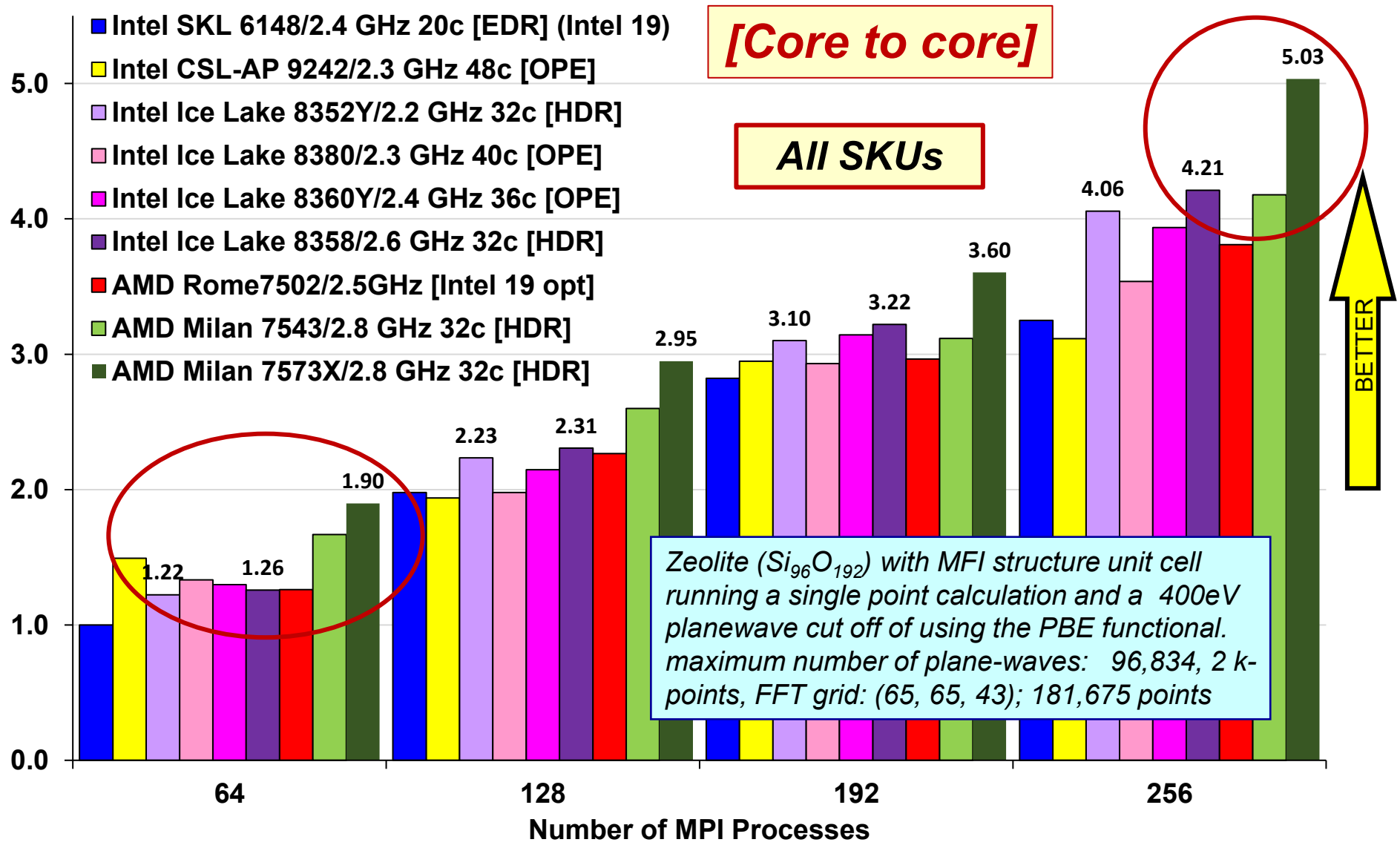
Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*



VASP 6.3 – Zeolite Benchmark - Parallelisation on k-points

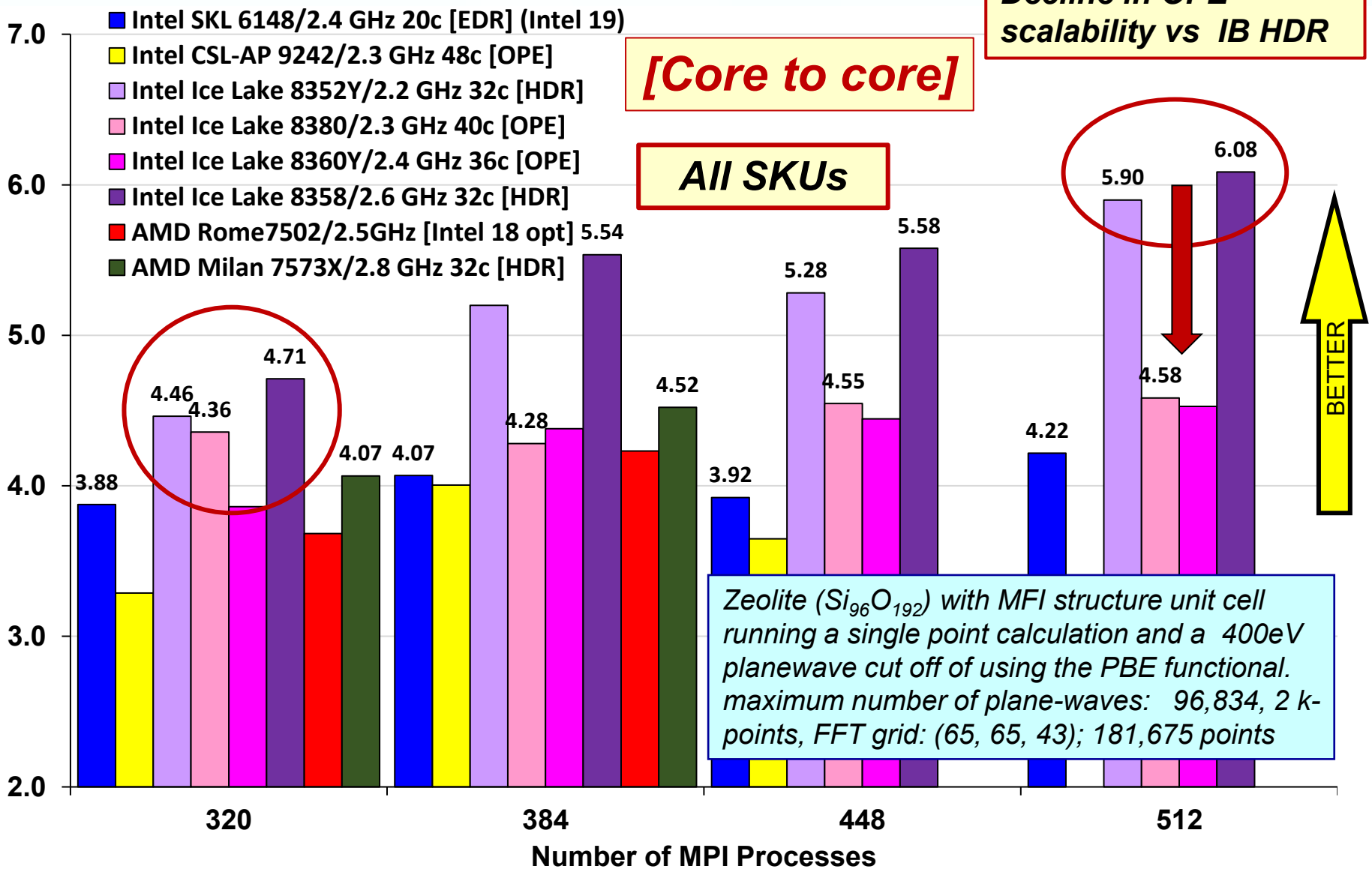
CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

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



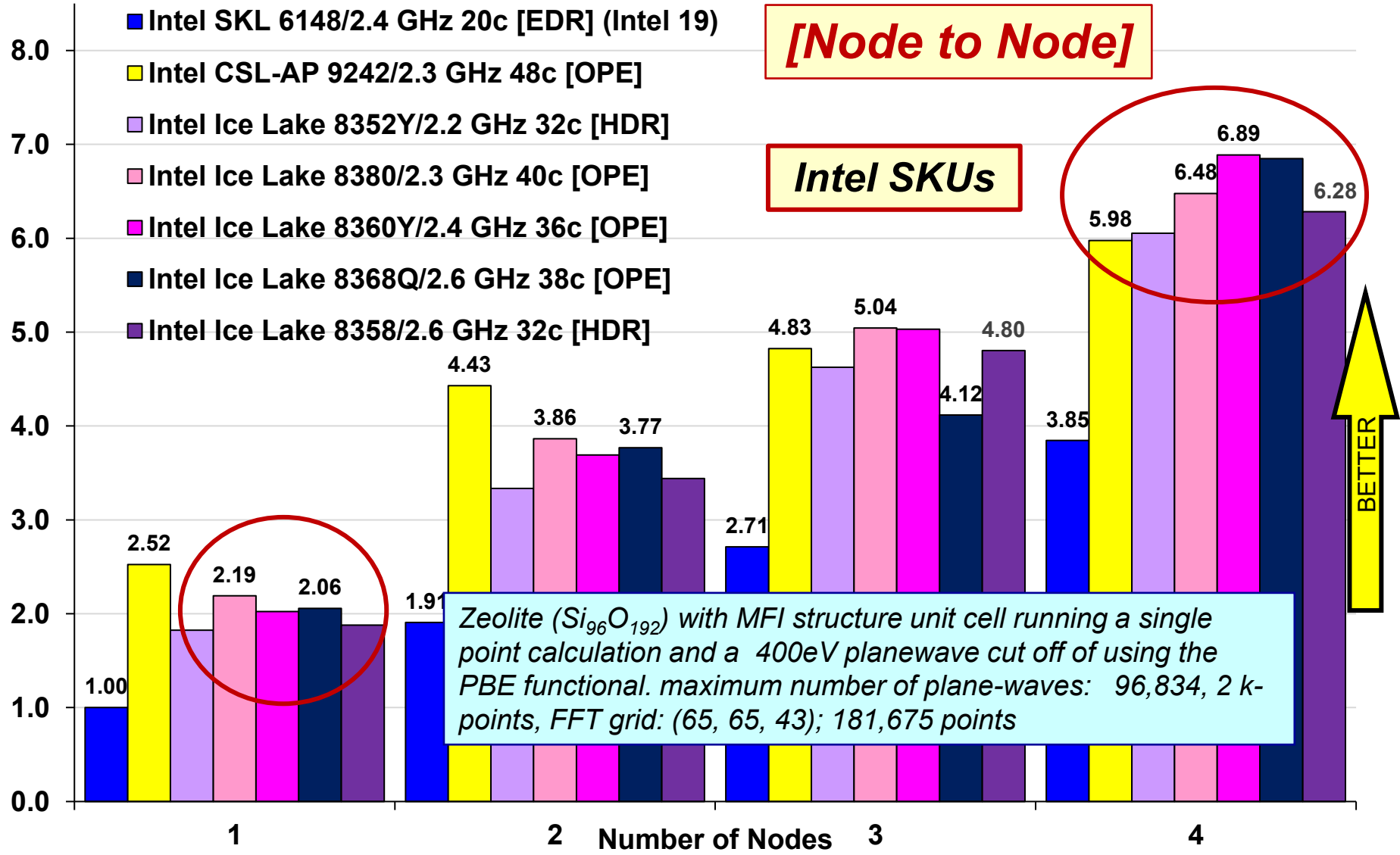
VASP 6.3 – Zeolite Benchmark - Parallelisation on k-points

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



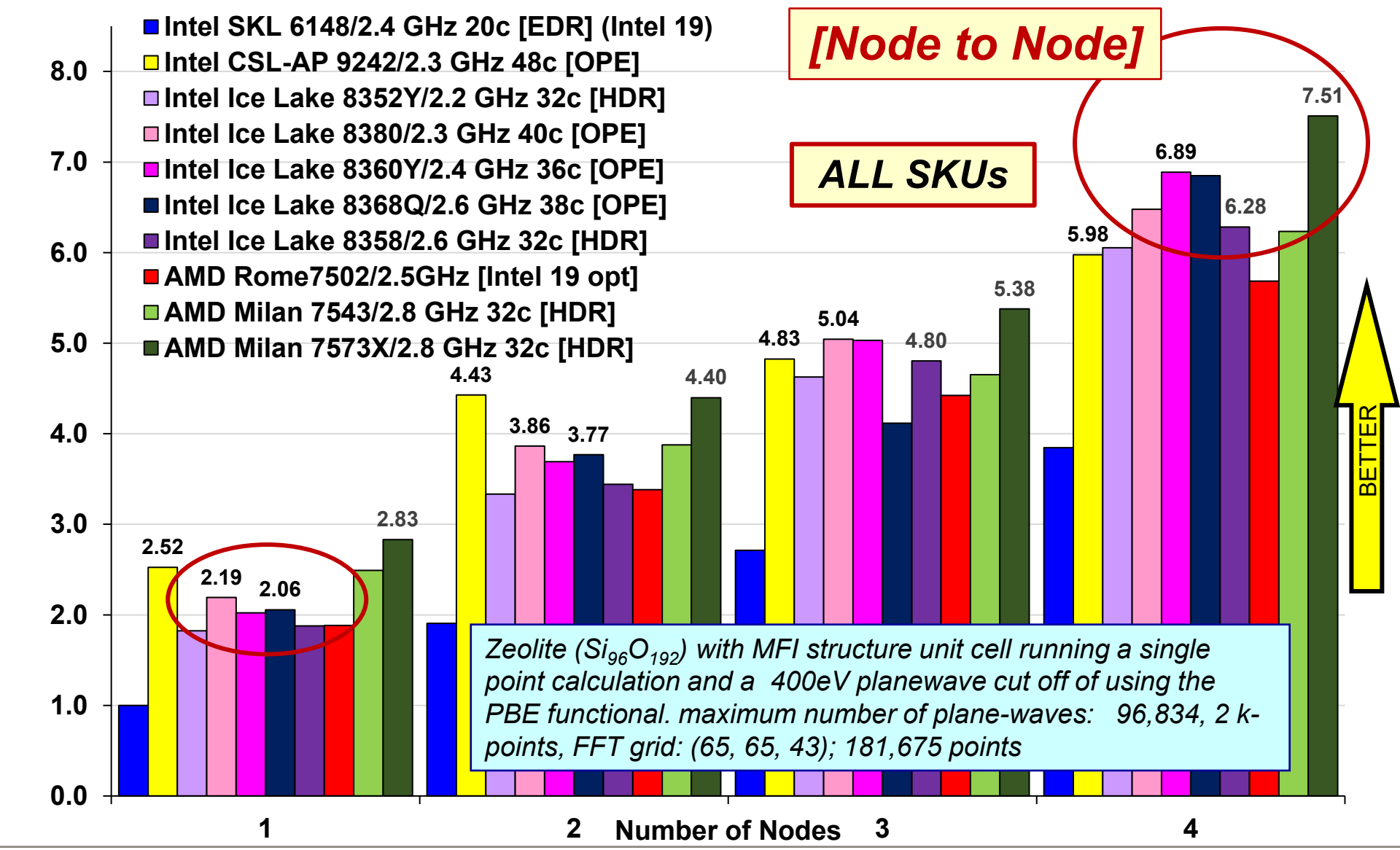
VASP 6.3 – Zeolite Benchmark - Parallelisation on k-points

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

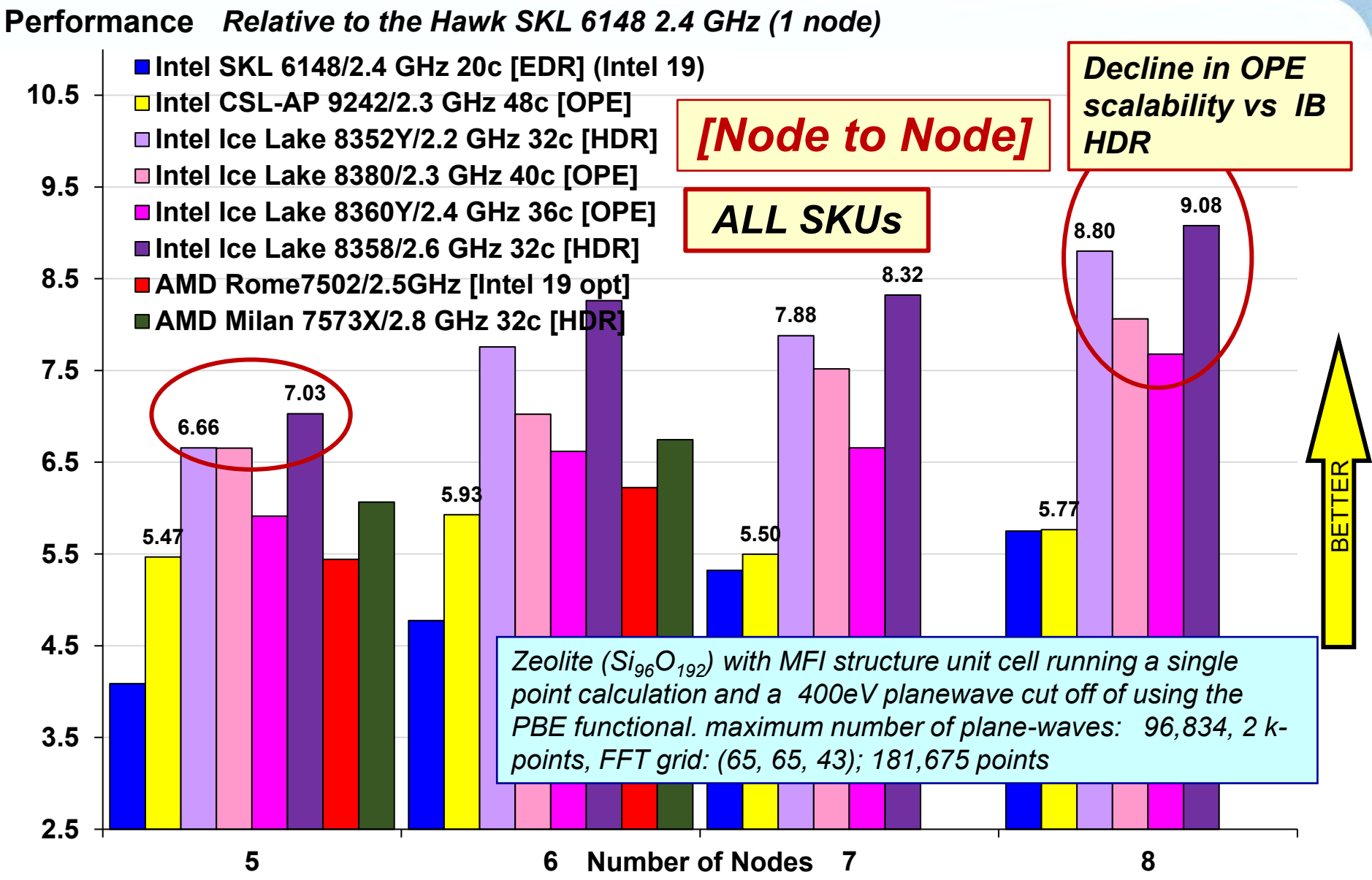


VASP 6.3 – Zeolite Benchmark - Parallelisation on k-points

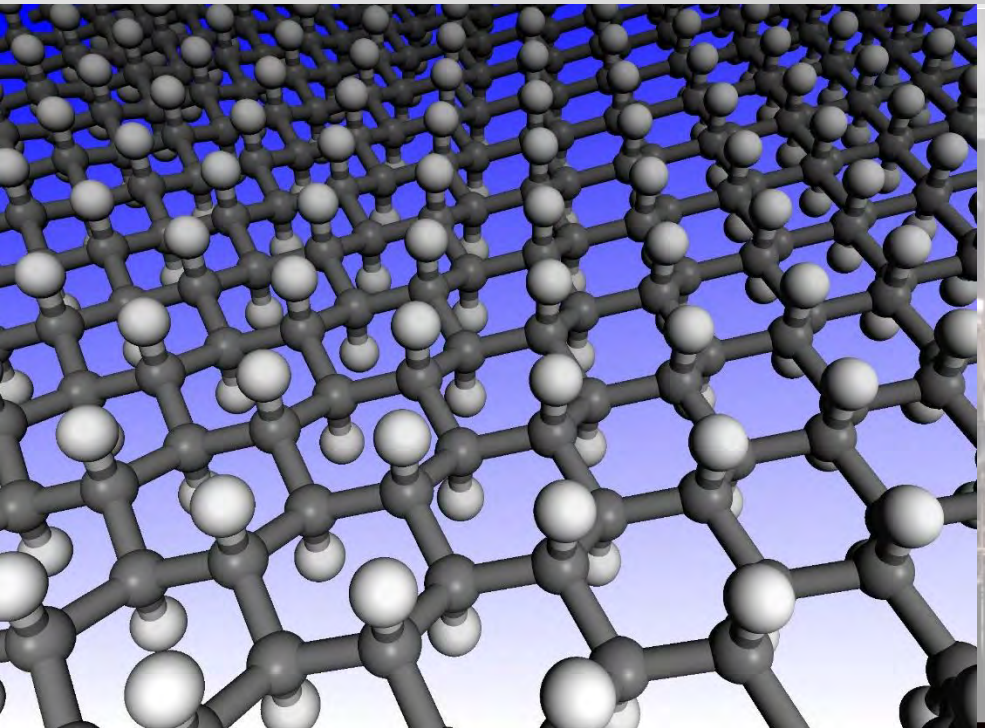
Performance *Relative to the Hawk SKL 6148 2.4 GHz (1 node)*



VASP 6.3 – Zeolite Benchmark - Parallelisation on k-points



Performance of Computational Chemistry and Ocean Modelling Codes



Advanced Materials Software:

2. CASTEP

CASTEP – Materials Modelling

- **CASTEP** is a full-featured materials modelling code based on a first-principles quantum mechanical description of electrons and nuclei. It uses the robust methods of a plane-wave basis set and pseudopotentials.
- Two versions of CASTEP used in this study, **Version 19.1.1** and the current academic release of CASTEP, **Version 21.1.1**.

- **Al3x3 Benchmark**

  The al3x3 simulation cell comprises a 270-atom sapphire surface, with a vacuum gap. There are only 2 k-points, so it is a good test of the performance of CASTEP's other parallelisation strategies.

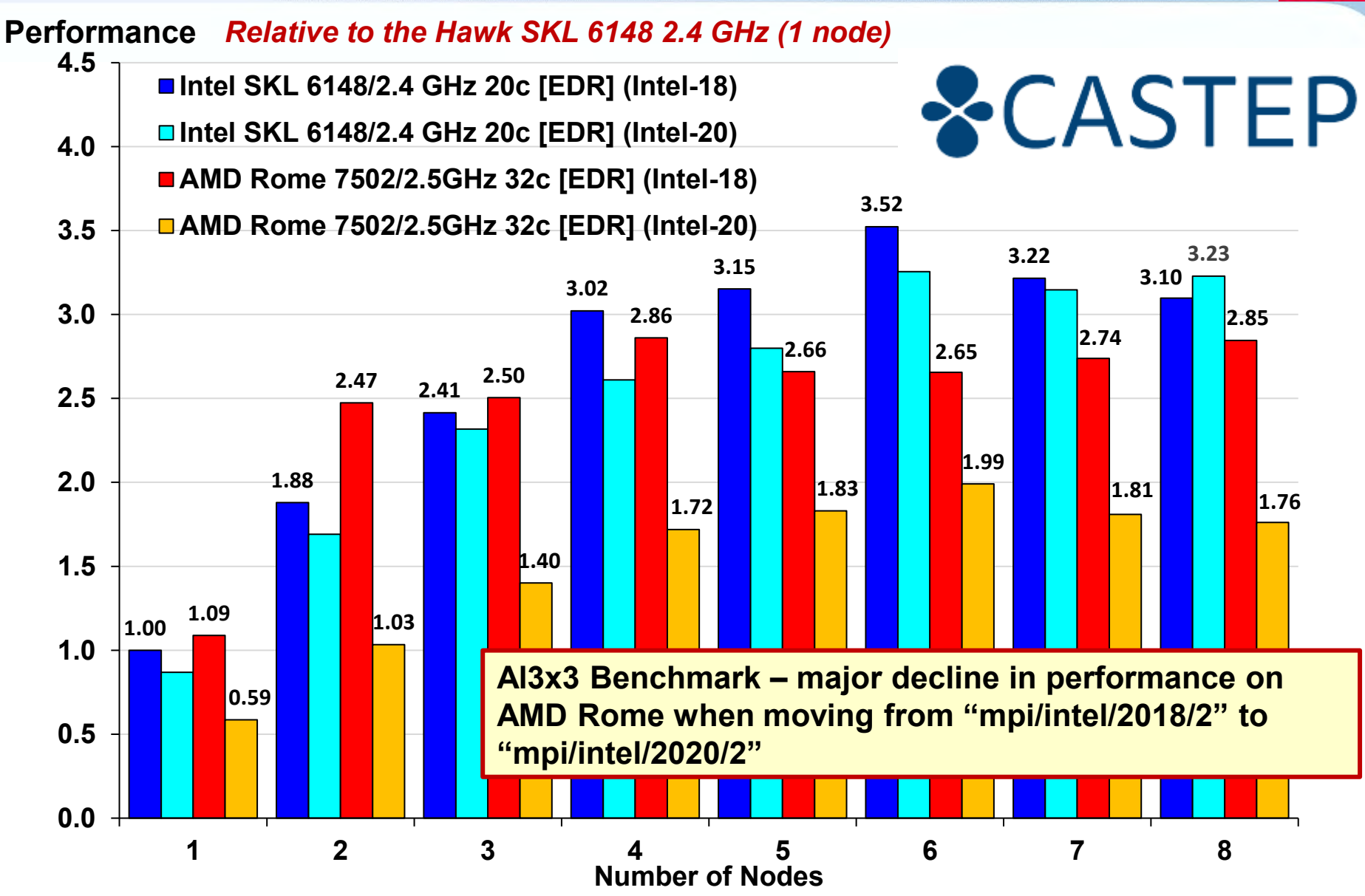
- **MnO_2 Benchmark**

  Bigger calculation (313 electrons and 64 ions) and involves MPI AllToAllV across all processors.

- **IDZ Benchmark**

  Longer MD calculation (1104 electrons and 404 ions) requiring several random initializations (16 MD iterations in total).

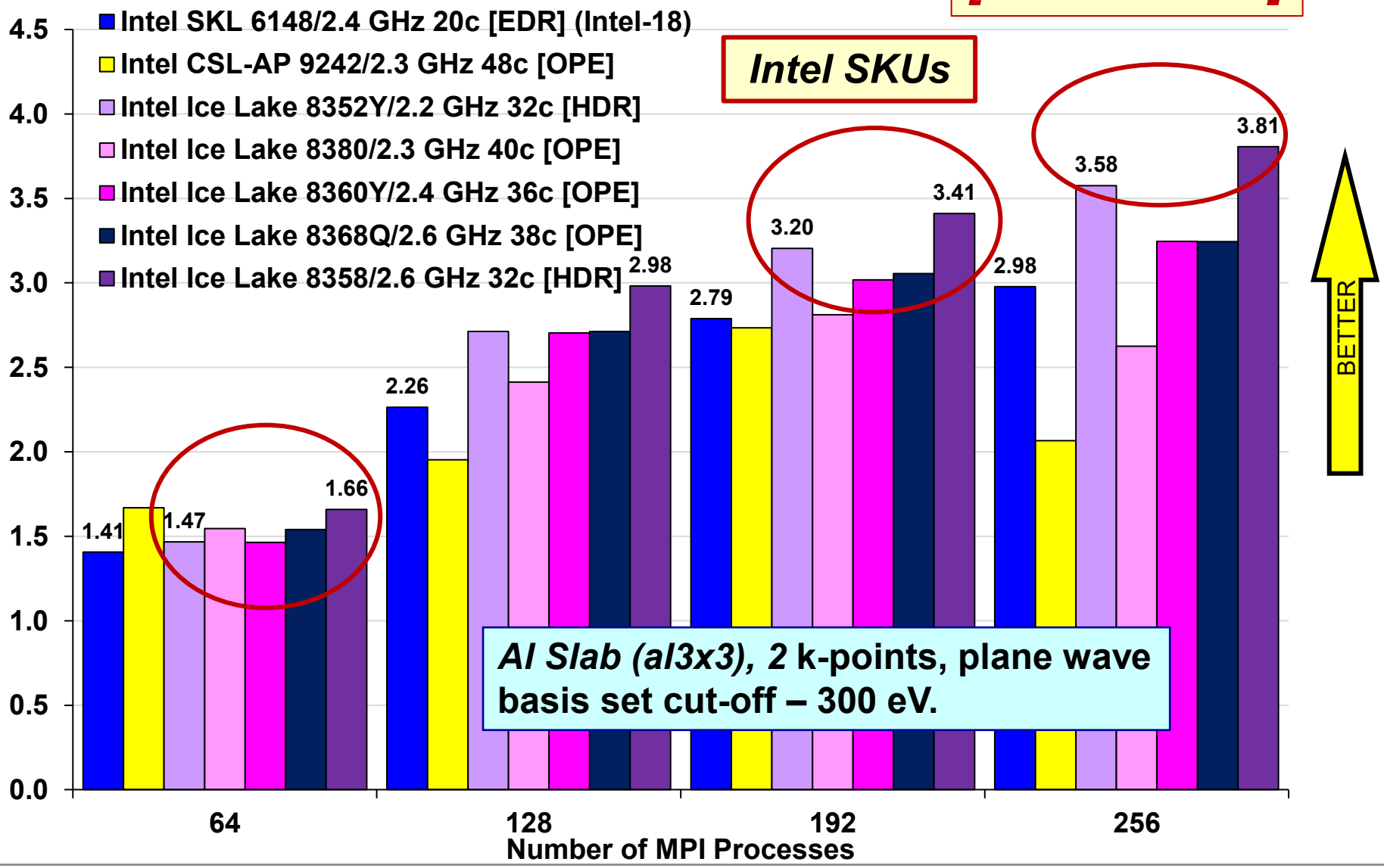
CASTEP – Impact of Intel MPI version on AMD clusters



CASTEP 19 – Al Slab (al3x3) Benchmark

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

[Core to core]

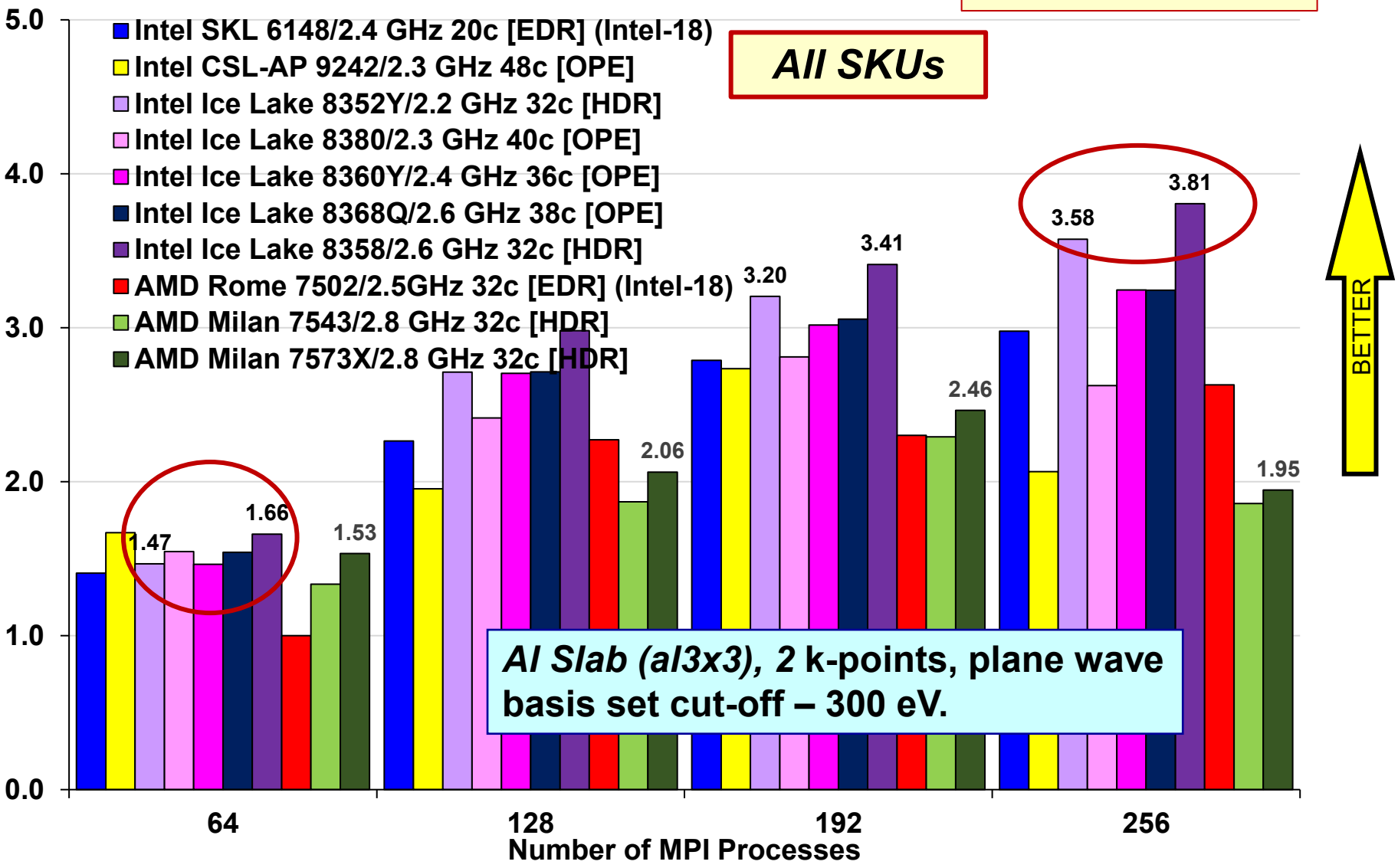


CASTEP 19 – Al Slab (al3x3) Benchmark

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

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

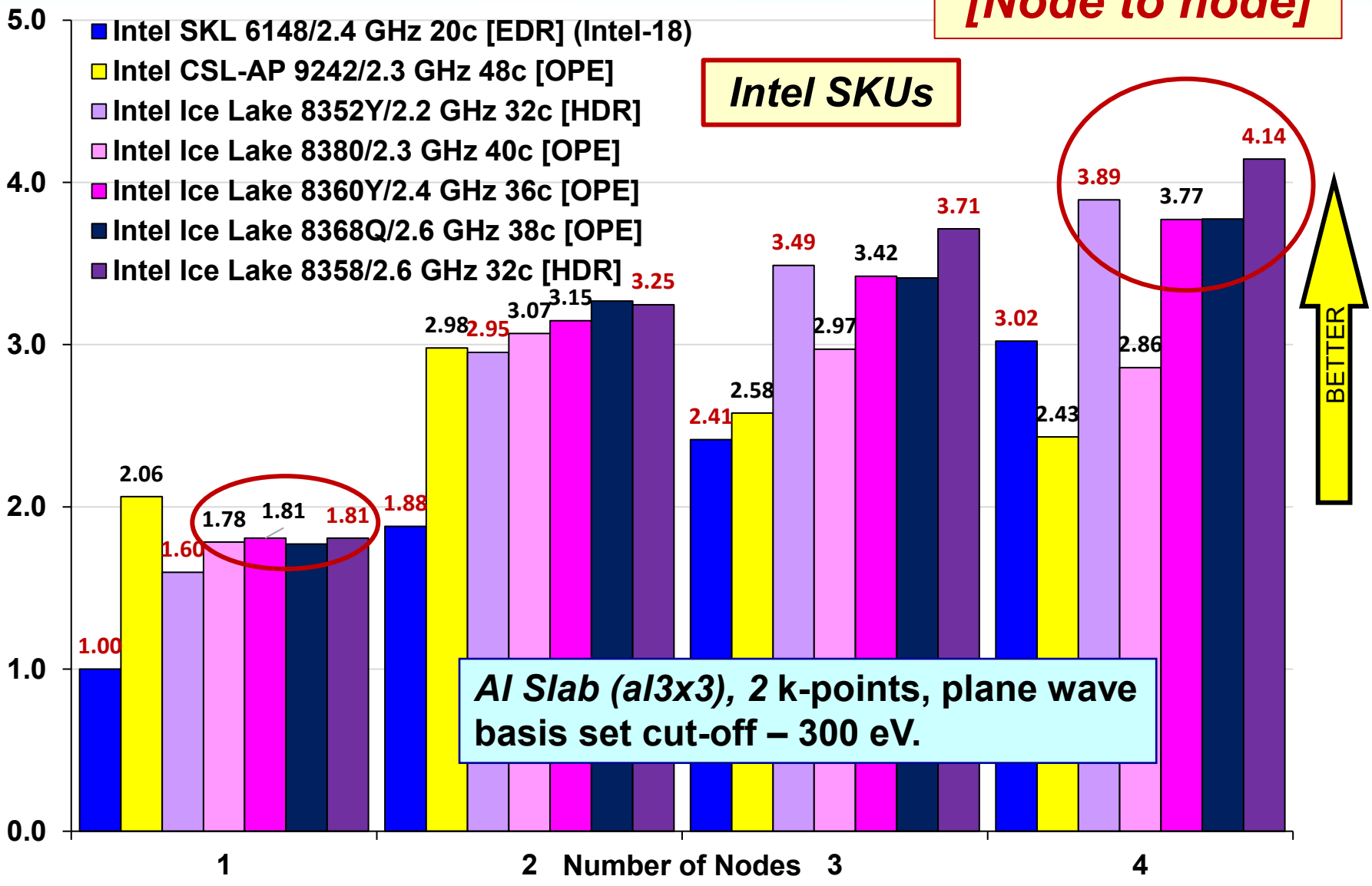
[Core to core]



CASTEP 19 – Al Slab (al3x3) Benchmark

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

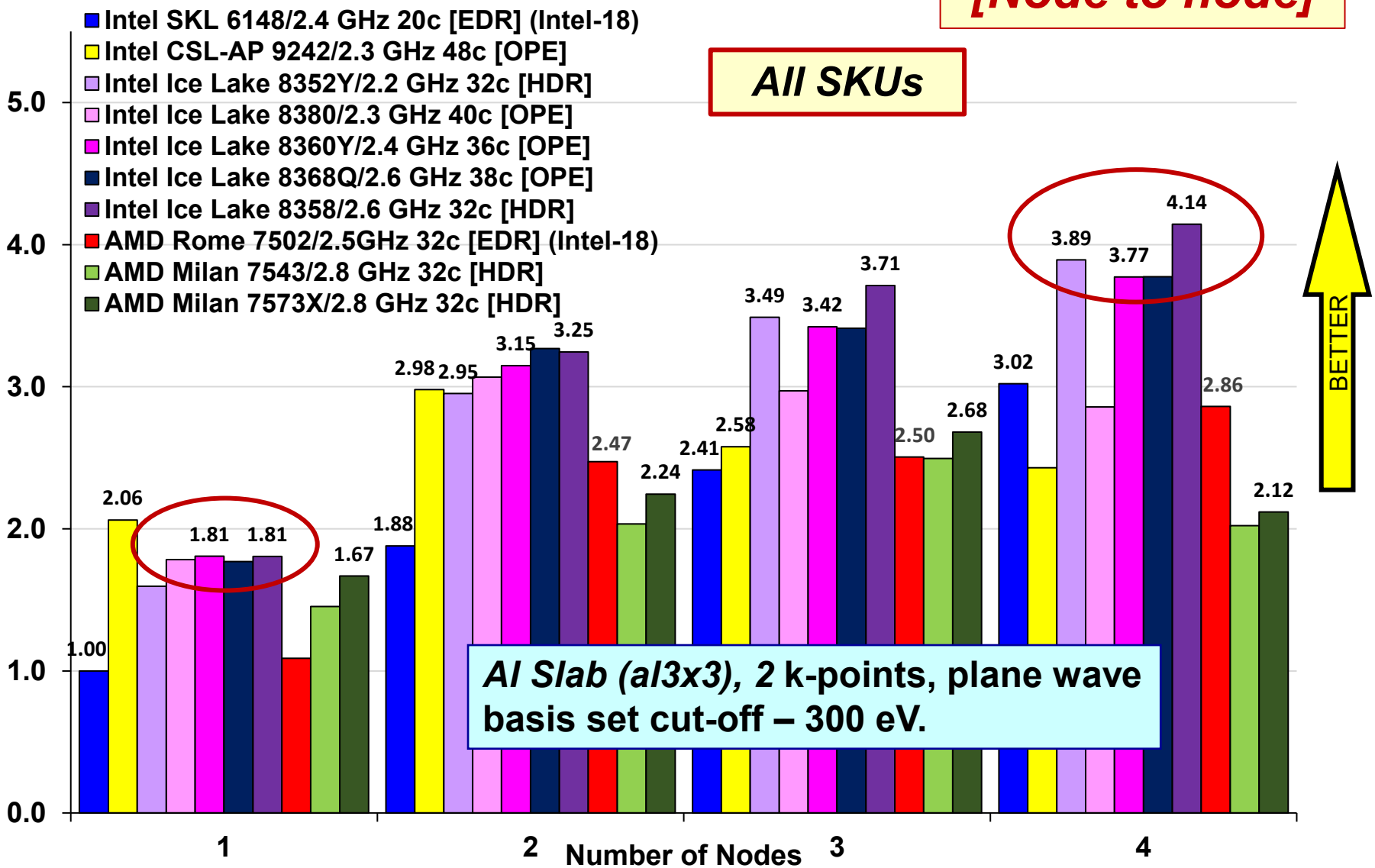
[Node to node]



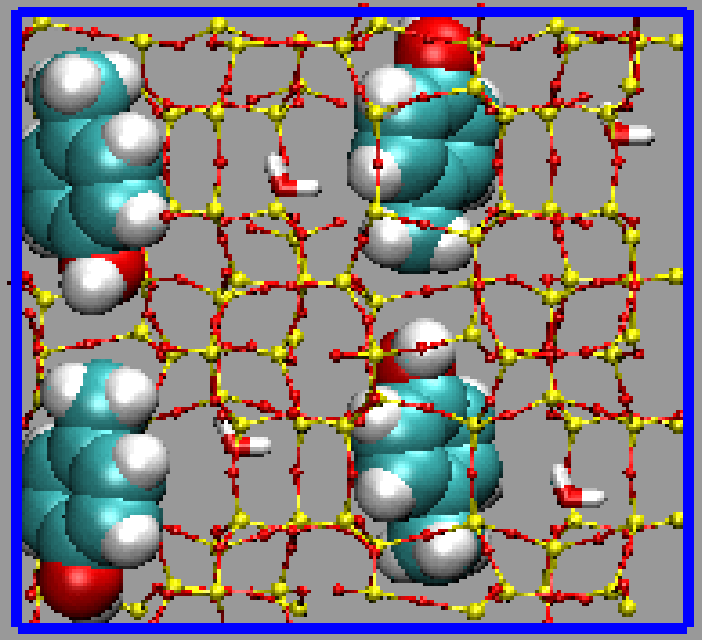
CASTEP 19 – Al Slab (al3x3) Benchmark

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

[Node to node]



Performance of Computational Chemistry and Ocean Modelling Codes



Electronic Structure GAMESS -UK

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

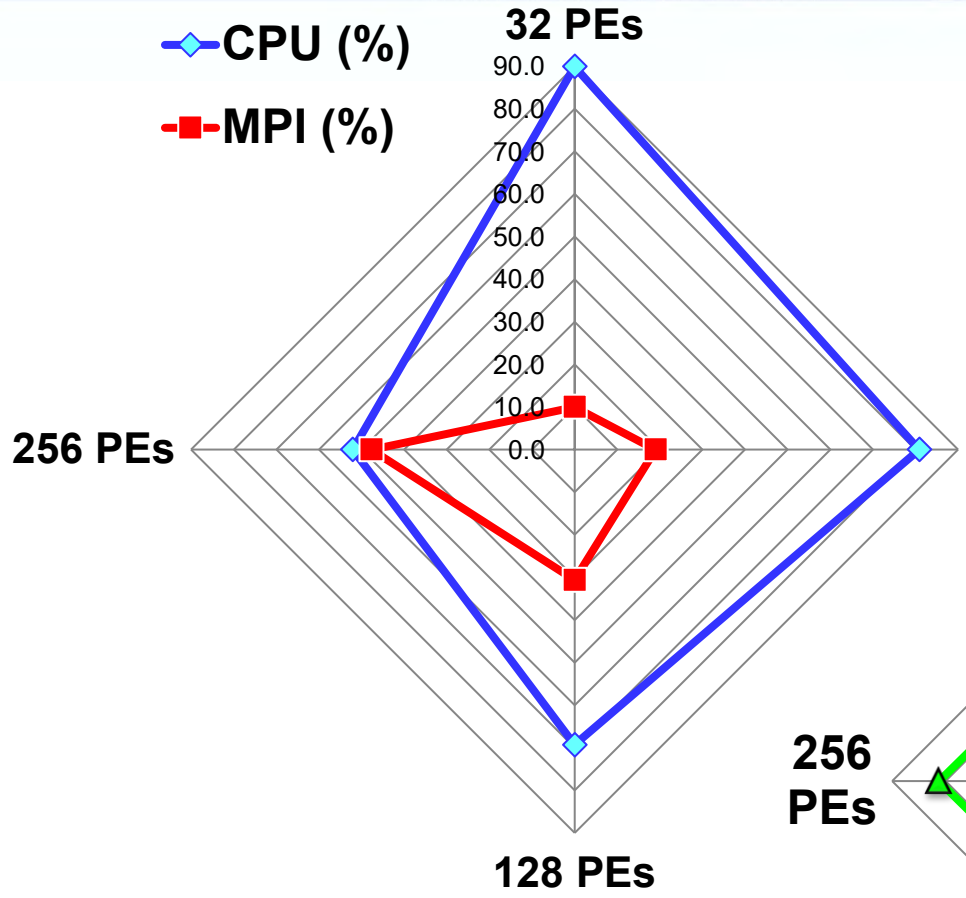
The MPI/ScaLAPACK Implementation of the GAMESS-UK SCF/DFT module

- Pragmatic approach to the replicated data constraints:
- MPI-based tools (such as ScaLAPACK) used in place of Global Arrays
- All data structures except those required for the Fock matrix build are fully distributed (F, P)

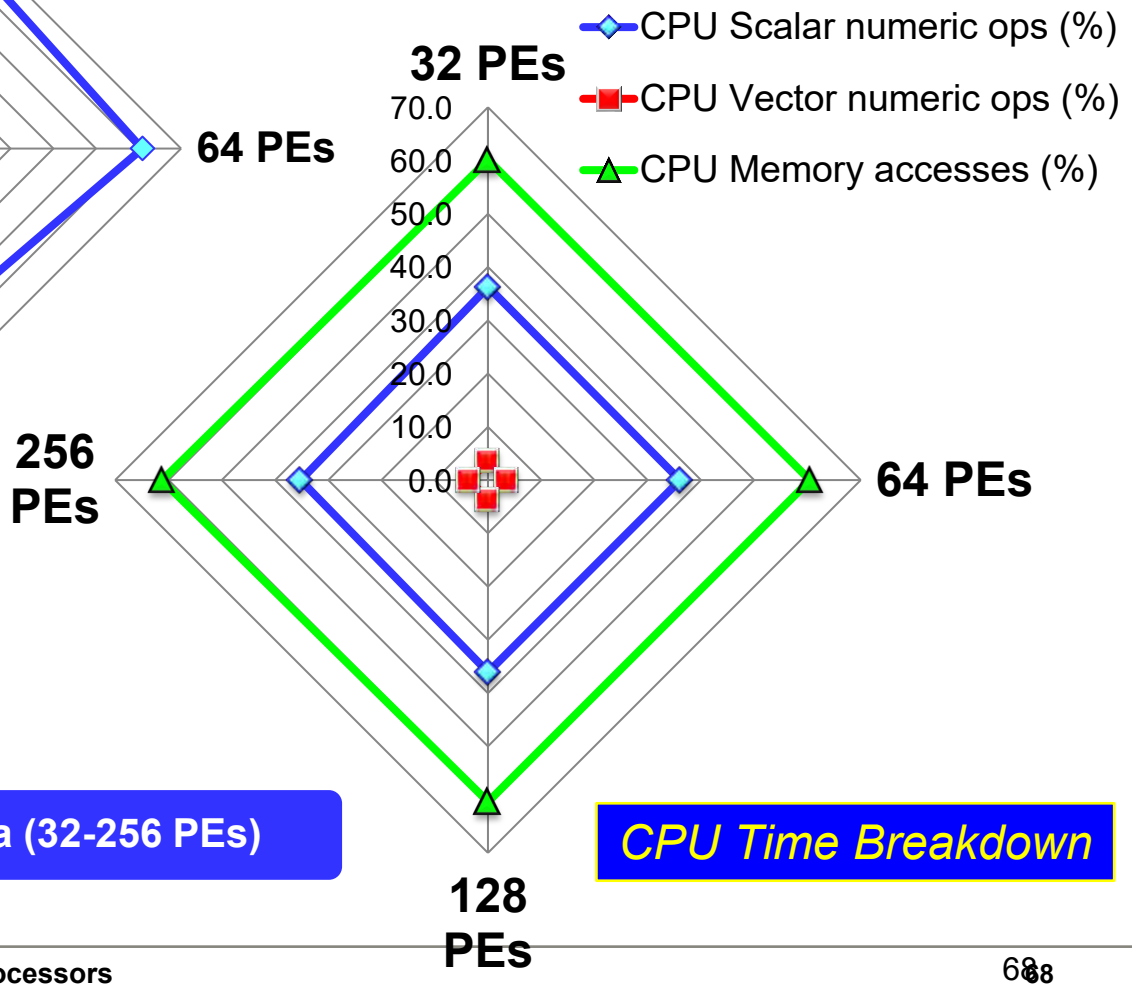
- Partially distributed model chosen because, in the absence of efficient one-sided communications it is difficult to efficiently load balance a distributed Fock matrix build.
- Obvious drawback - some large replicated data structures are required.
  - These are kept to a minimum. For a closed shell HF or DFT calculation only **2 replicated matrices** are required, 1 × Fock and 1 × Density (doubled for UHF).

"The GAMESS-UK electronic structure package: algorithms, developments and applications"
M.F. Guest, I. J. Bush, H.J.J. van Dam, P. Sherwood, J.M.H. Thomas, J.H. van Lenthe, R.W.A Havenith, J. Kendrick, Mol. Phys. 103, No. 6-8, 2005, 719-747.

GAMESS-UK.MPI DFT – DFT Performance Report



Cyclosporin 6-31G** basis (1855 GTOs); DFT B3LYP

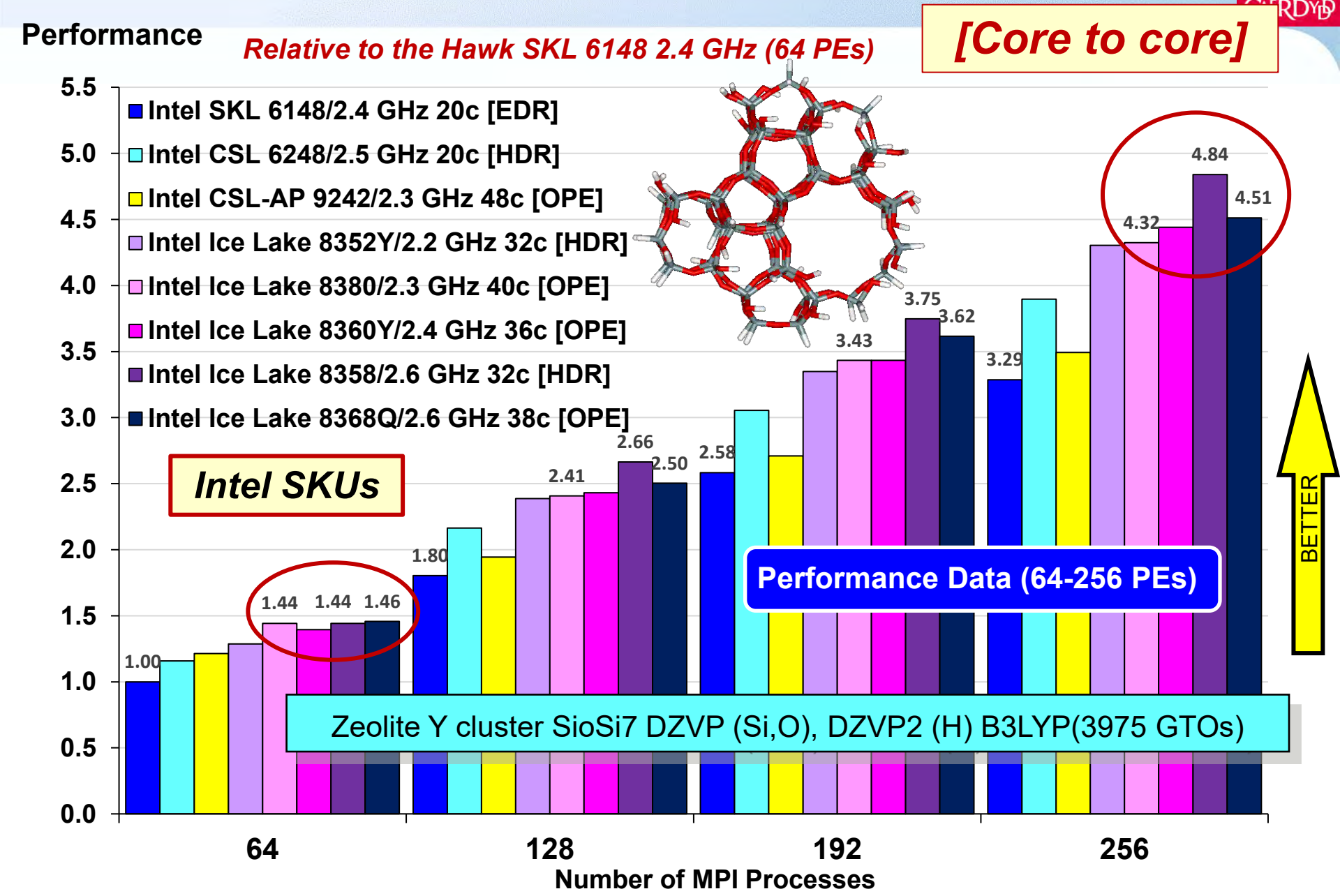


Total Wallclock Time Breakdown

Performance Data (32-256 PEs)

CPU Time Breakdown

GAMESS-UK Performance - Zeolite Y cluster



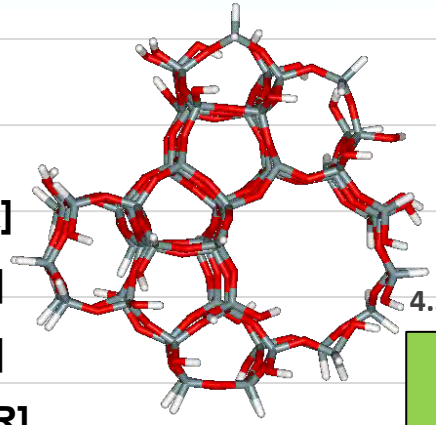
GAMESS-UK Performance - Zeolite Y cluster

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

[Core to core]

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

Performance



- Intel SKL 6148/2.4 GHz 20c [EDR]
- AMD Rome 7502/2.5GHz 32c [EDR]
- AMD Rome 7702/2.0 GHz 64c [HDR]
- AMD Milan 7713/2.0 GHz 64c [HDR]
- AMD Milan 7543/2.8 GHz 32c [HDR]
- AMD Milan 7573X/2.6 GHz 32c [HDR]

AMD SKUs

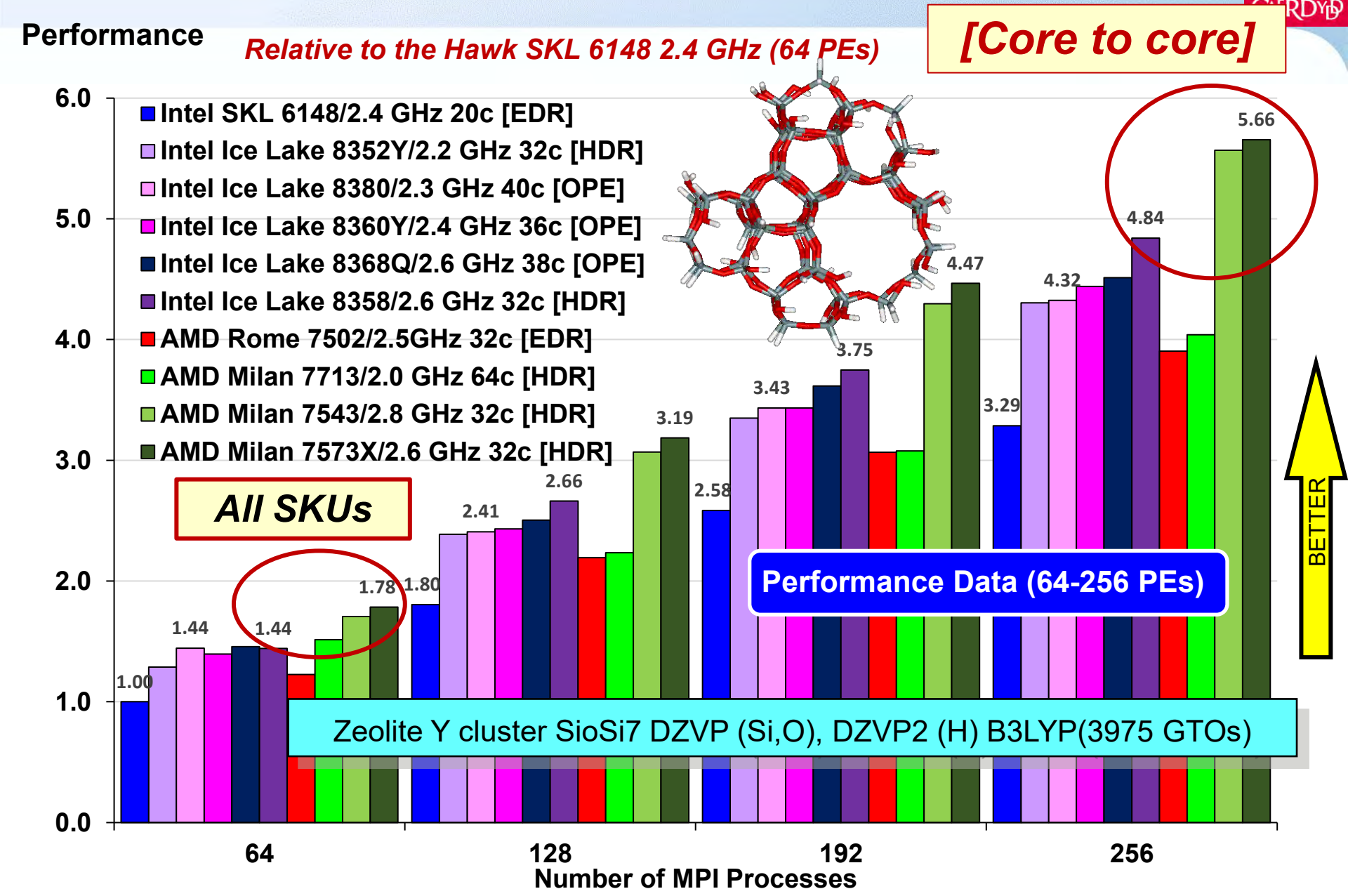
Number of MPI Processes	64	128	192	256
Intel SKL 6148/2.4 GHz 20c [EDR]	1.00	1.80	2.58	3.29
AMD Rome 7502/2.5GHz 32c [EDR]	1.23	2.19	3.07	3.90
AMD Rome 7702/2.0 GHz 64c [HDR]				
AMD Milan 7713/2.0 GHz 64c [HDR]				
AMD Milan 7543/2.8 GHz 32c [HDR]	1.71	1.99	3.07	4.30
AMD Milan 7573X/2.6 GHz 32c [HDR]	1.78	2.24	3.19	4.47
				5.57 / 5.66

Performance Data (64-256 PEs)

BETTER

Zeolite Y cluster SioSi7 DZVP (Si,O), DZVP2 (H) B3LYP(3975 GTOs)

GAMESS-UK Performance - Zeolite Y cluster

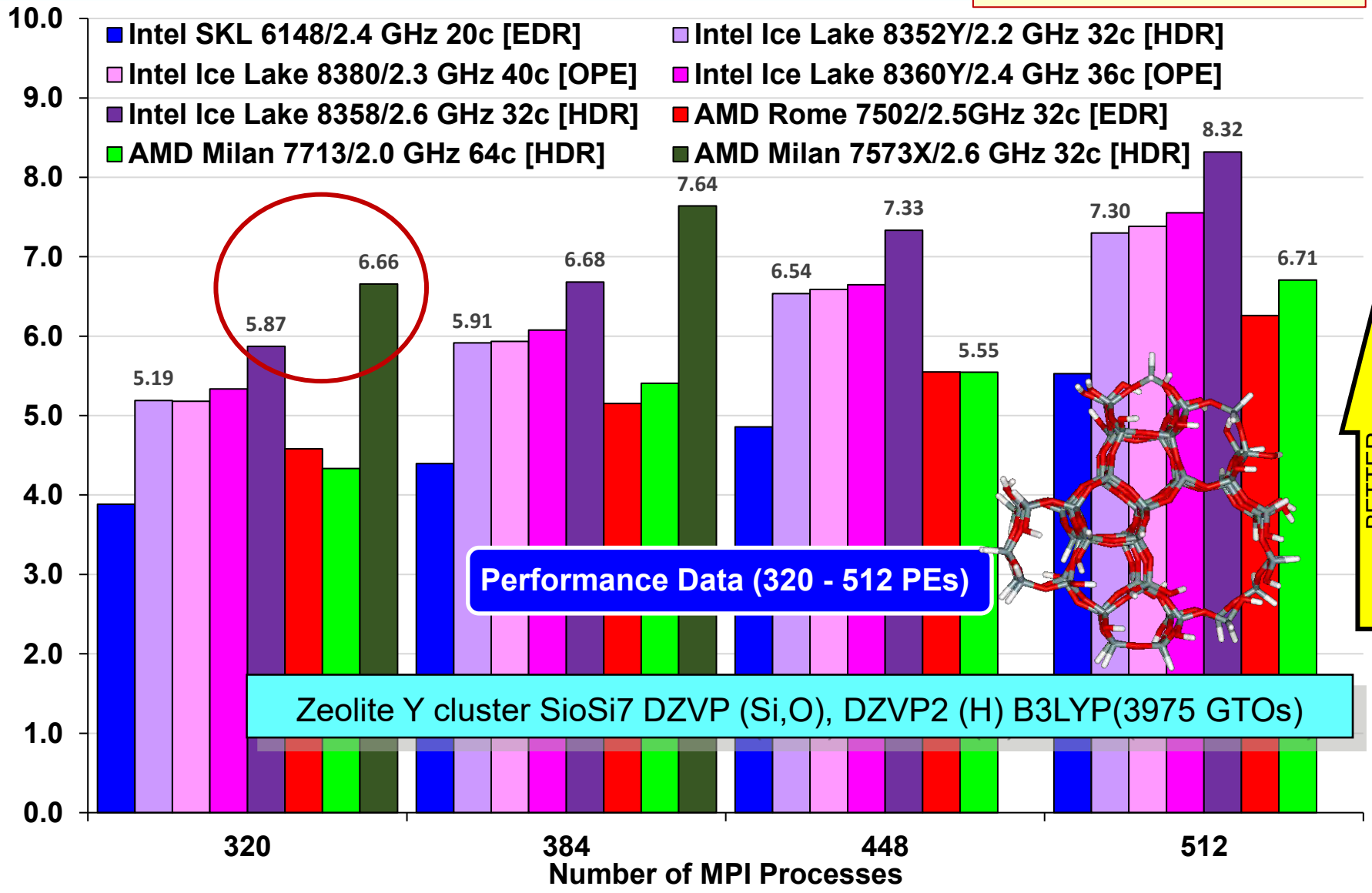


GAMESS-UK Performance - Zeolite Y cluster

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

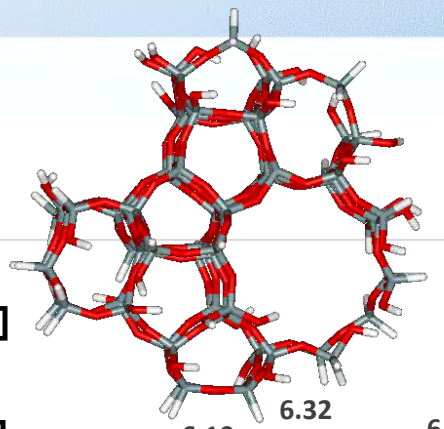
Performance *Relative to the Hawk SKL 6148 2.4 GHz (64 PEs)*

[Core to core]



GAMESS-UK Performance - Zeolite Y cluster

Relative to the Hawk SKL 6148 2.4 GHz (40 PEs)



Intel SKUs

- Intel SKL 6148/2.4 GHz 20c [EDR]
- Intel CSL 6248/2.5 GHz 20c [HDR]
- Intel CSL-AP 9242/2.3 GHz 48c [OPE]
- Intel Ice Lake 8352Y/2.2 GHz 32c [HDR]
- Intel Ice Lake 8380/2.3 GHz 40c [OPE]
- Intel Ice Lake 8360Y/2.4 GHz 36c [OPE]
- Intel Ice Lake 8358/2.6 GHz 32c [HDR]
- Intel Ice Lake 8368Q/2.6 GHz 38c [OPE]

Performance

8.0
6.0
4.0
2.0
0.0

1.00 1.16 2.44 2.54 2.31 2.42
1.88 2.24 4.35 4.72 4.35 4.64
2.60 3.12 6.10 6.32 5.90 6.16
3.51 4.23 7.64 8.35 7.78 8.14

1 2 Number of Nodes 3 4

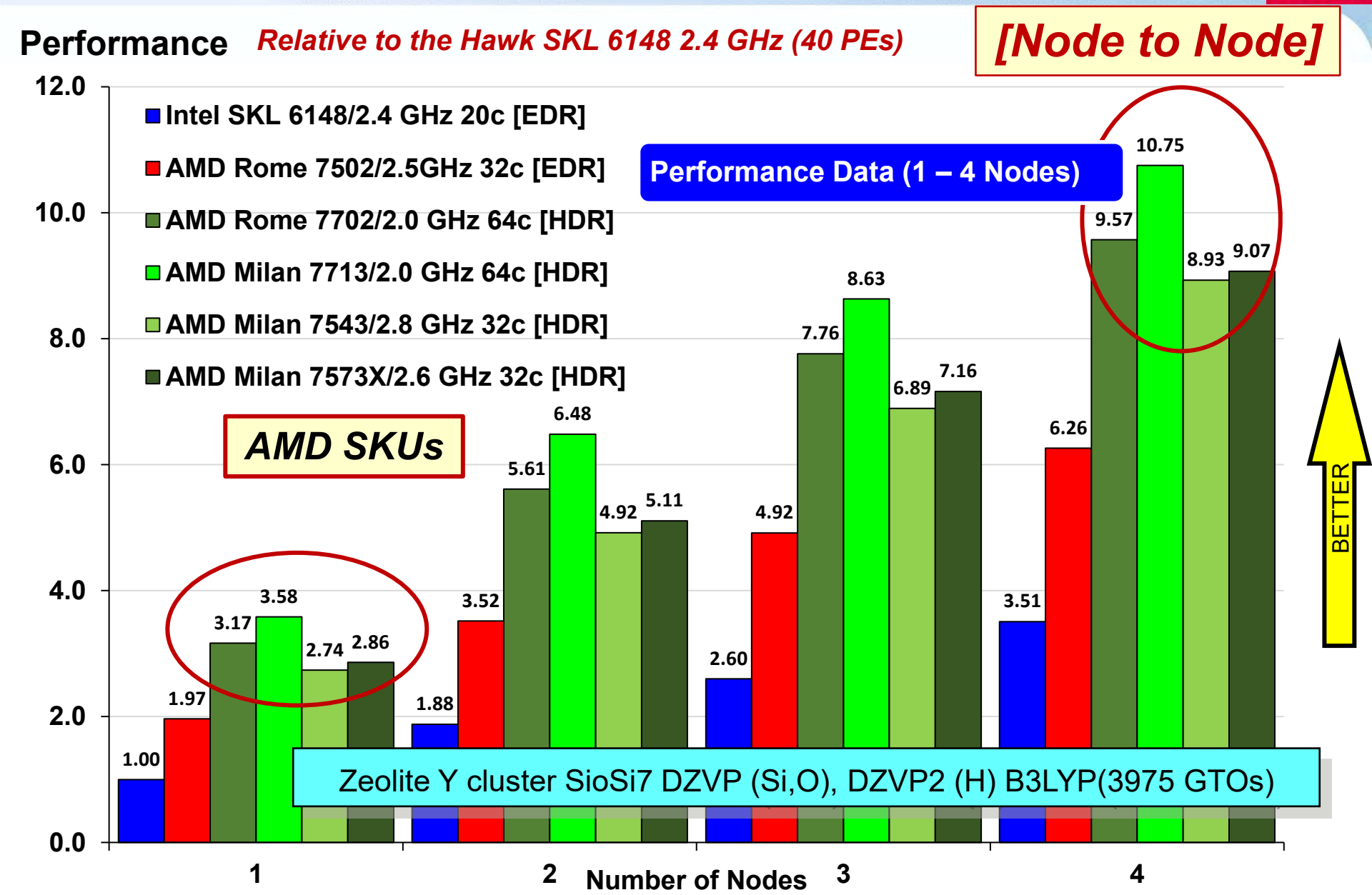
BETTER

[Node to Node]

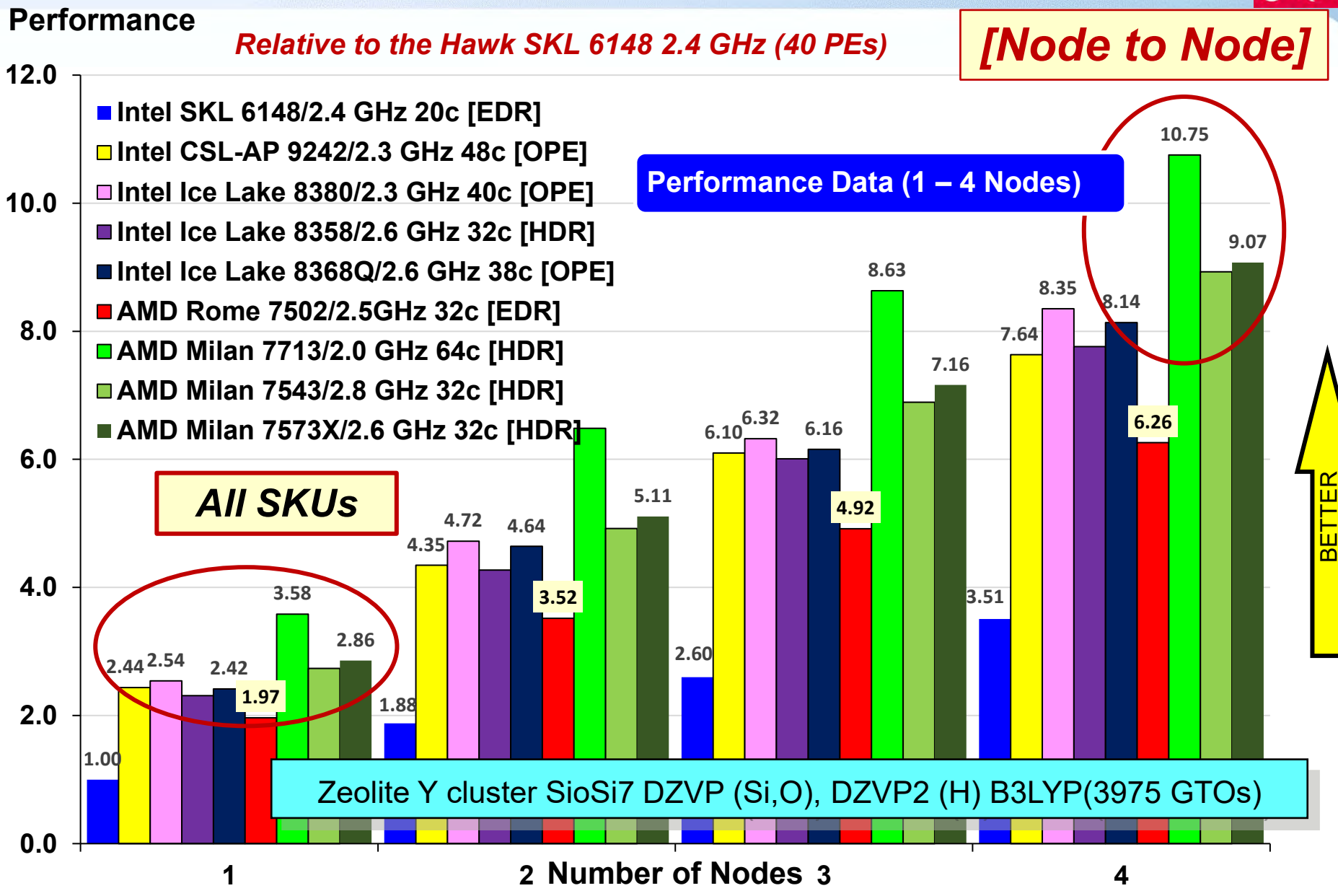
Performance Data (1 – 4 Nodes)

Zeolite Y cluster SioSi7 DZVP (Si,O), DZVP2 (H) B3LYP(3975 GTOs)

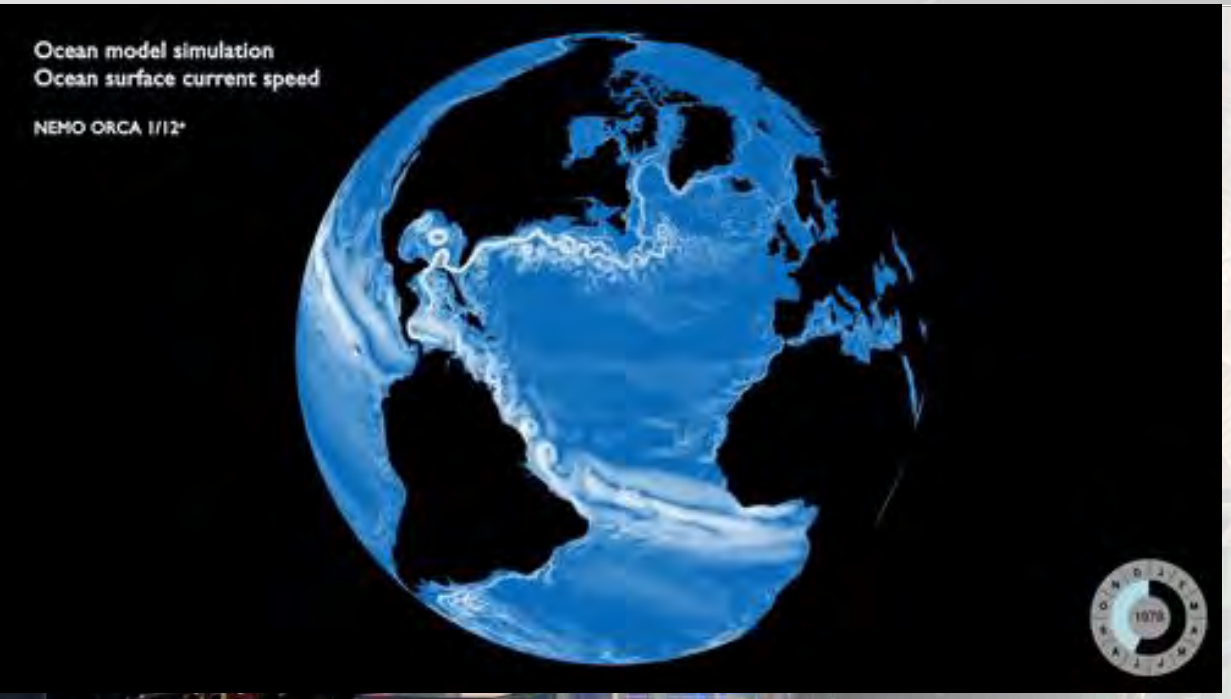
GAMESS-UK Performance - Zeolite Y cluster



GAMESS-UK Performance - Zeolite Y cluster



Performance of Computational Chemistry and Ocean Modelling Codes



Ocean Modelling: NEMO and FVCOM

Performance of Ocean Modelling Codes

- Assistance provided to **The Marine Systems Modelling Group at Plymouth Marine Laboratory.**
- At the heart of much of the group's work are two numerical models of the ocean's circulation:

  The NEMO Community Ocean Model

  A prognostic, primitive equation ocean circulation model for studying problems relating to both the global ocean and marginal seas. Uses a ***structured* model grid.**

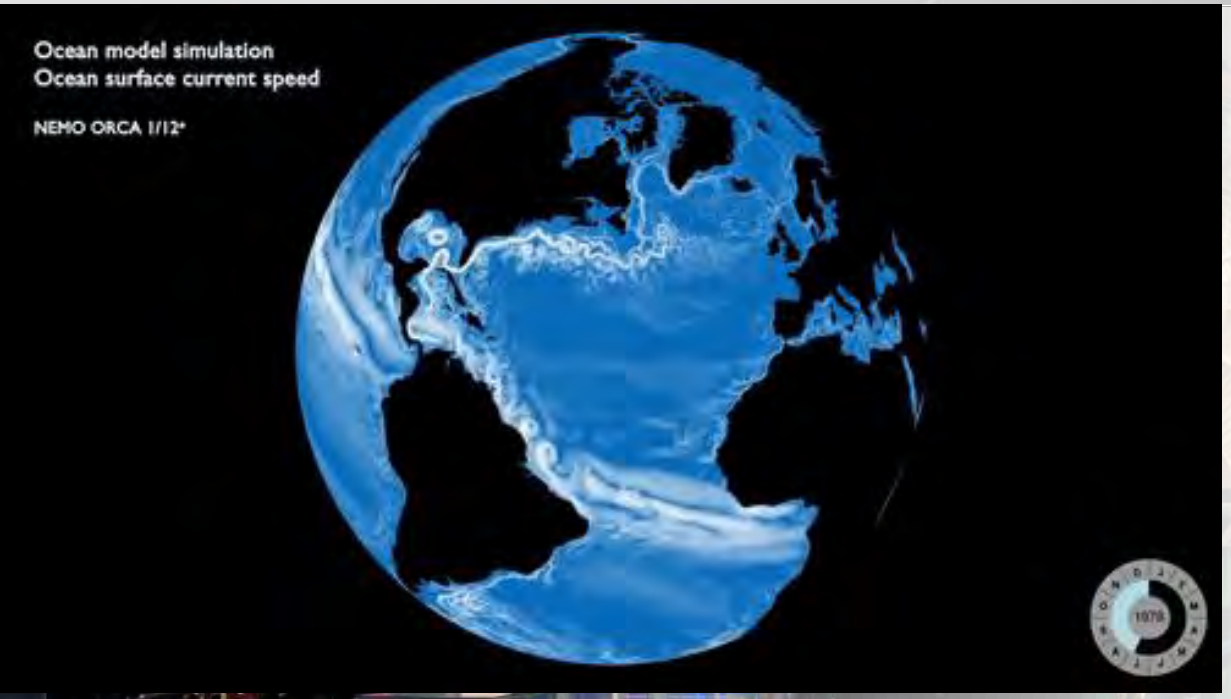
  The Finite Volume Community Ocean Model (FVCOM)

  A prognostic, primitive equation ocean circulation model for (mainly) studying problems relating to estuarine and coastal environments. **Uses an *unstructured* model grid.**

- Both models are often run with a **biogeochemical model called ERSEM -** significantly increases the compute & memory requirements.
- To be run efficiently, both models require a CPU based HPC system

Performance of Ocean Modelling Codes

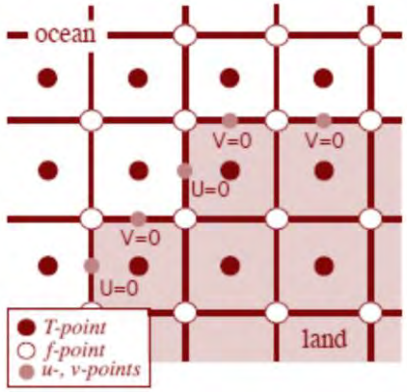
NEMO - Nucleus for European Modelling of the Ocean



1. NEMO

Primary Requirements for PML

NEMO* - Data parallelism through domain decomposition

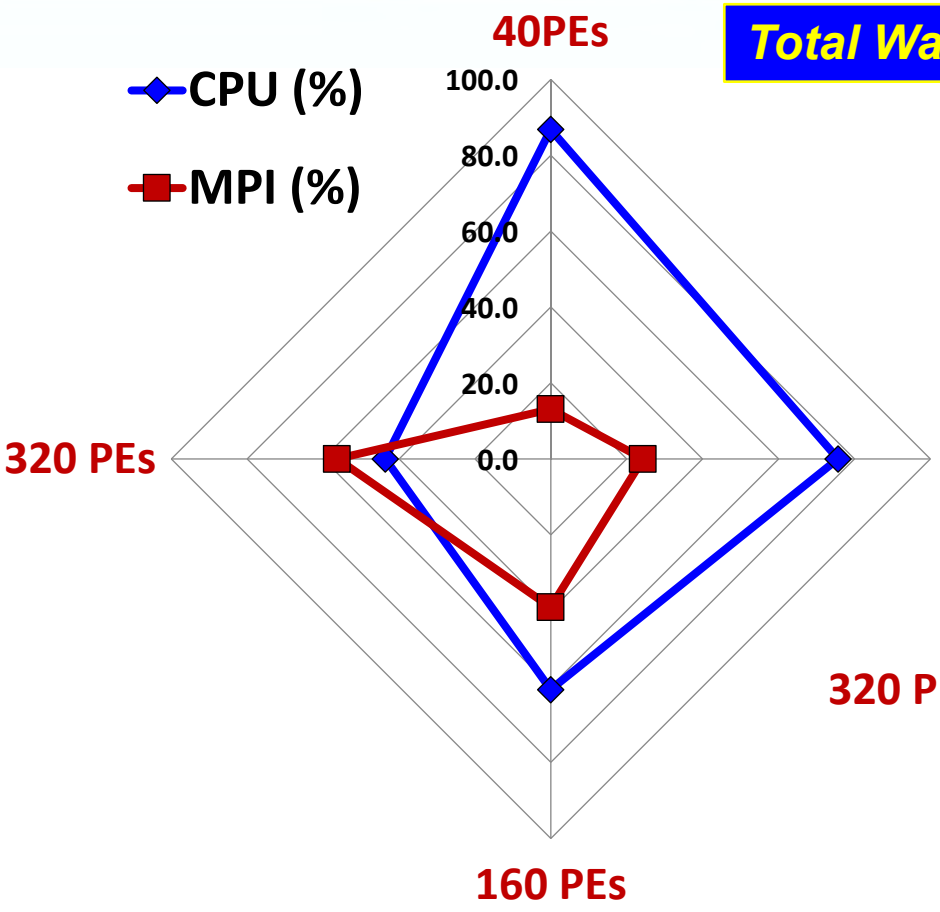


Example: The North Atlantic Ocean

- 773 x 1236 horizontal grid points, multiplied by 'k' depth levels.
- Full horizontal domain split into 9 x 20 sub-domains.
- Each subdomain is handled by a separate core during parallel runs.
- MPI for handling communication between subdomains.
- Known memory B/W issues – avoid full node occupancy

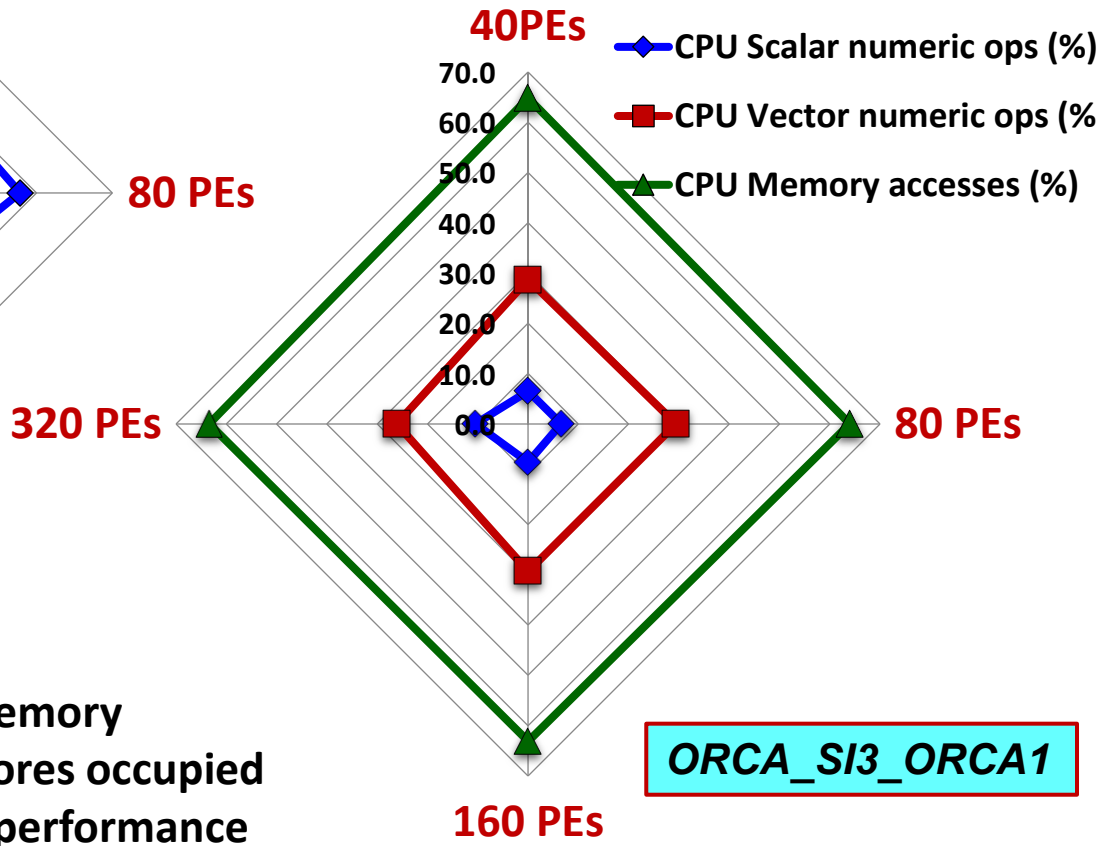
***FVCOM** employs a similar approach to parallelism, albeit based upon an unstructured, triangular horizontal mesh.

NEMO – ORCA_SI3 Model 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.

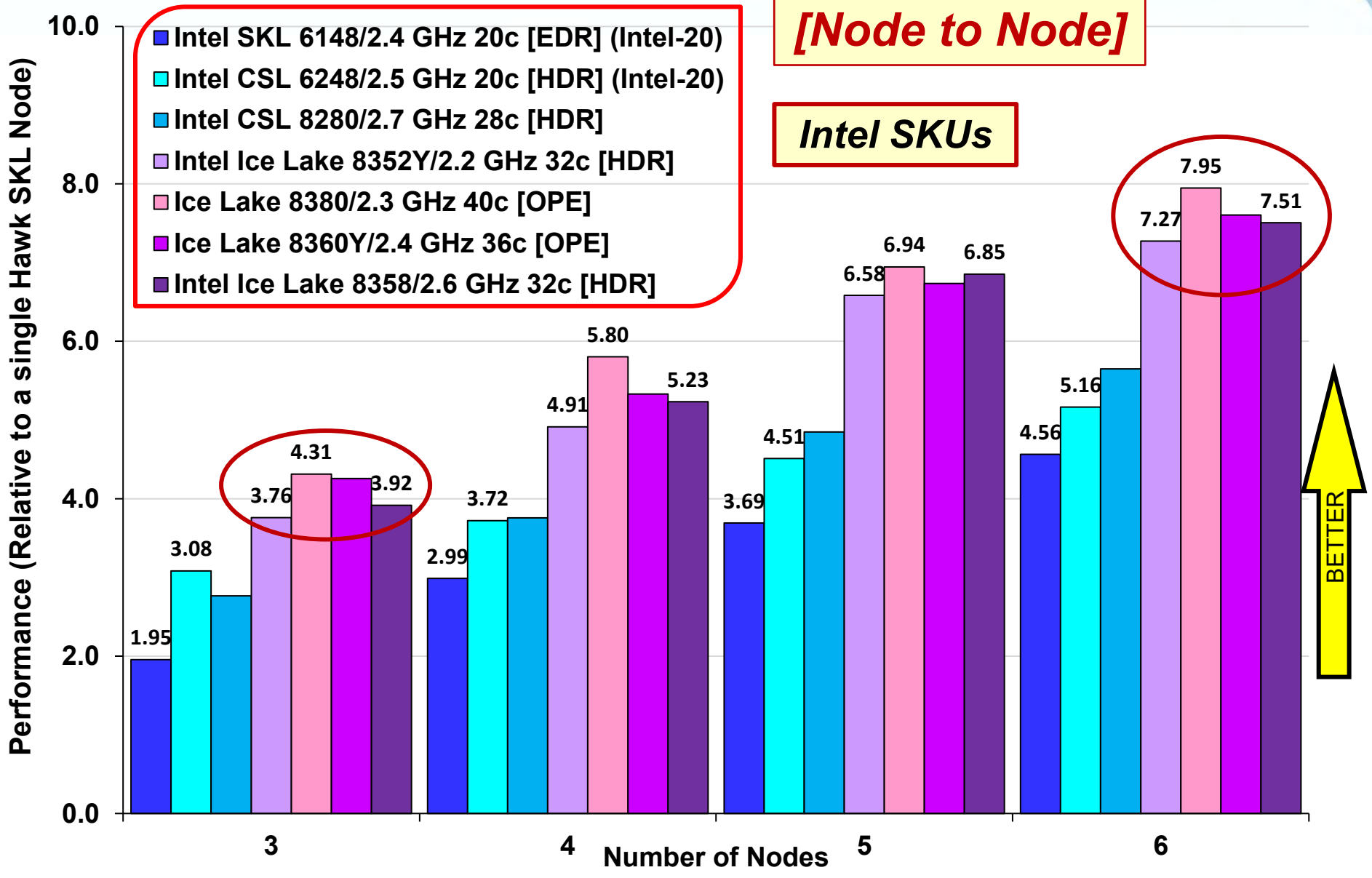
CPU Time Breakdown

The NEMO-ERSEM Benchmark

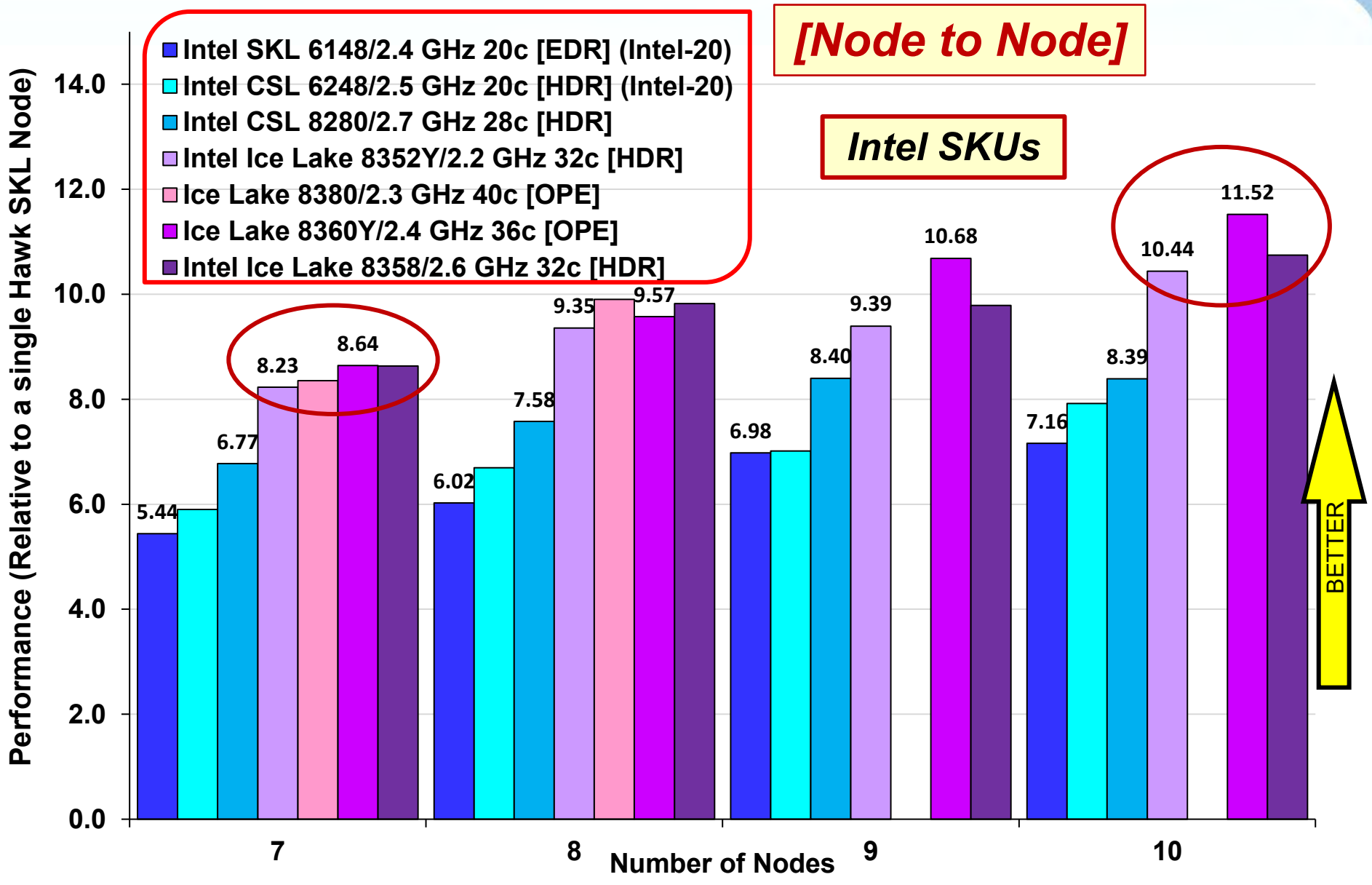
- NEMO, "Nucleus for European Modelling of the Ocean" is a modelling framework for research activities and forecasting services in ocean and climate sciences, developed by a European consortium. (https://www.nemo-ocean.eu)
- NEMO is a **memory-bandwidth limited code** where performance can be improved by part-populating nodes.
- ERSEM, "European Regional Seas Ecosystem Model" is a bio-geochemical and ecosystem mode, developed at PML (https://github.com/pmlmodelling/ersem)
- **Benchmark Case:** NEMO-FABM-ERSEM on the AMM7 domain covering the NW European shelf at ca. 7 km resolution. Four elements to the code (a) **XIOS**: an I/O library, (b) **ERSEM:** Biogeochemical model code, (c) **FABM:** Interface between ERSEM and NEMO and (d) **NEMO**.
- Compilation requires **parallel netcdf and hdf5 libraries**. Several cores are allocated to the I/O server XIOS, with remainder allocated to NEMO:

```
mpirun -n $XIOSCORES $code_xios : -n $OCEANCORES $code_nemo
```

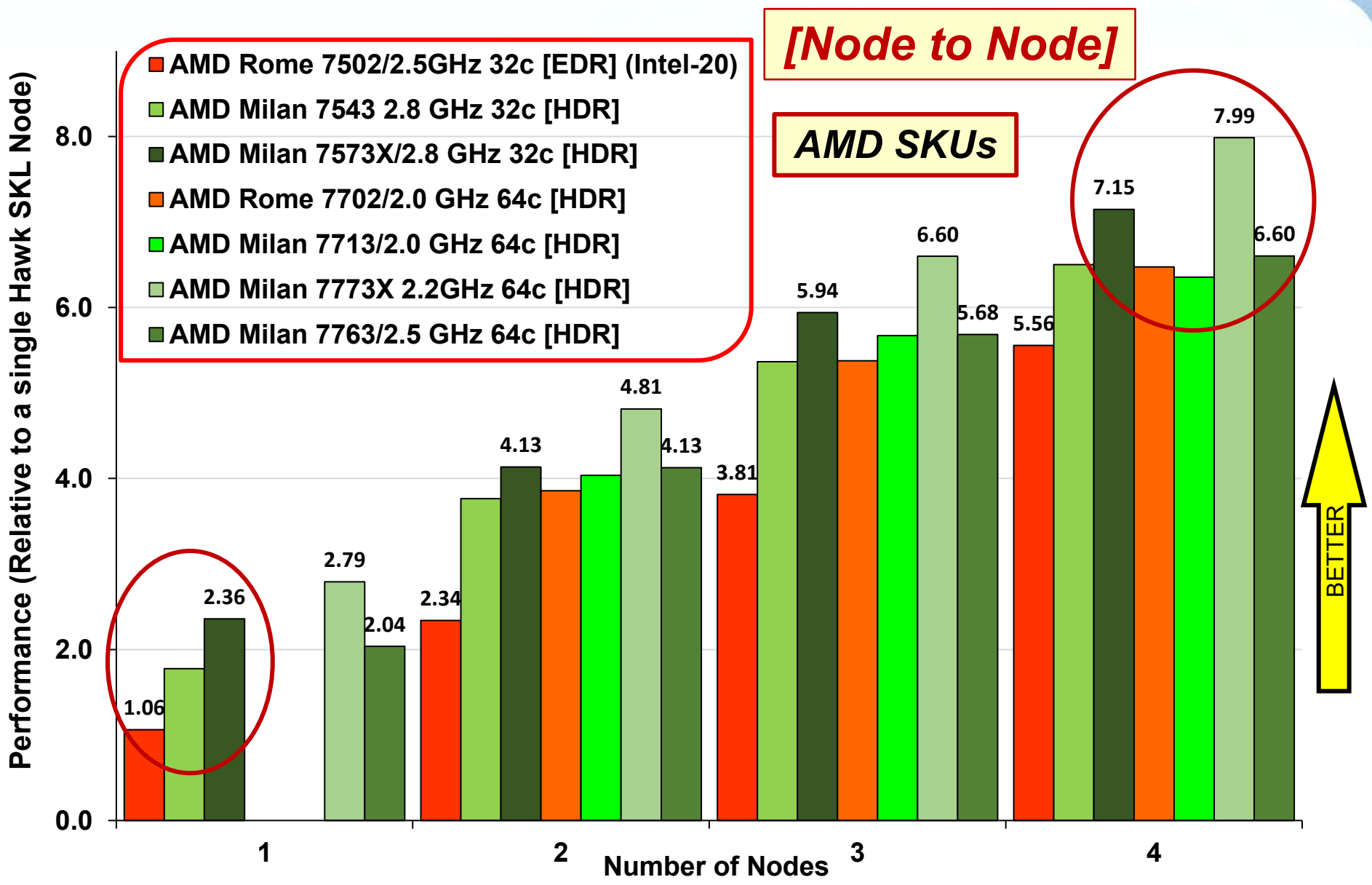
NEMO-FABM-ERSEM (AMM7) – Node Performance



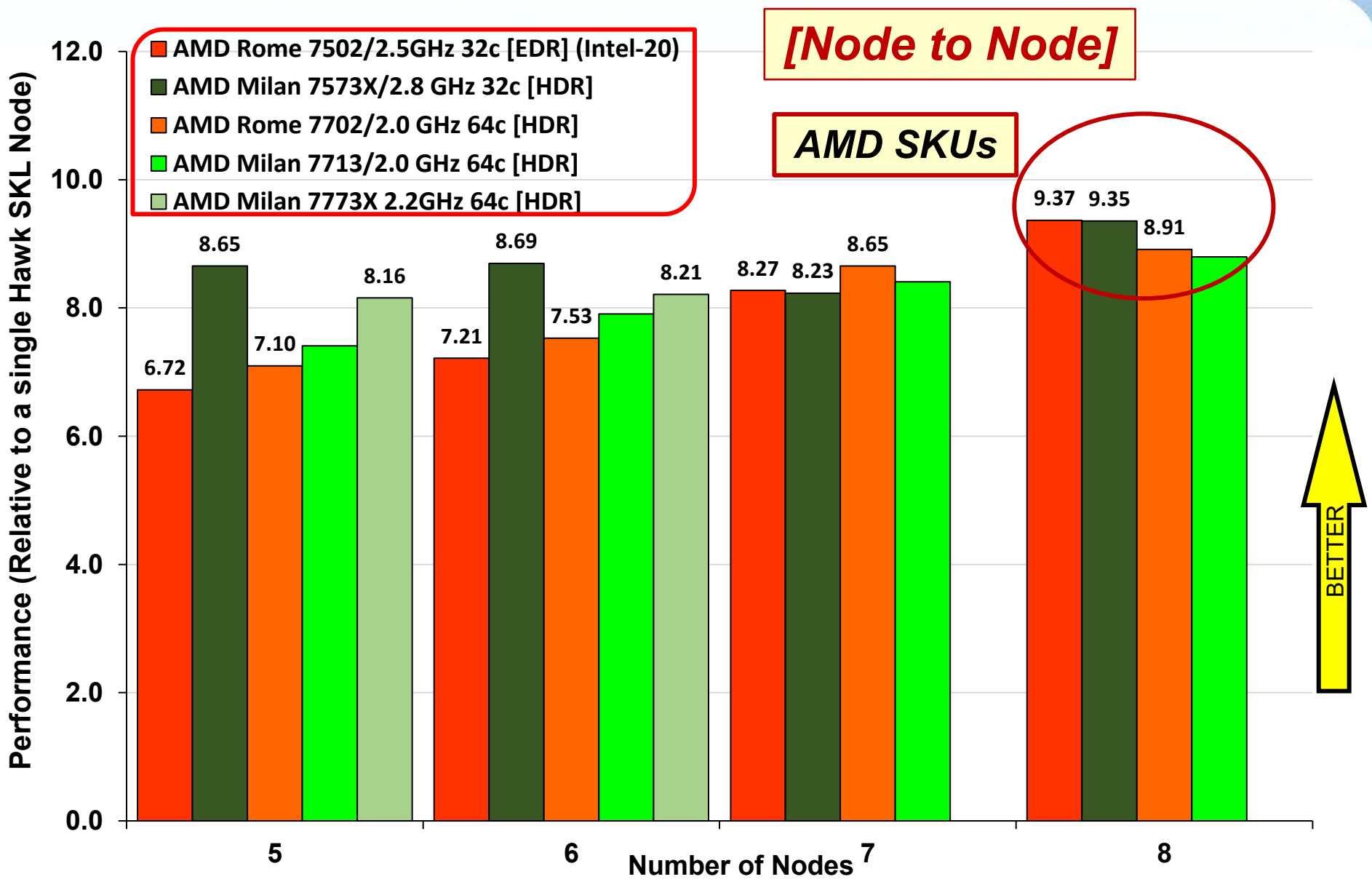
NEMO-FABM-ERSEM (AMM7) – Node Performance



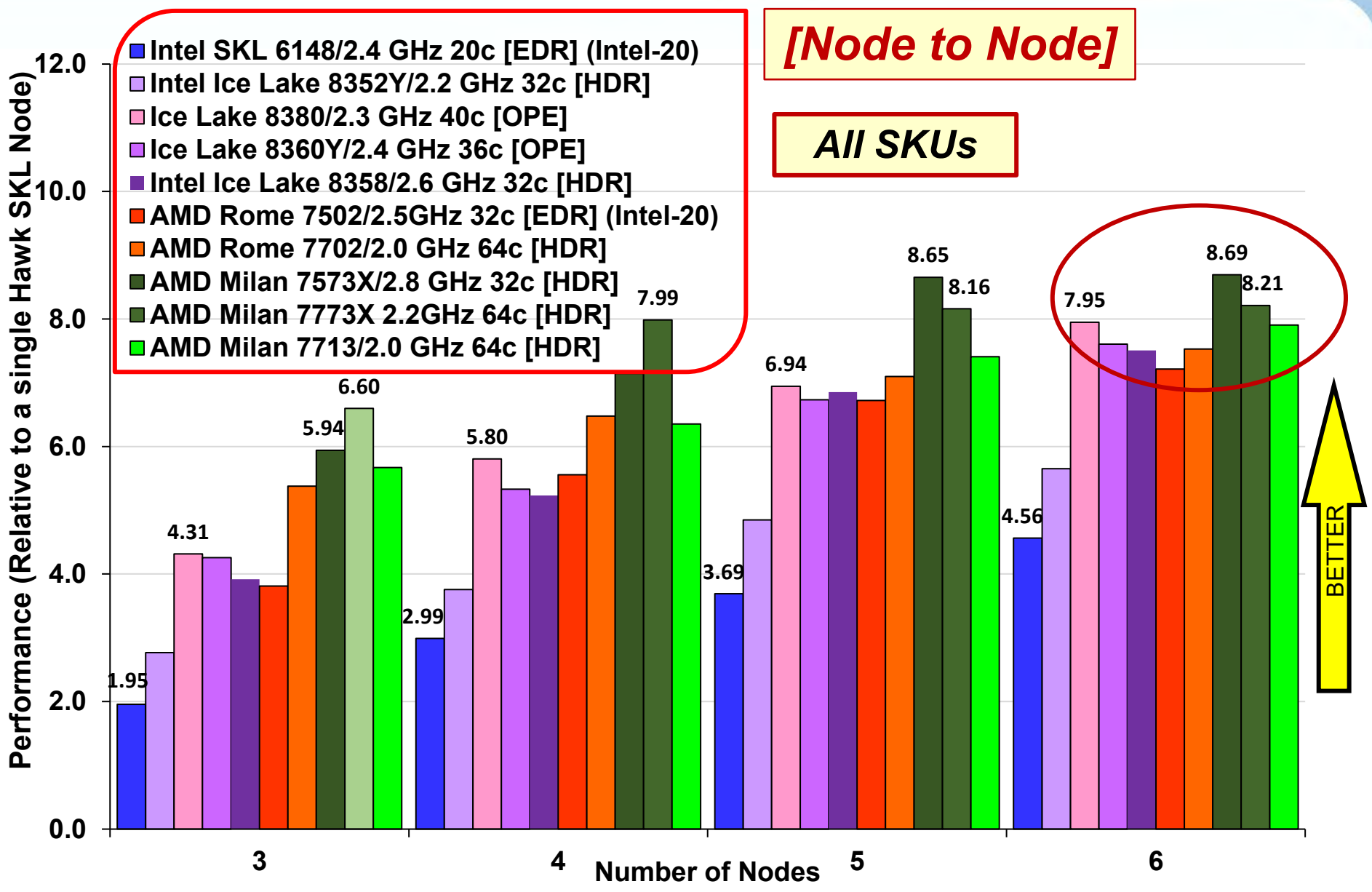
NEMO-FABM-ERSEM (AMM7) – Node Performance



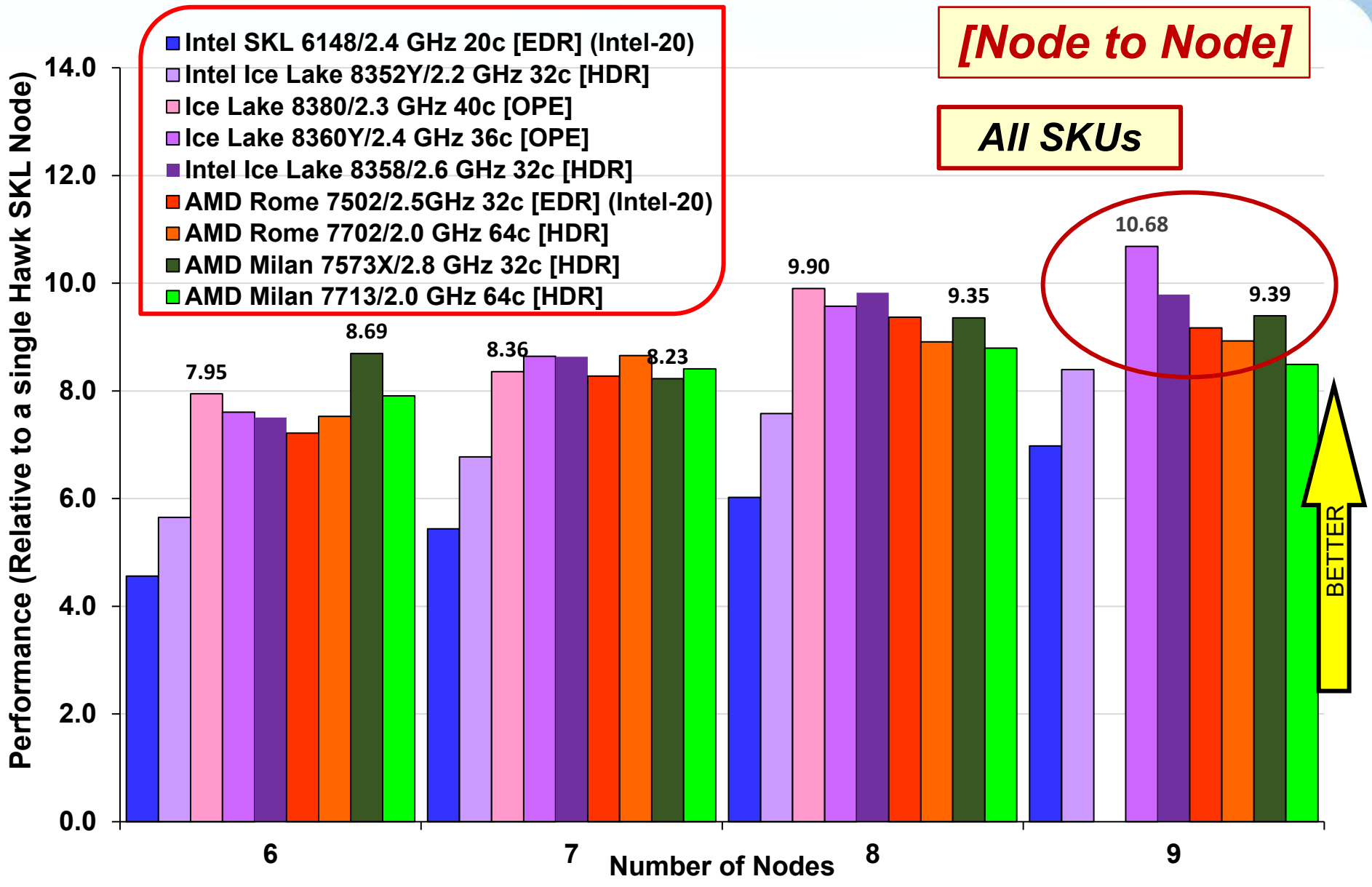
NEMO-FABM-ERSEM (AMM7) – Node Performance



NEMO-FABM-ERSEM (AMM7) – Node Performance

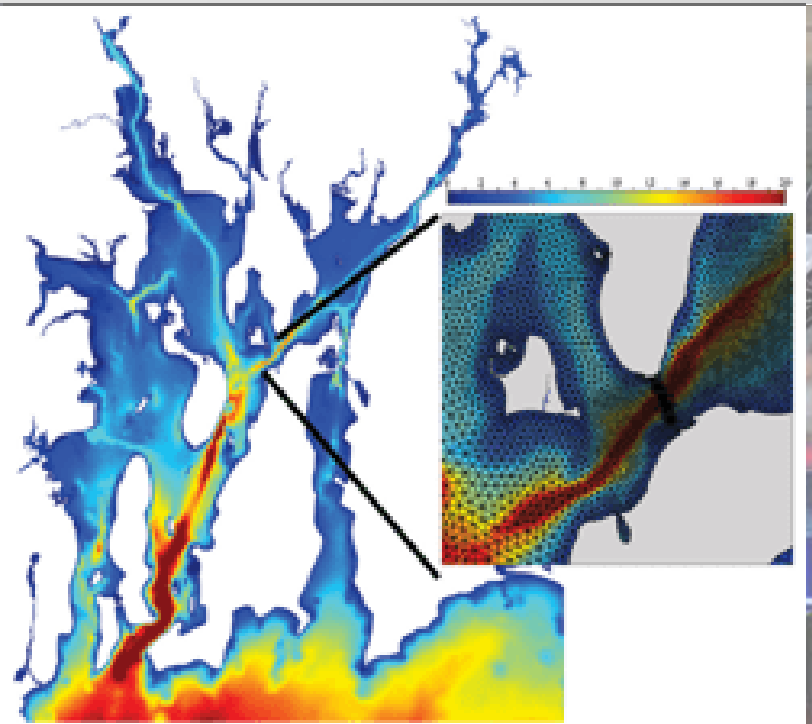


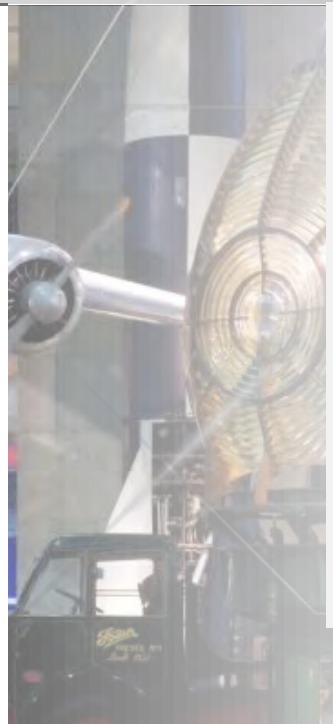
NEMO-FABM-ERSEM (AMM7) – Node Performance



Performance of Ocean Modelling Codes

The Finite Volume Community Ocean Model (FVCOM)





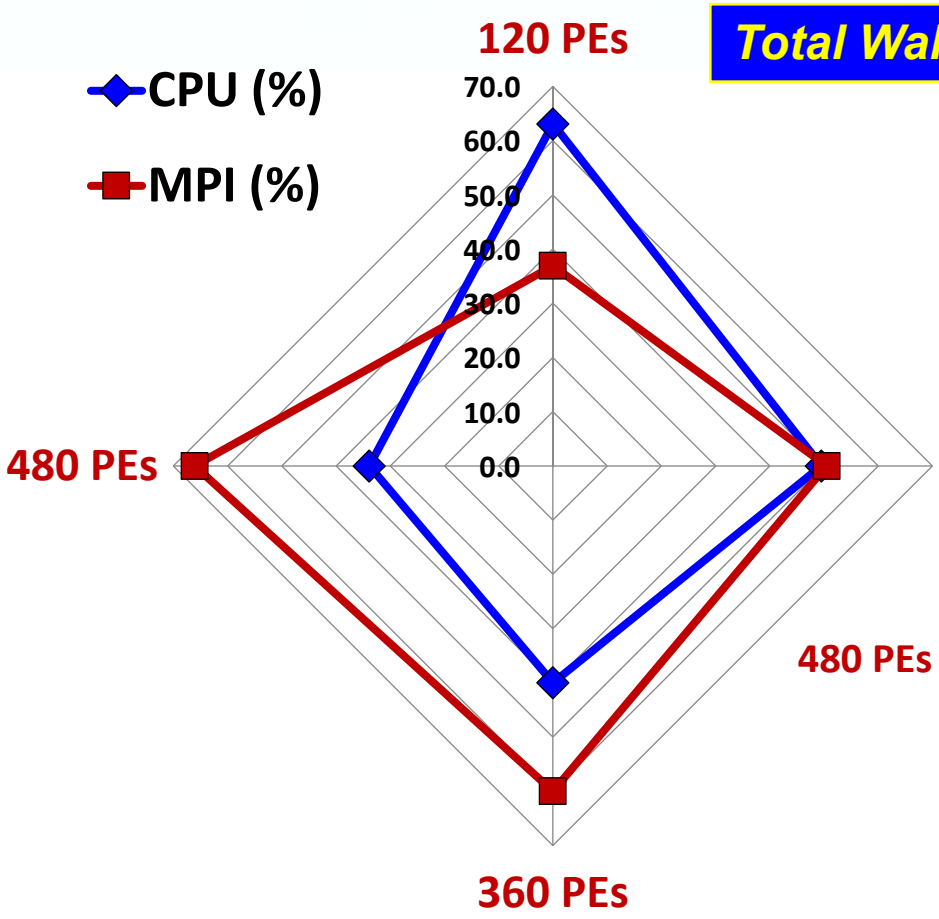
2. FVCOM

An ocean circulation model for (mainly) studying problems relating to estuarine and coastal environments.
Uses an *unstructured* model grid.

The FVCOM-ERSEM Benchmark

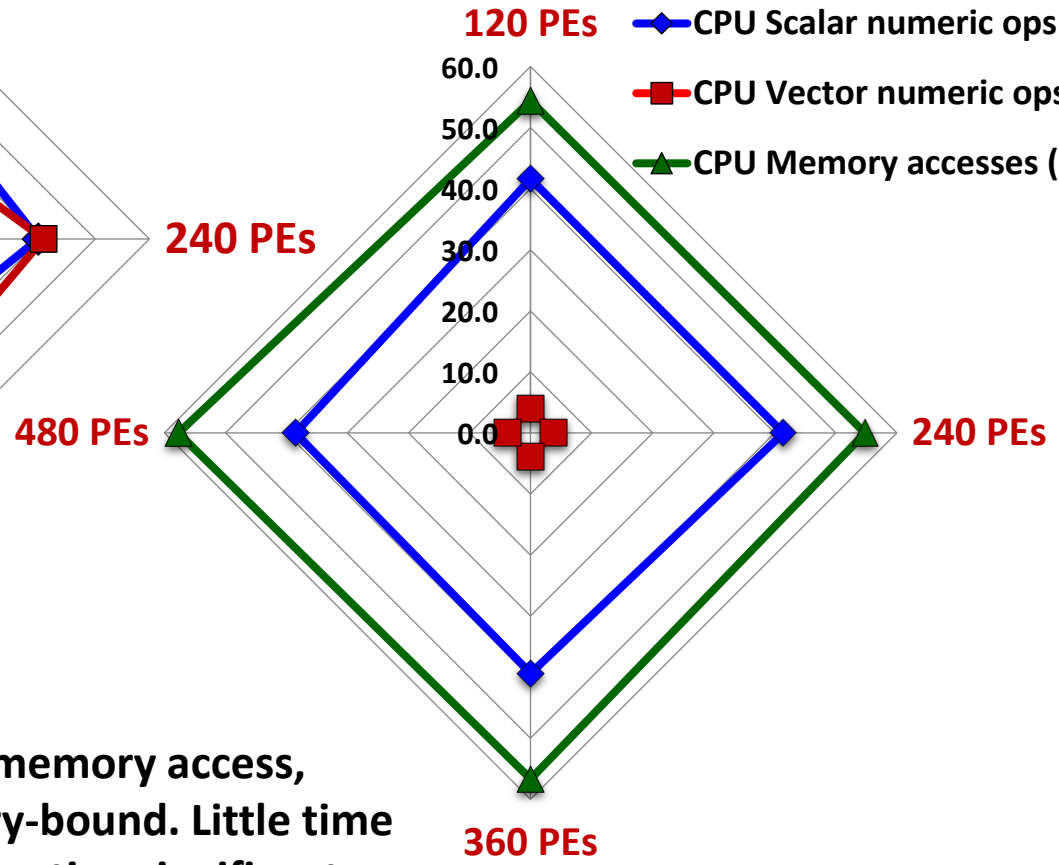
- **FVCOM**, "Finite Volume Community Ocean model" is a prognostic, unstructured-grid, finite-volume, free-surface, 3-D primitive equation coastal ocean circulation model developed by UMASSD-WHOI in the US and based on a triangular mesh. (http://fvcom.smast.umassd.edu/fvcom/)
- **ERSEM**, "European Regional Seas Ecosystem Model" is a bio-geochemical and ecosystem mode, developed at PML (https://github.com/pmlmodelling/ersem)
- Compilation requires **parallel netcdf and hdf5 libraries**.
- *Performance Report* highlights major features of the code. Performance **dominated by memory access**, with the per-core performance memory-bound.
- **Little time spent in vectorized instructions**, suggesting significant opportunities for improving code performance.

FVCOM – Performance Report



Total Wallclock Time Breakdown

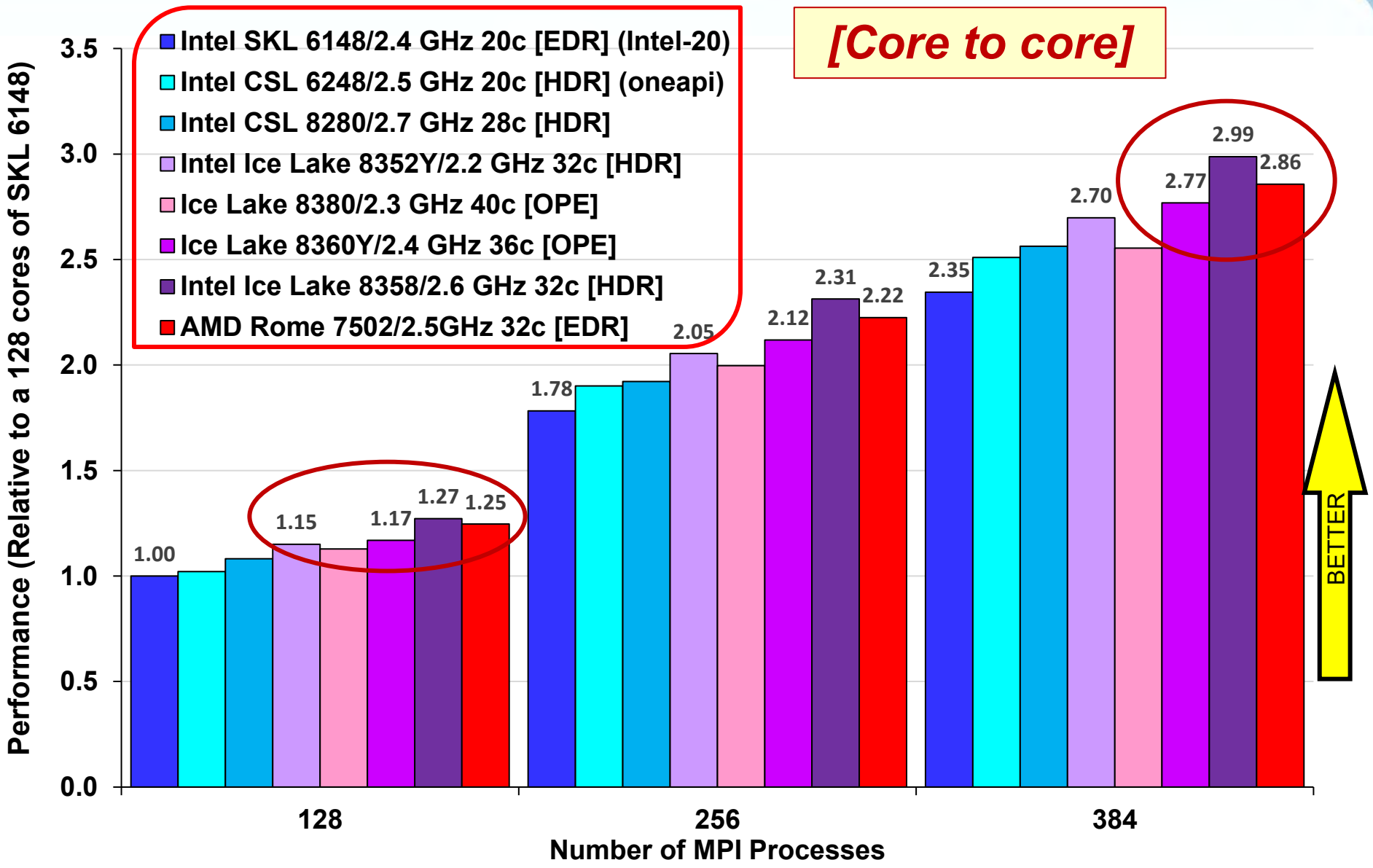
Performance Data (120-480 PEs)



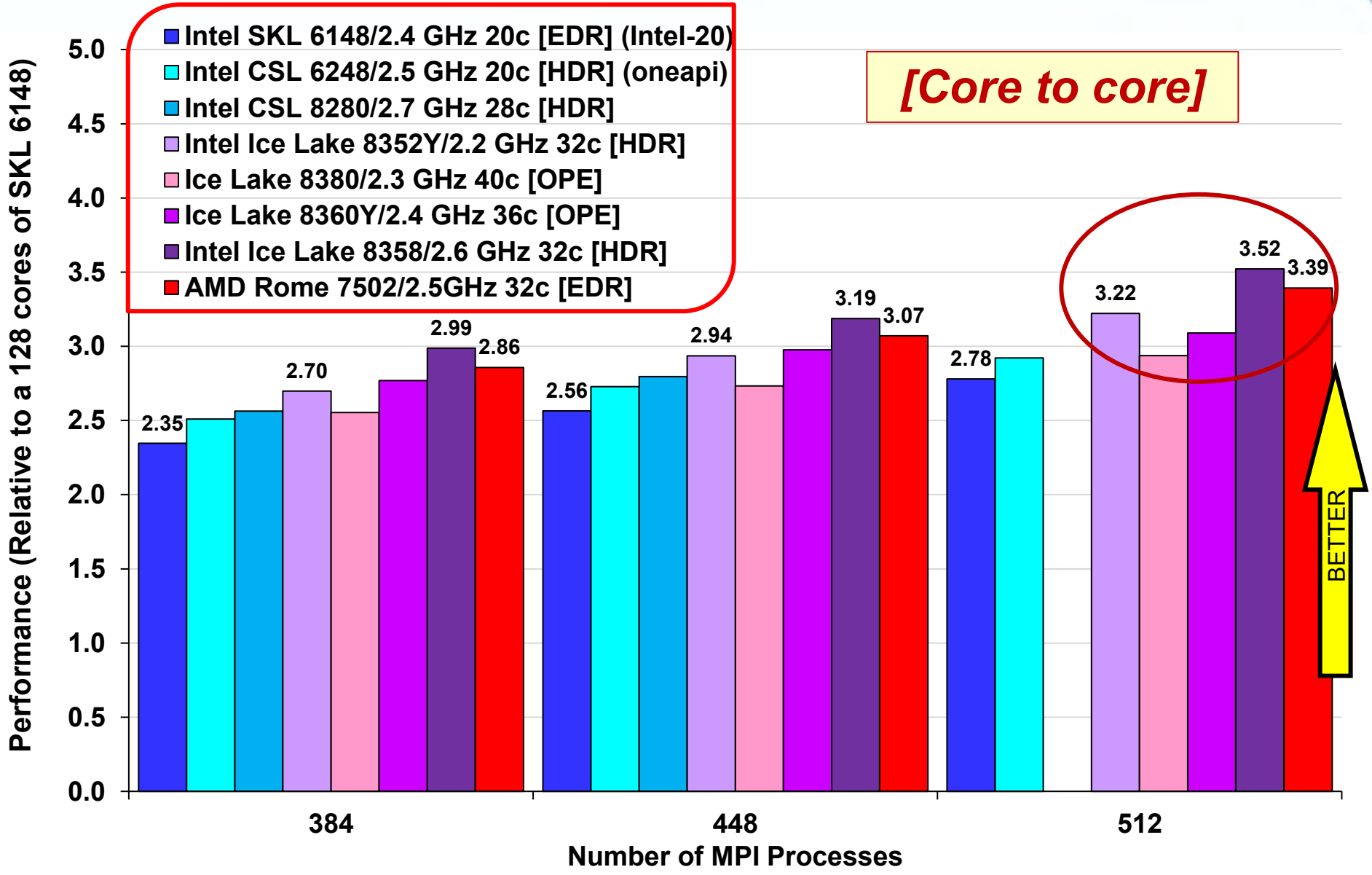
CPU Time Breakdown

FVCOM performance is dominated by memory access, with the per-core performance memory-bound. Little time is spent in vectorized instructions, suggesting significant opportunities for improving code performance.

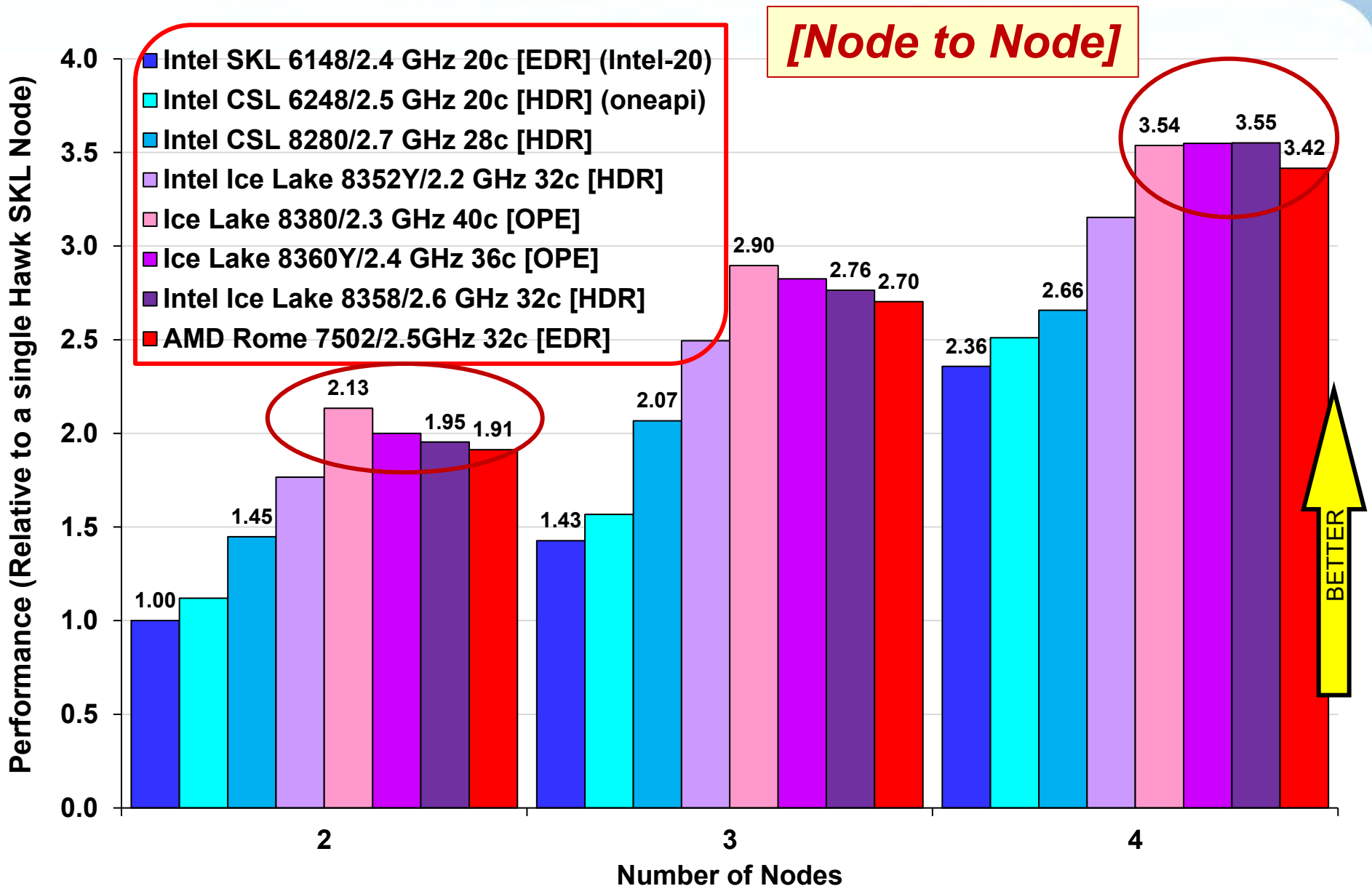
FVCOM – Performance Analysis [Core to Core]



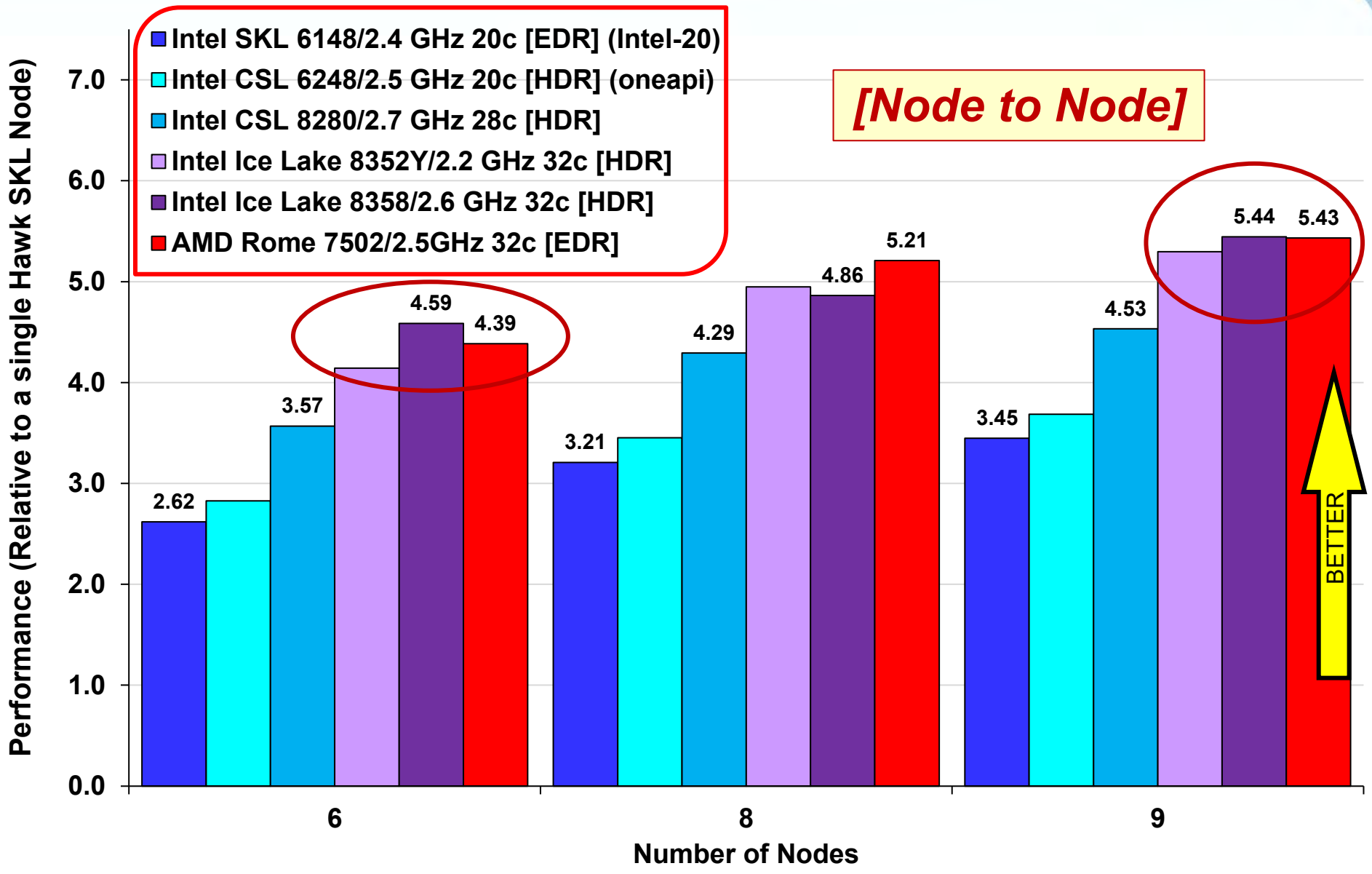
FVCOM – Performance Analysis [Core to Core]



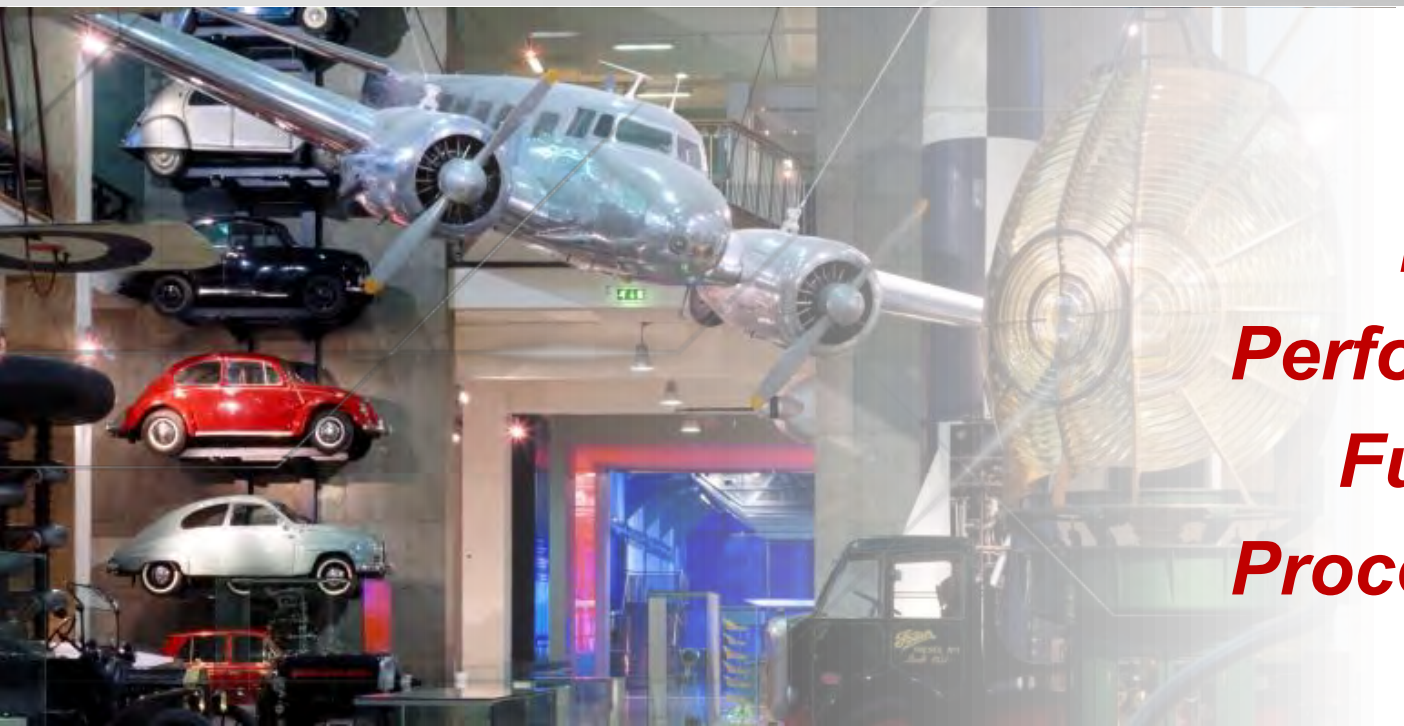
FVCOM – Performance Analysis [Node to Node]



FVCOM – Performance Analysis [Node to Node]

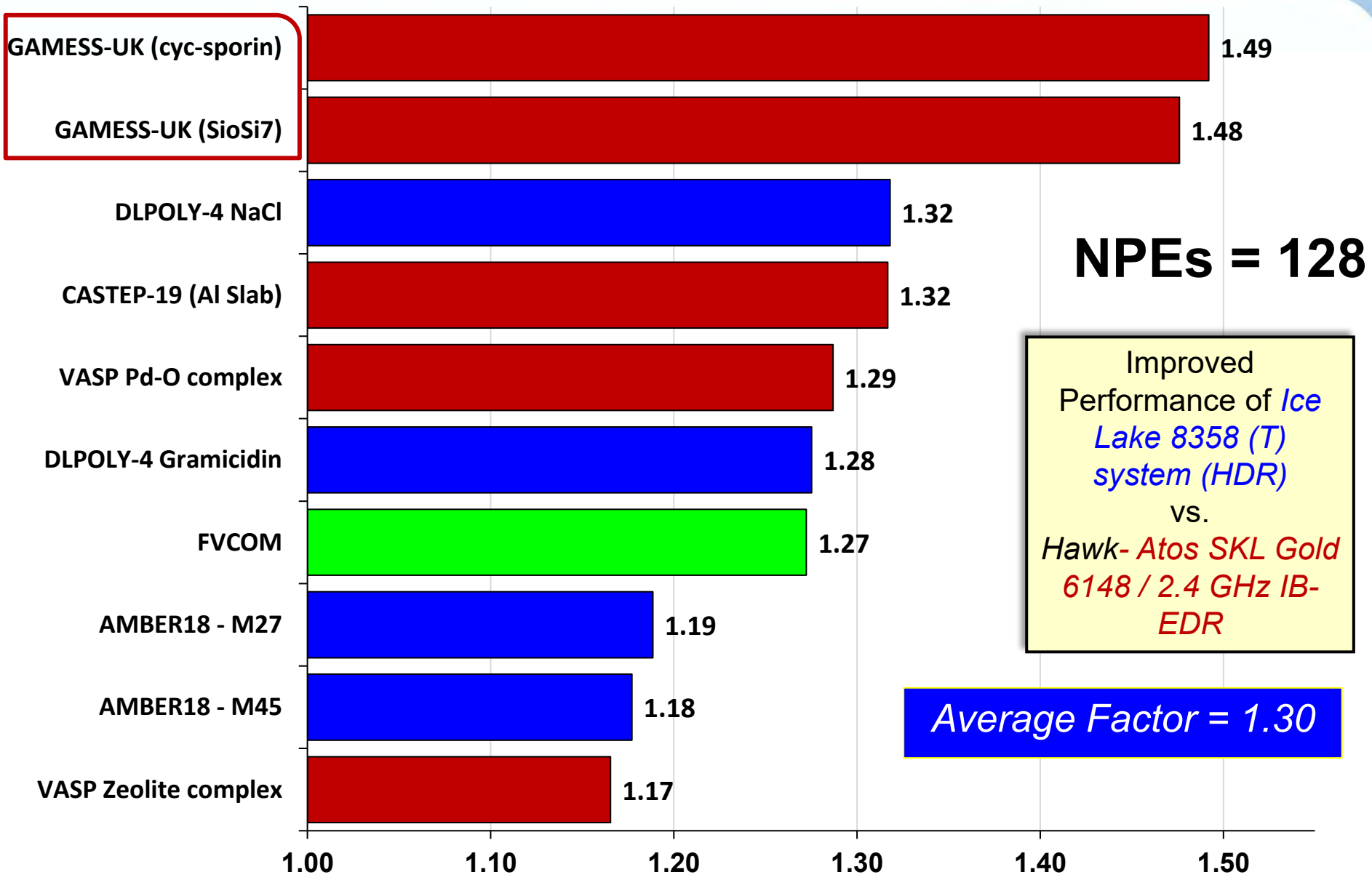


Performance of Computational Chemistry and Ocean Modelling Codes

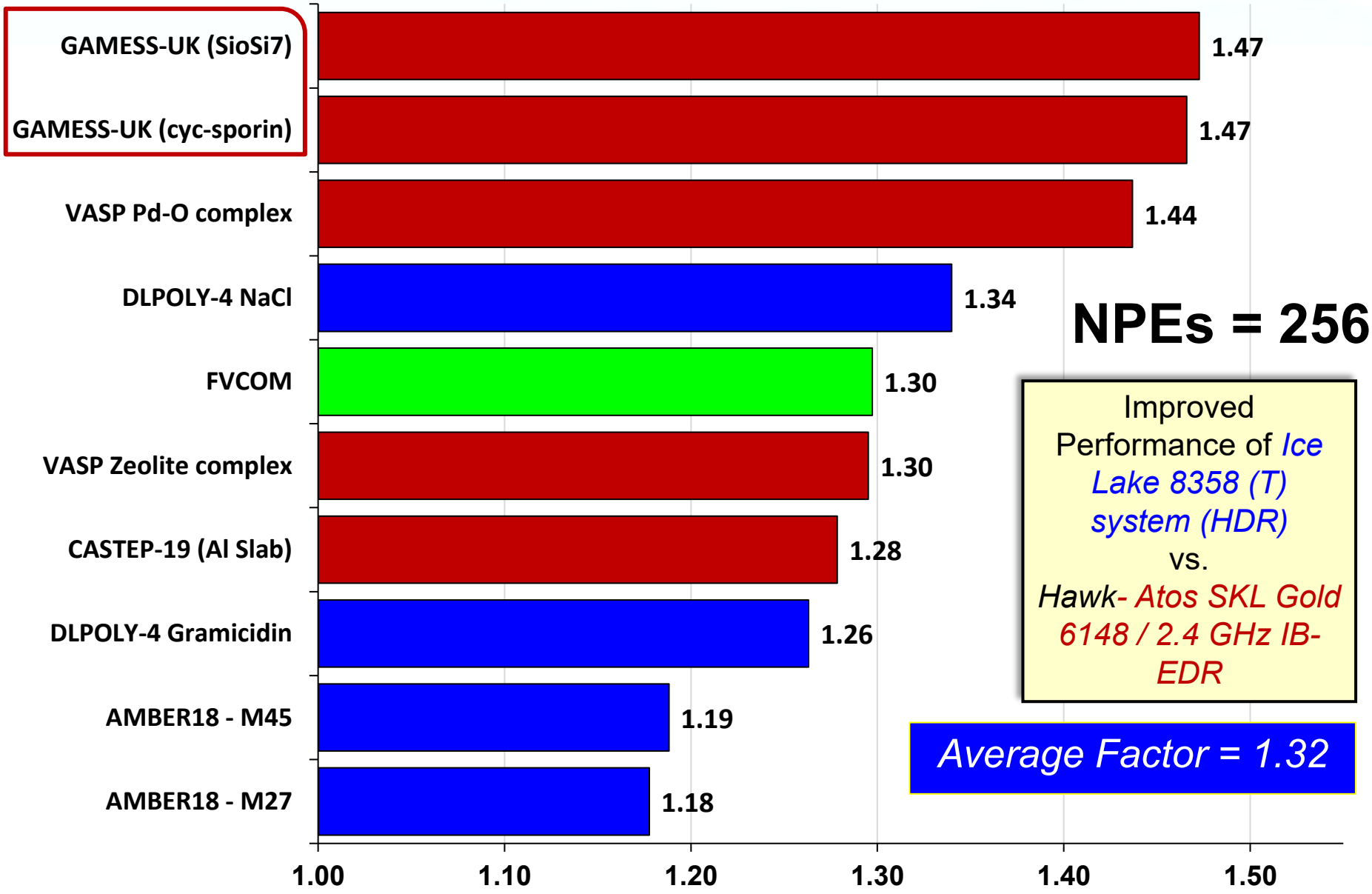


Relative Performance as a Function of Processor Family

Ice Lake 8358 2.6 GHz HDR vs. SKL 6148 2.4 GHz EDR



Ice Lake 8358 2.6 GHz HDR vs. SKL 6148 2.4 GHz EDR

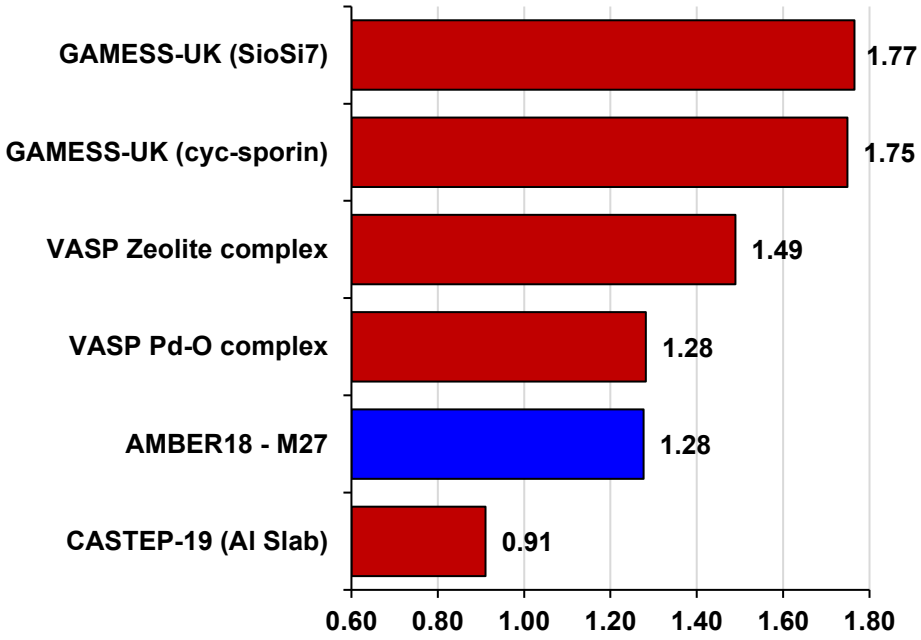


Performance of the AMD Milan 7573X 2.5 GHz HDR

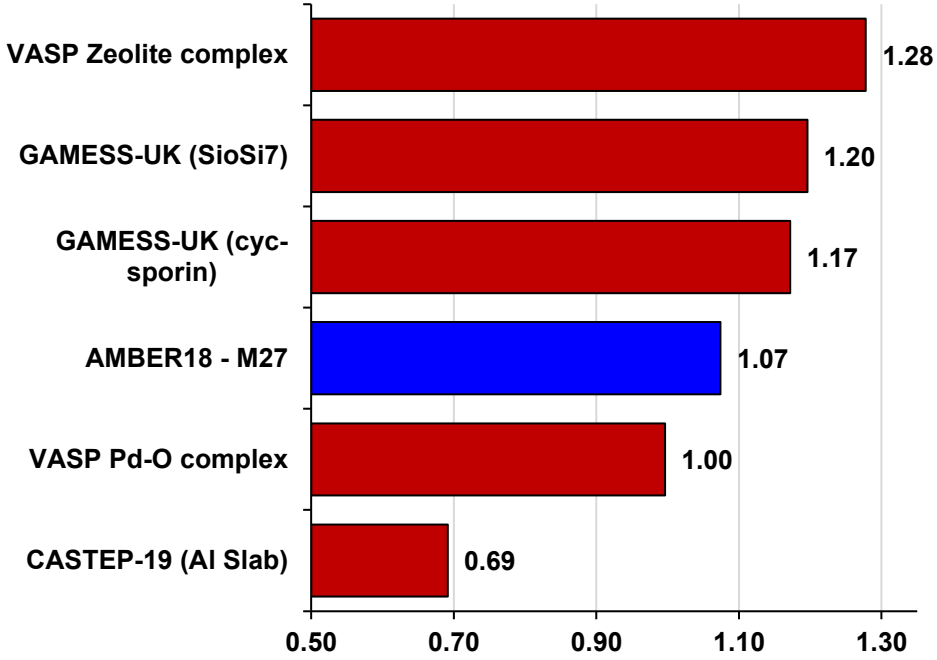
Performance of the AMD Milan 7573X / 2.8 GHz *(HDR)* vs. *Intel SKL Gold 6148 / 2.4 GHz (EDR)*

NPEs = 128

Performance of the AMD Milan 7573X / 2.8 GHz *(HDR)* vs. *Intel Ice Lake 8358 / 2.6 GHz system (HDR)*



Average Factor = 1.41



Average Factor = 1.07

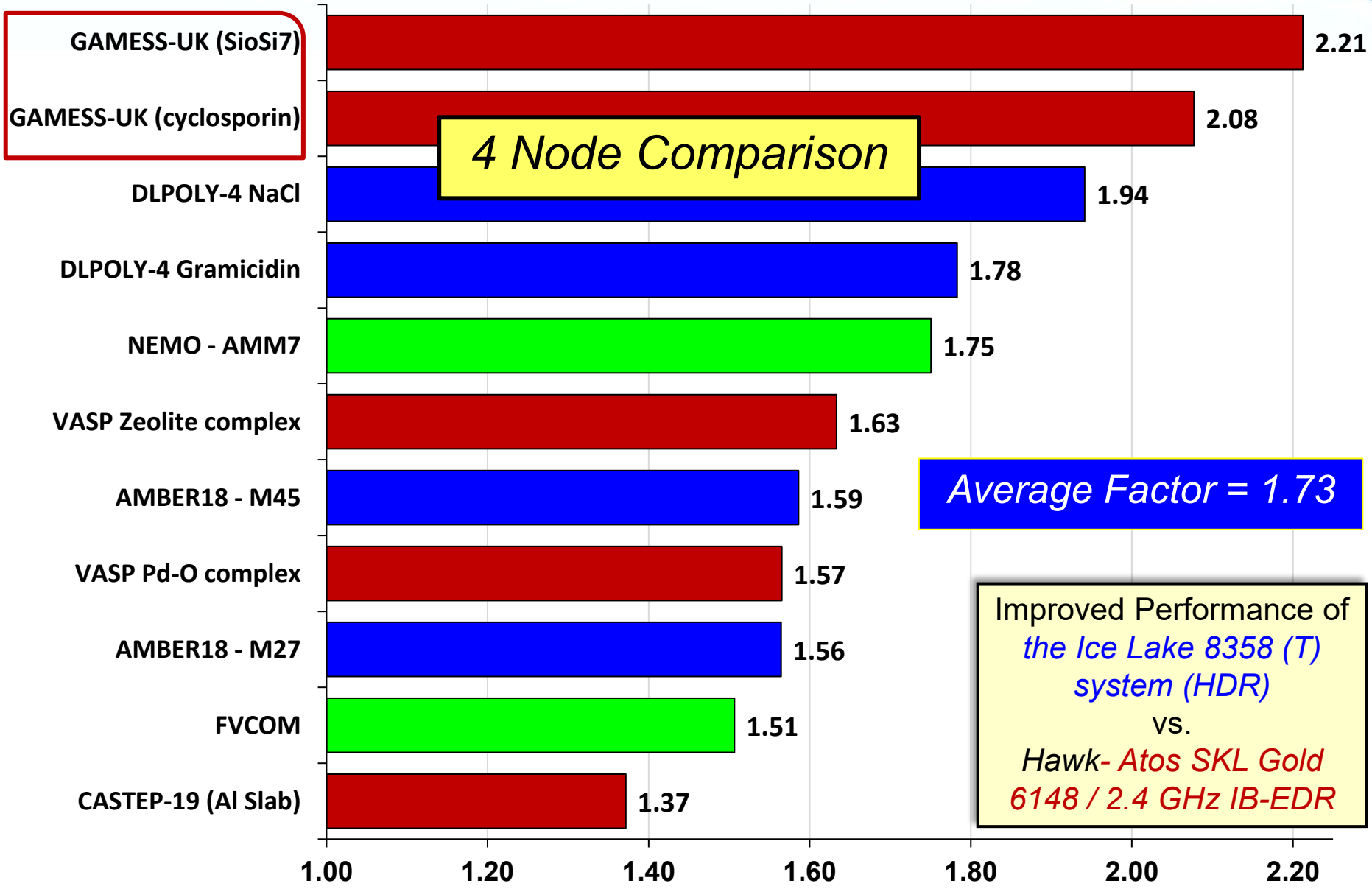
Summary – Core-to-Core Comparisons

- A ***Core-to-Core* comparison** suggests on average that the Intel Ice Lake 8358 2.6 GHz SKU outperforms all other Intel SKUs, although relative performance is sensitive to effective use of the AVX instructions.
- Low utilisation of AVX-512 leads to weaker performance of the SKL, CSL and Ice Lake CPUs and **better performance of the Milan-based clusters** e.g. DLPOLY, GAMESS-UK
- With significant AVX-512 utilisation, Ice Lake Lake systems outperform the AMD Milan systems in core-to-core comparisons e.g. Gromacs, notwithstanding the use of AVX2-256.
- **Exception** is the **AMD Milan 7573X / 2.8 GHz that outperforms the Intel Ice Lake** SKUs in a number of applications.
- With the possible exception of the Intel Ice Lake 8358, there is little to choose between the variety of Intel-based SKUs used in this study, the 36c 8360Y/2.4 GHz, the 38c 8368Q/2.6 GHz & 40c 8380/2.3 GHz.
- Baselined in part across **P100** and **V100** NVIDIA GPU performance.

Summary – Node-to-Node Comparisons

- Given superior core performance, a ***Node-to-Node comparison*** typical of the performance when running a workload shows the Ice Lake 8358 delivering **superior performance** compared to (i) the SKL Gold 6148 (64 cores vs. 40 cores) by a factor of between 1.4 – 2.2 across all applications.
- The **AMD Milan 7713, 7763 and 7773X (128 core nodes)** are the dominant systems given the “high” core counts. e,g,. GROMACS and GAMESS-UK.
- In contrast to the core-to-core comparisons, the higher core count Ice Lake systems – the **38c 8368Q and 40c 8380** – are now performing on a par with the 32c 8358.
- The 32c AMD Milan 7573X is ranked first in four of the 4-node application benchmarks.
- **Pricing** – remains of course a key issue, but lies outside the scope of this presentation.

Ice Lake 8358 2.6 GHz HDR vs. SKL 6148 2.4 GHz EDR



Acknowledgements

- **Joseph Stanfield and Joshua Weage,**, Dave Coughlin, Derek Rattansey for access to, and assistance with, the variety of AMD EPYC and Ice Lake SKUs at the Dell Benchmarking Centre.
- **Toby Smith, Ian Lloyd and Adam Roe** for access to and assistance with the CXL-AP and Ice Lake clusters at the Swindon Benchmarking Lab
- **Erwin James and John Swinburne** for implementing the NETCDF and XIOS-5 libraries on the Endeavour cluster for testing both the NEMO and FVCOM applications
- ***Okba Hamitou, Luis Cebamanos and Chrisophe Bertherlot*** and access to the SPARTAN and Ice Lake & Milan systems (Genji) at the Atos HPC, AI & QLM Benchmarking Centre
- **Jim Clark, Dale Partridge, Gary Holder and Jerry Blackford** at Plymouth Marine Laboratory for discussions on NEMO & FVCOM performance.

Summary

Focus here on systems featuring **processors from AMD** (EPYC Milan SKUs) and **Intel** (Ice Lake SKUs) with IB and Cornelis Networks interconnects.

- Baseline cluster: the Skylake (SKL) **Gold 6148/2.4 GHz** and **AMD EPYC Rome 7502 2.5Gz** cluster – "Hawk" – at Cardiff University.
- **Five** Intel Xeon Ice Lake clusters, the 32-core Platinum **8358** (2.6 GHz) and **8352Y** (2.2 GHz), the 40-core **8380** (2.3 GHz), 38-core **8368Q** (2.6 GHz), 36-core **8360Y** (2.4GHz) plus other Cascade Lake & Cascade Lake-AP systems.
- **Four** AMD EPYC Milan clusters featuring the 64-core **7713** (2.0 GHz) and **7773X** (2.2 GHz) and the 32-core **7543** (2.8 GHz) and **7573X** (2.8 GHz).
- Consider performance of both synthetic and **end-user applications**. Latter include molecular simulation (**DL_POLY, AMBER**), materials modelling (**CASTEP, VASP**), & electronic structure (**GAMESS-UK**), plus representative ocean modelling codes including **NEMO** and **FVCOM**.
- Scalability analysis by **processing elements (cores)** and by **nodes** (ARM Performance Reports). Baselined against **P100** & **V100** NVIDIA GPUs.

Any Questions?

Martyn Guest *029-208-79319*

Jose Munoz *029-208-70626*

Thomas Green *029-208-79269*

Keynote Presentation

Professor Michèle Weiland (EPCC, The University of Edinburgh)

Net Zero HPC - noble dream or inevitable goal?

Abstract: In the face of the climate emergency, the term 'Net Zero' has sprung up in every context of daily life, and HPC is no exception. But how far away are we from achieving Net Zero? Where is the community doing well already, where are we falling short, and why? How hard and how quickly should we push for Net Zero HPC - what sacrifices (if any) are acceptable? In this talk, I will explore what the concept of 'Net Zero' might mean for users, operators and hosts of HPC systems going forward, and show examples of research and activities that are trying to push HPC in the direction of Net Zero and sustainability.

Bio: Prof Michèle Weiland is the Director of Research and Met Office Joint Chair at EPCC, the supercomputing centre at the University of Edinburgh. She leads Edinburgh's involvement in the Met Office Academic Partnership, as well as the technical work in the UKRI-funded ASiMoV Strategic Prosperity Partnership with Rolls-Royce, building and performing highly complex multi-physics simulations of aircraft engines. She led the EU H2020 project NEXTGenIO which, in collaboration with Intel and Fujitsu, developed a new HPC platform using non-volatile memory to accelerate I/O performance, and the ExCALIBUR ELEMENT project that investigated meshing for and at Exascale. She also led the EU FP7 project Adept, which investigated methods for energy and power efficiency measurements on parallel hardware. Michèle is a member of the EPSRC Strategic Advisory Team for e-Infrastructure.

NET ZERO HPC - NOBLE DREAM OR INEVITABLE GOAL?

Prof Michèle Weiland
Director of Research & Met Office Joint Chair

epcc

IPCC's definition of Net Zero

Net zero CO_2 emissions

Net zero *carbon dioxide (CO_2)* emissions are achieved when *anthropogenic* CO_2 emissions are balanced globally by anthropogenic CO_2 removals over a specified period. Net zero CO_2 emissions are also referred to as carbon neutrality. See also *Net zero emissions* and *Net negative emissions*.

IPCC, 2018: Annex I: Glossary [Matthews, J.B.R. (ed.)]. In: *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 541-562, doi:10.1017/9781009157940.008.

epcc

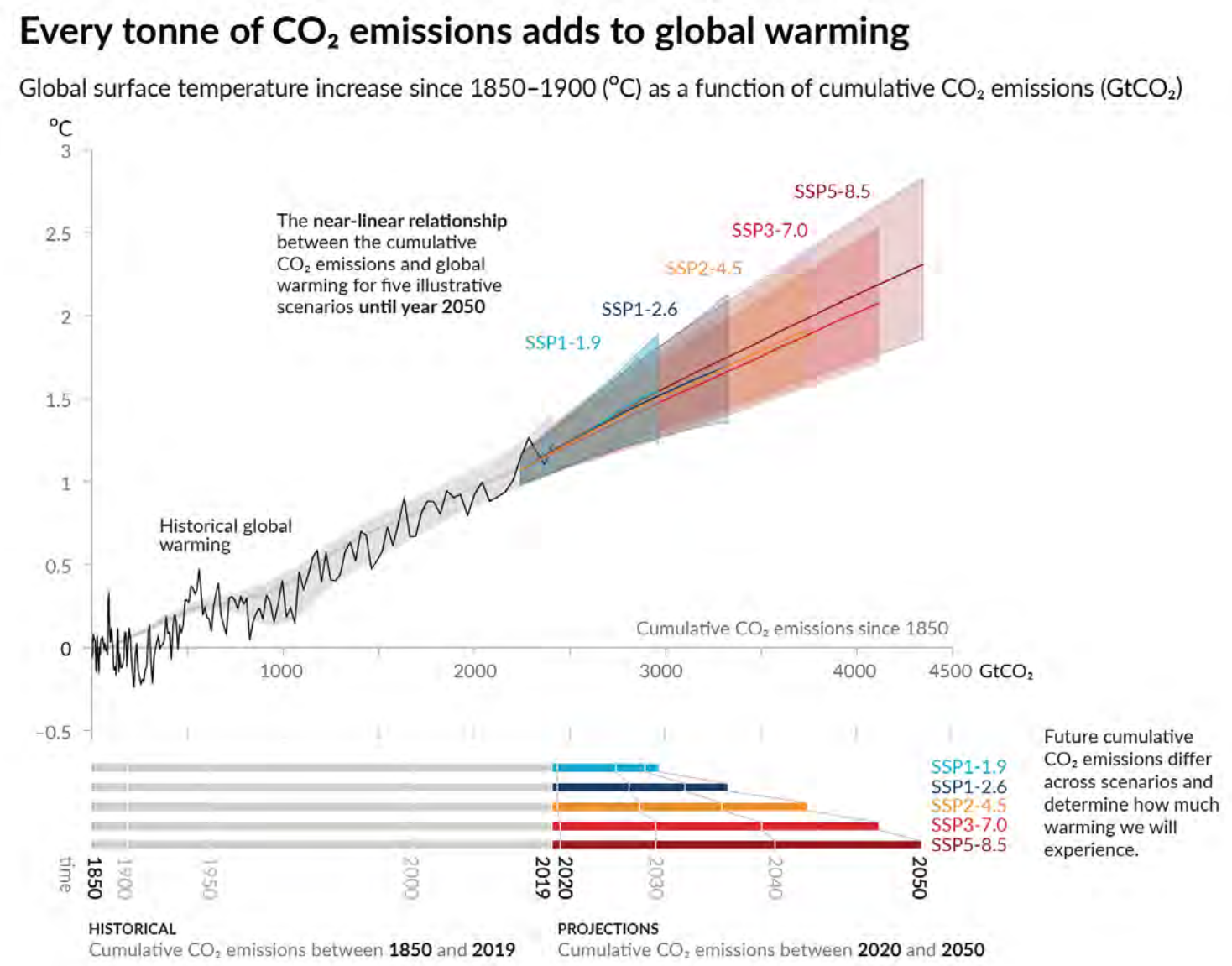
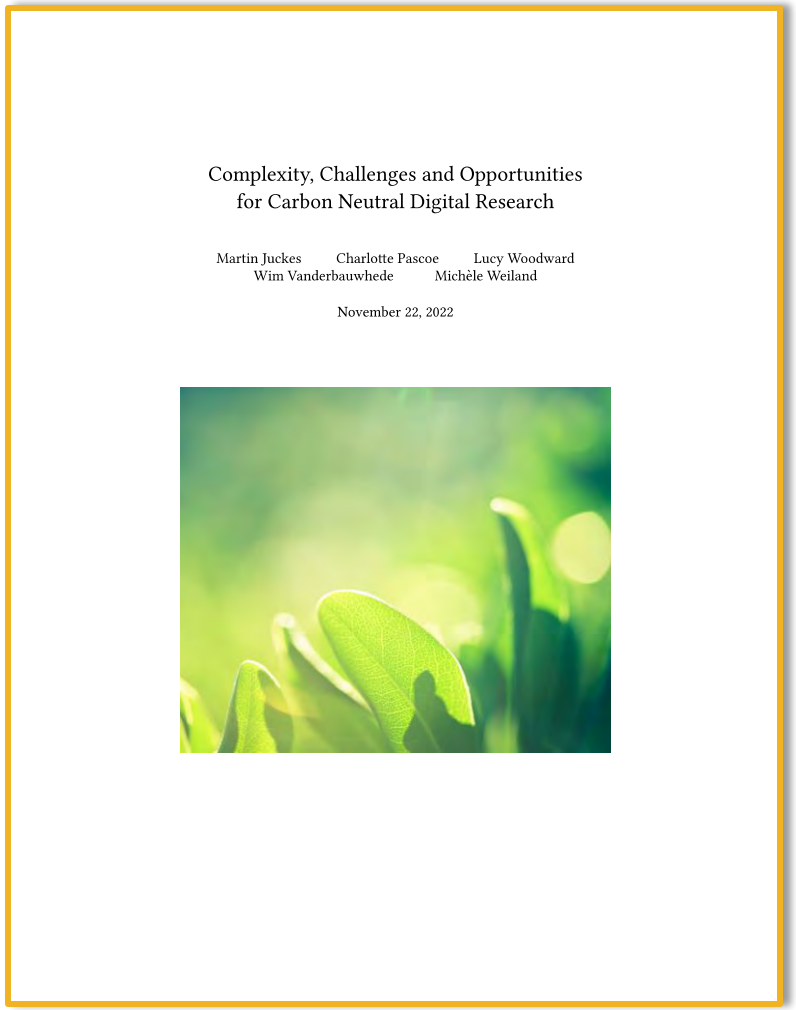


Figure SPM.10 in IPCC, 2021: Summary for Policymakers. In: ***Climate Change 2021: The Physical Science Basis***. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 3–32, doi: 10.1017/9781009157896.001

Net Zero Digital Research Infrastructure


Complexity, Challenges and Opportunities
for Carbon Neutral Digital Research

Martin Juckes Charlotte Pascoe Lucy Woodward
Wim Vanderbauwhede Michèle Weiland

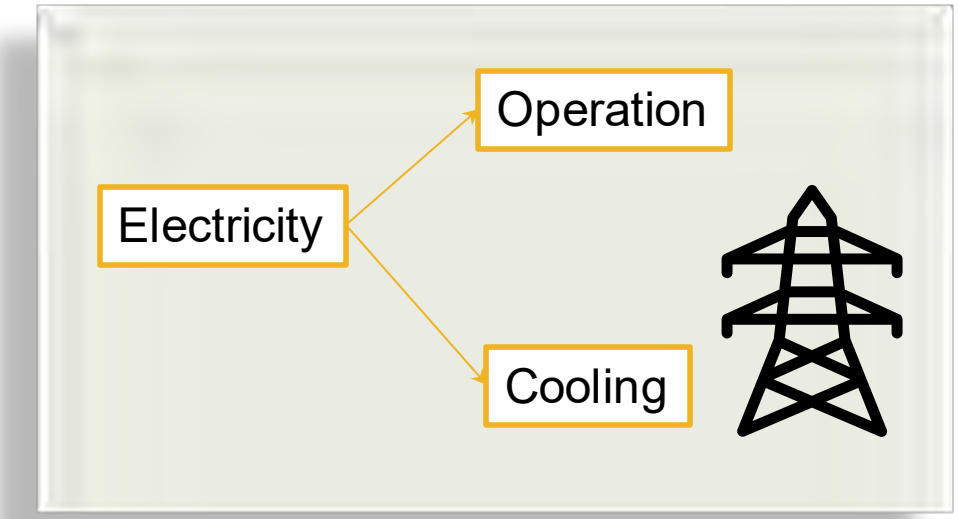
November 22, 2022

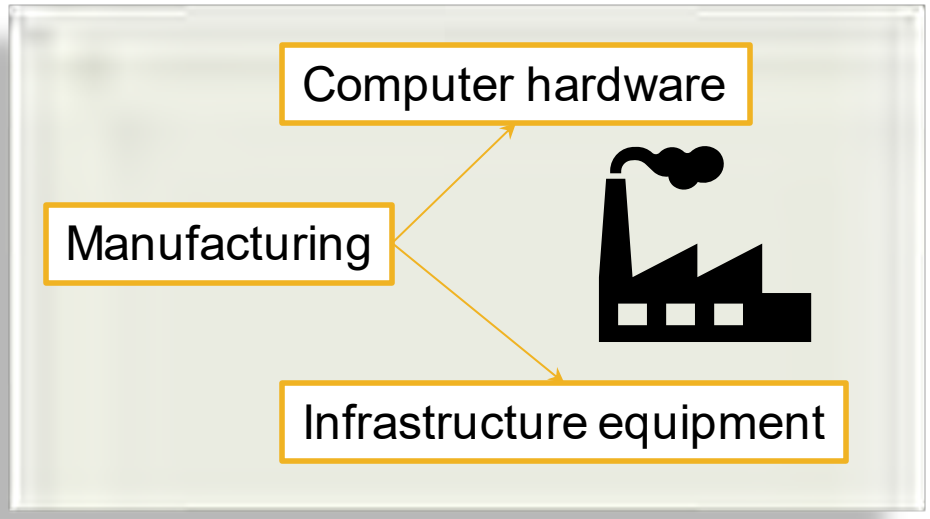
> This report is addressed to UKRI Digital Research Infrastructure (DRI) stakeholders in UK Research and Innovation (UKRI), its constituent Research Councils and in the Universities and other institutions which own and operate many digital research facilities which are majority funded by UKRI.
>
> The aim of the UKRI Net Zero Digital Research Infrastructure Scoping Project (hereafter "the project") is to report on evidence and make recommendations for a roadmap to a Net Zero UKRI DRI by 2040 or sooner. The project, via this interim report and future reports, provides recommendations for reducing and avoiding carbon emissions. It also reviews options for dealing with unavoided emissions through carbon capture, biochar and offsetting.

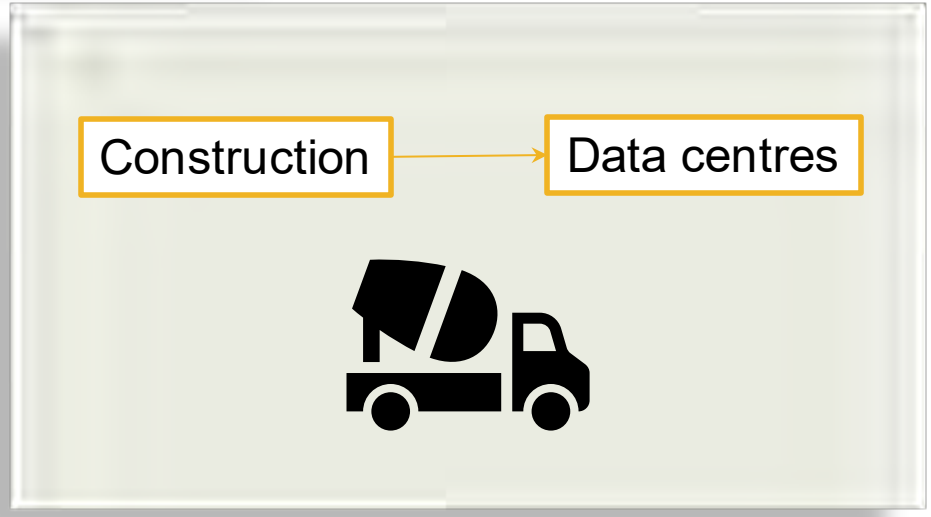
https://net-zero-dri.ceda.ac.uk

epcc

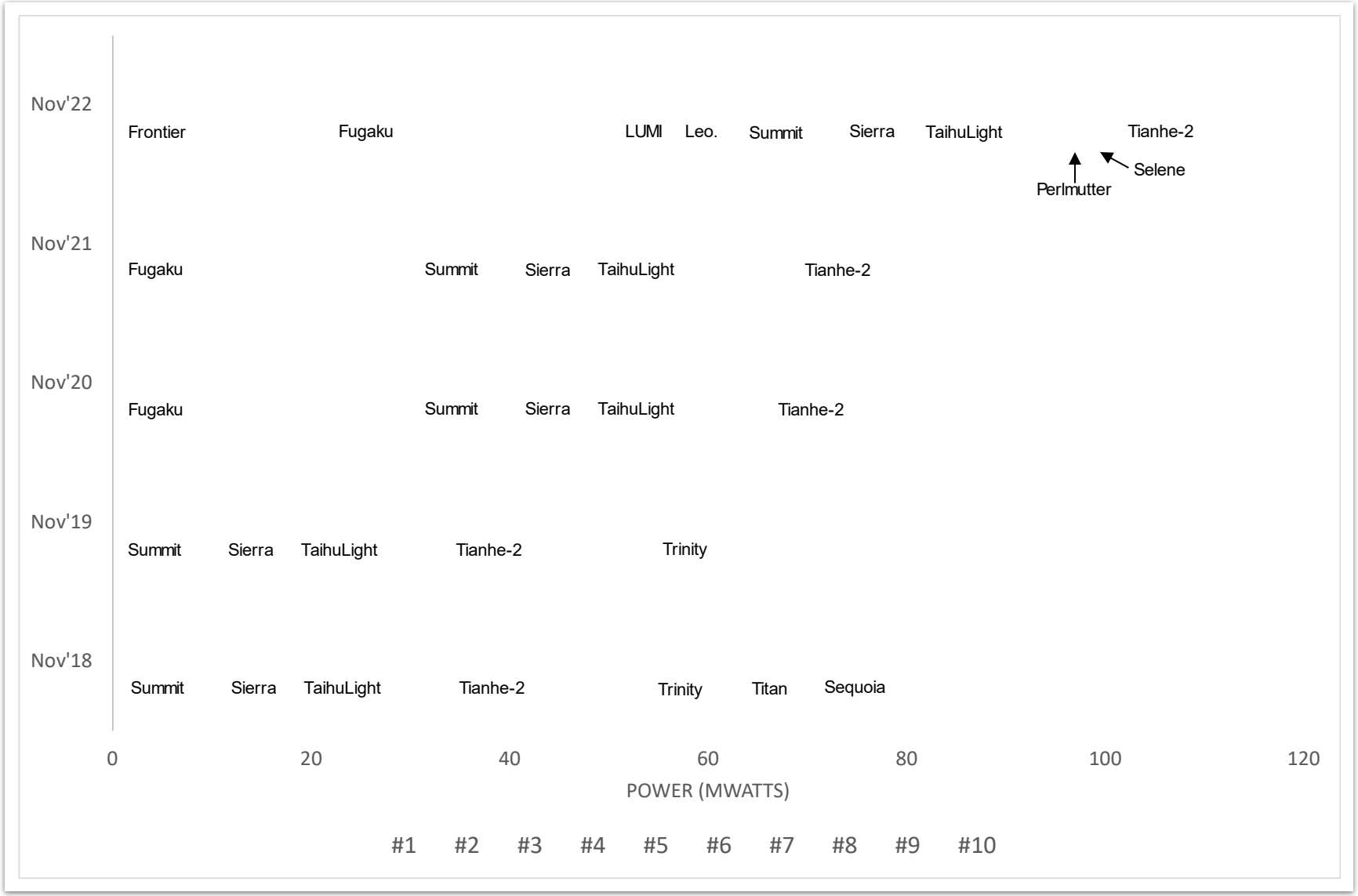
HPC & CO^2 emissions



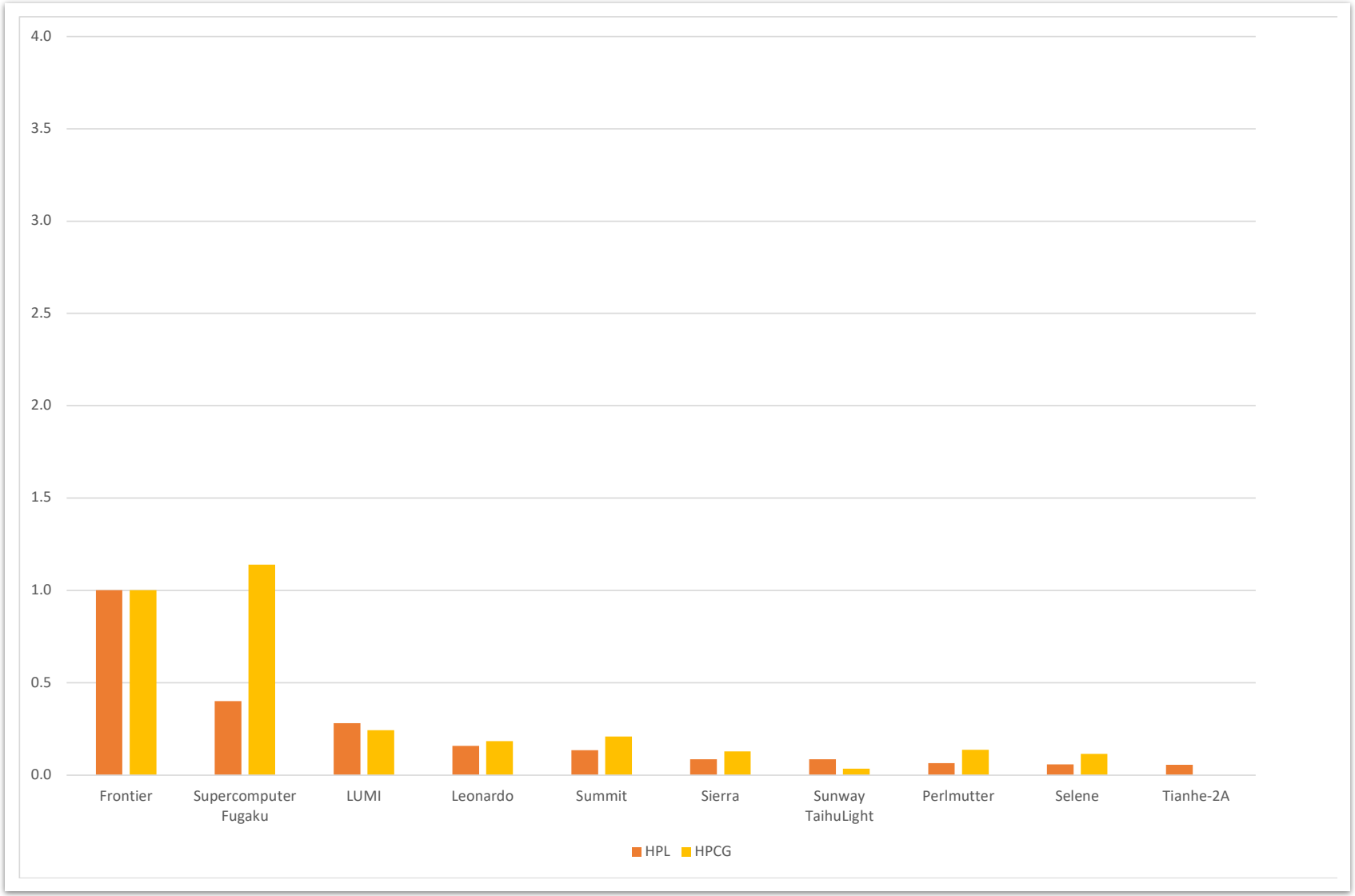




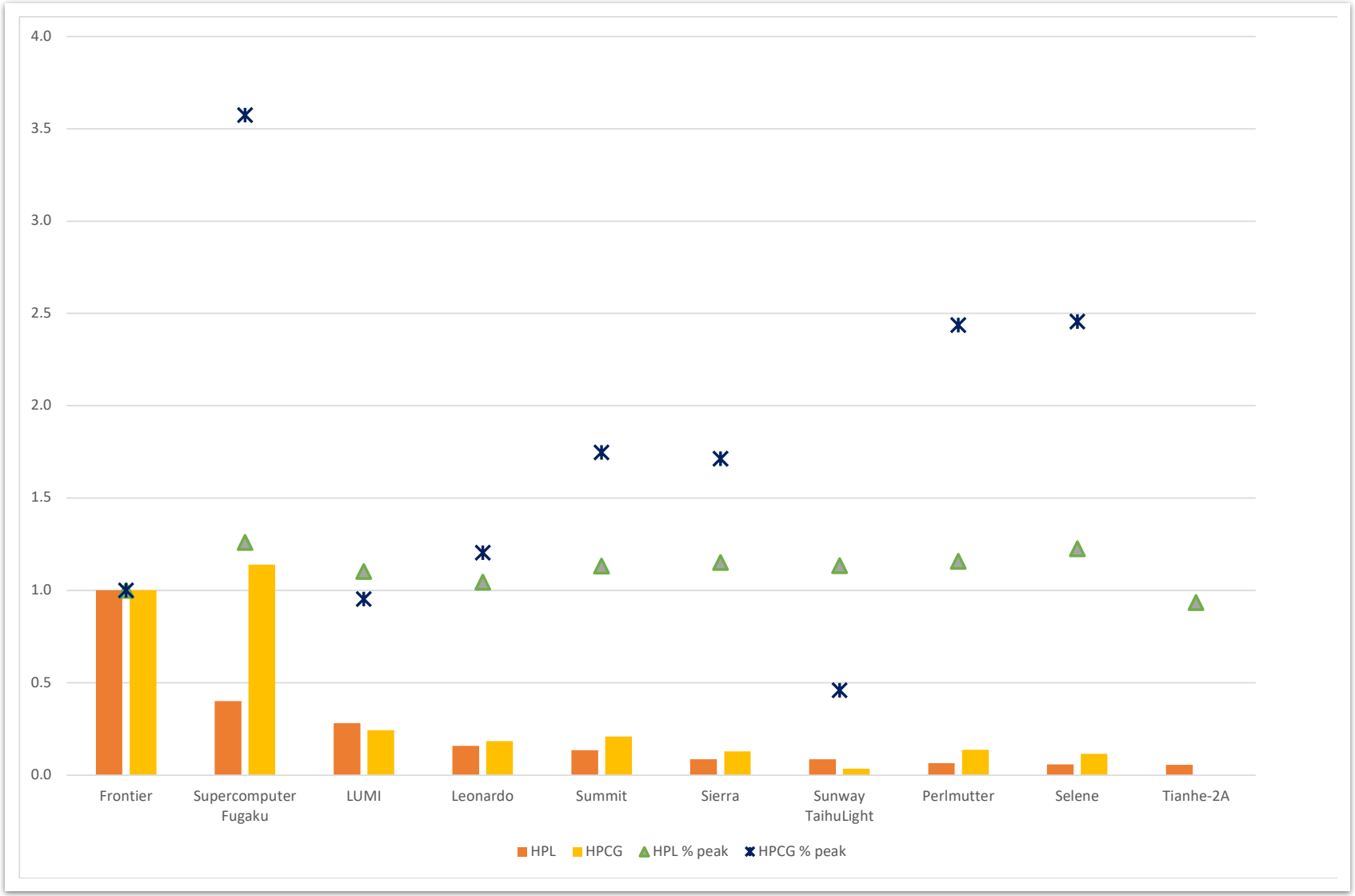
- Power reported for Top10 systems over the past 5 years
- #1 system ranges from 9.7MW to 30MW
- #10 system in Nov'22 has 3rd highest power draw
- Only in 2022 all Top10 systems submitted power
- Accumulative power for Top10 in 2022: 120MW



- HPL & HPCG performance normalized to Frontier
- >1 means “better”

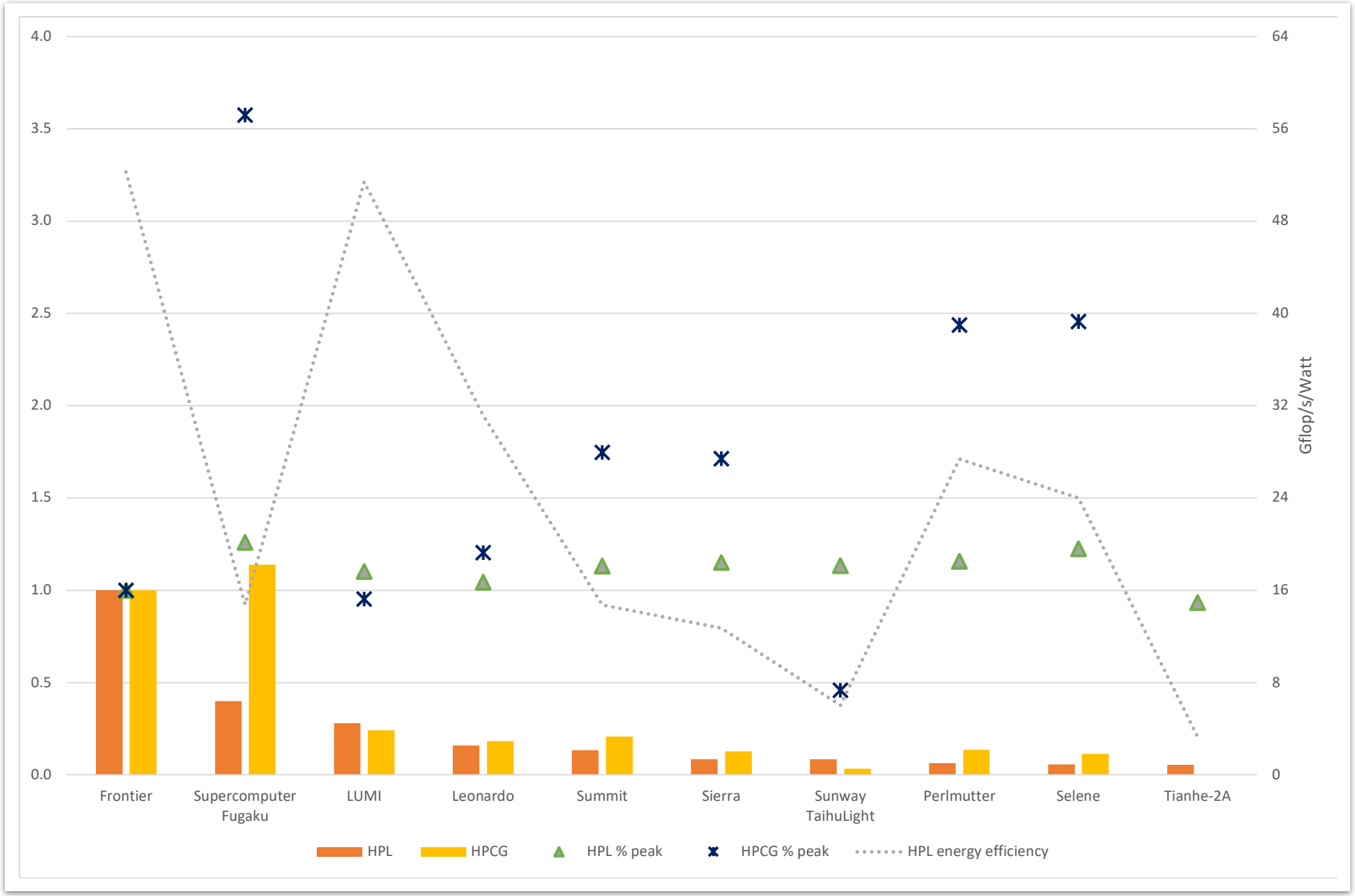


- HPL & HPCG performance normalized to Frontier
- >1 means “better”
- HPL % of peak close to Frontier for all systems
  - 70-75% on average



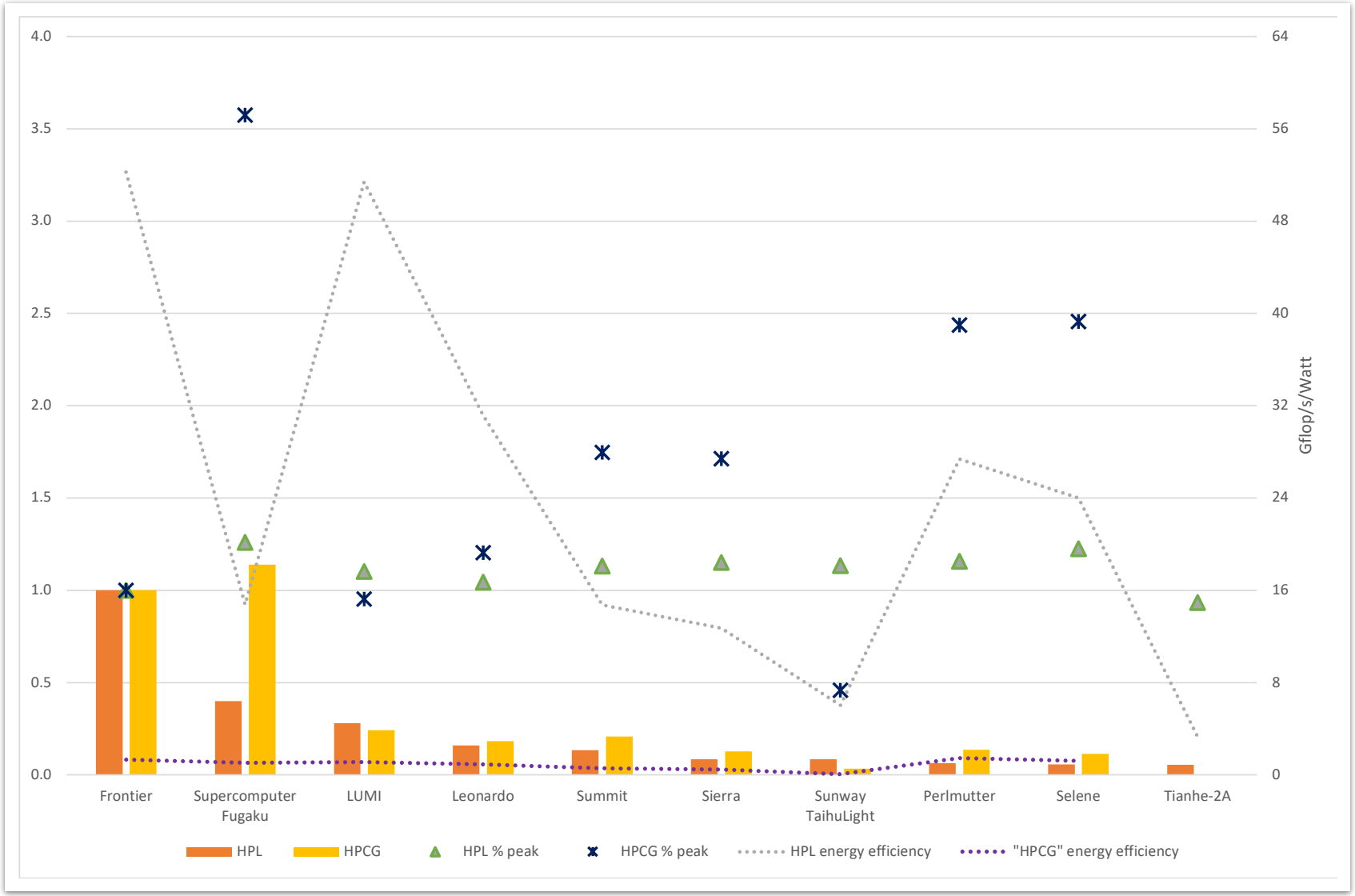
epcc

- HPL & HPCG performance normalized to Frontier
- >1 means “better”
- HPL % of peak close to Frontier for all systems
  - 70-75% on average
- Green500 list uses HPL power to evaluate efficiency



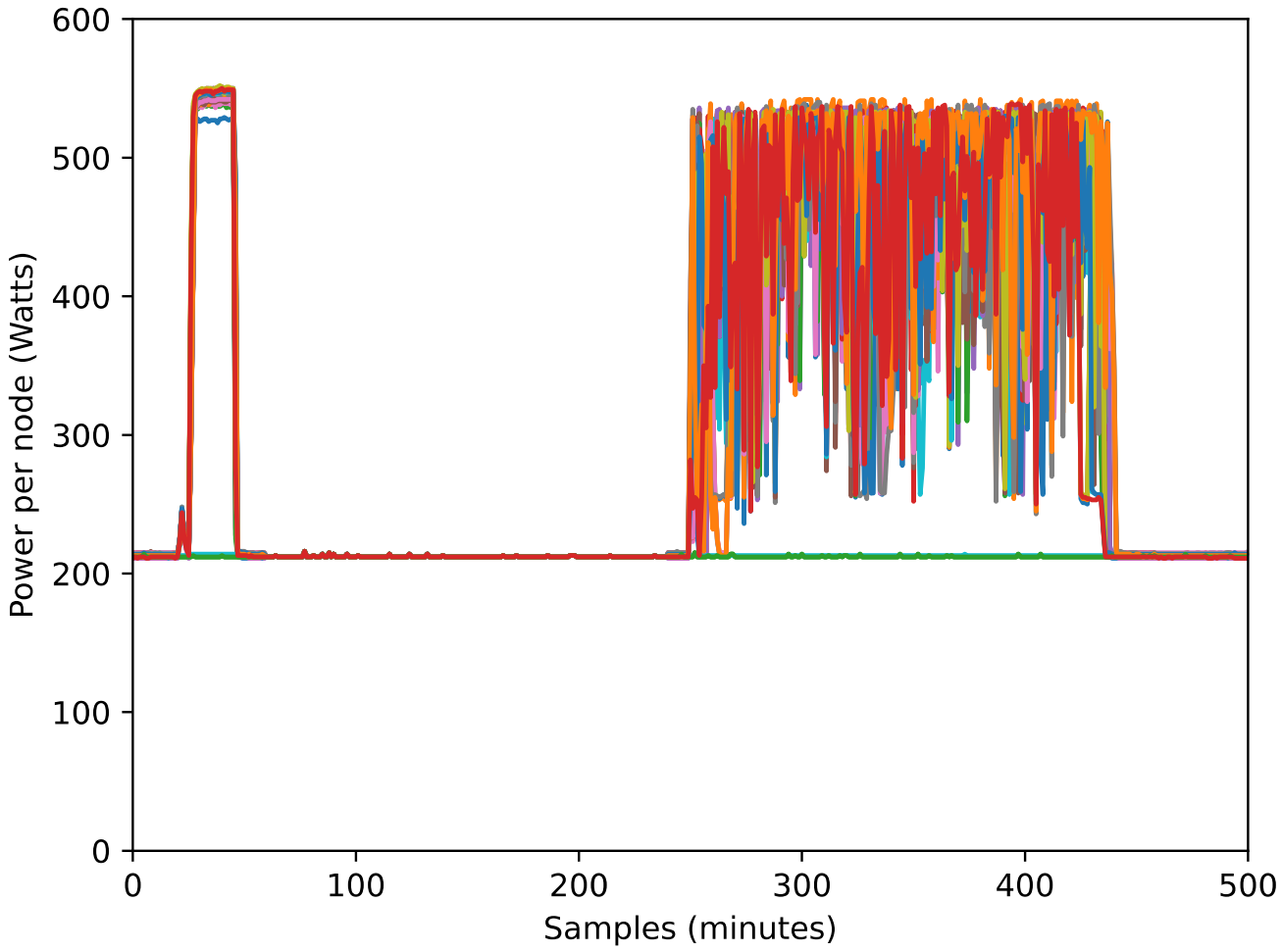
epcc

- HPL & HPCG performance normalized to Frontier
- >1 means “better”
- HPL % of peak close to Frontier for all systems
  - 70-75% on average
- Green500 list uses HPL power to evaluate efficiency
- “HPCG” energy efficiency assumes 50% of HPL power



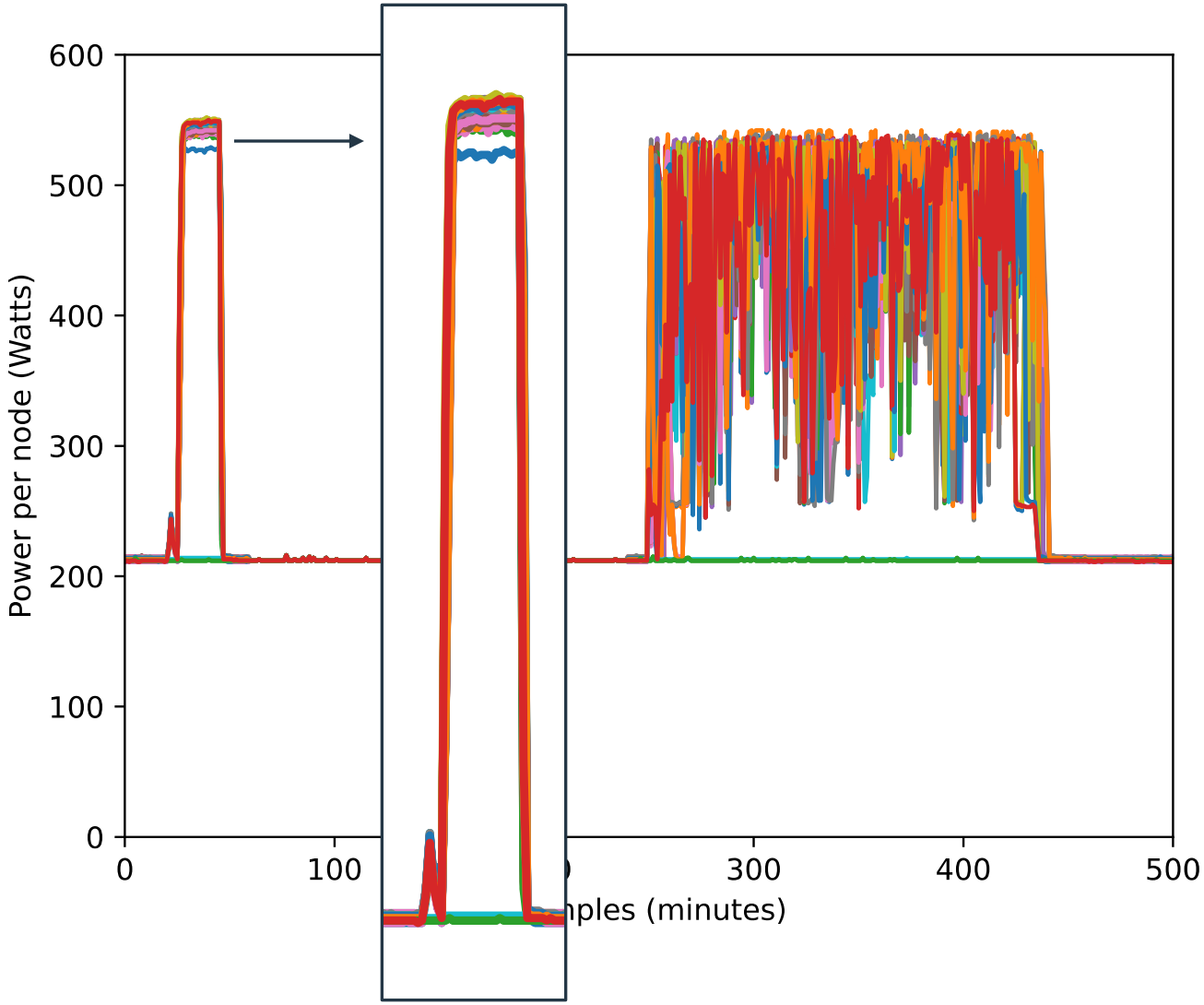
epcc

Impact of inefficient I/O configuration



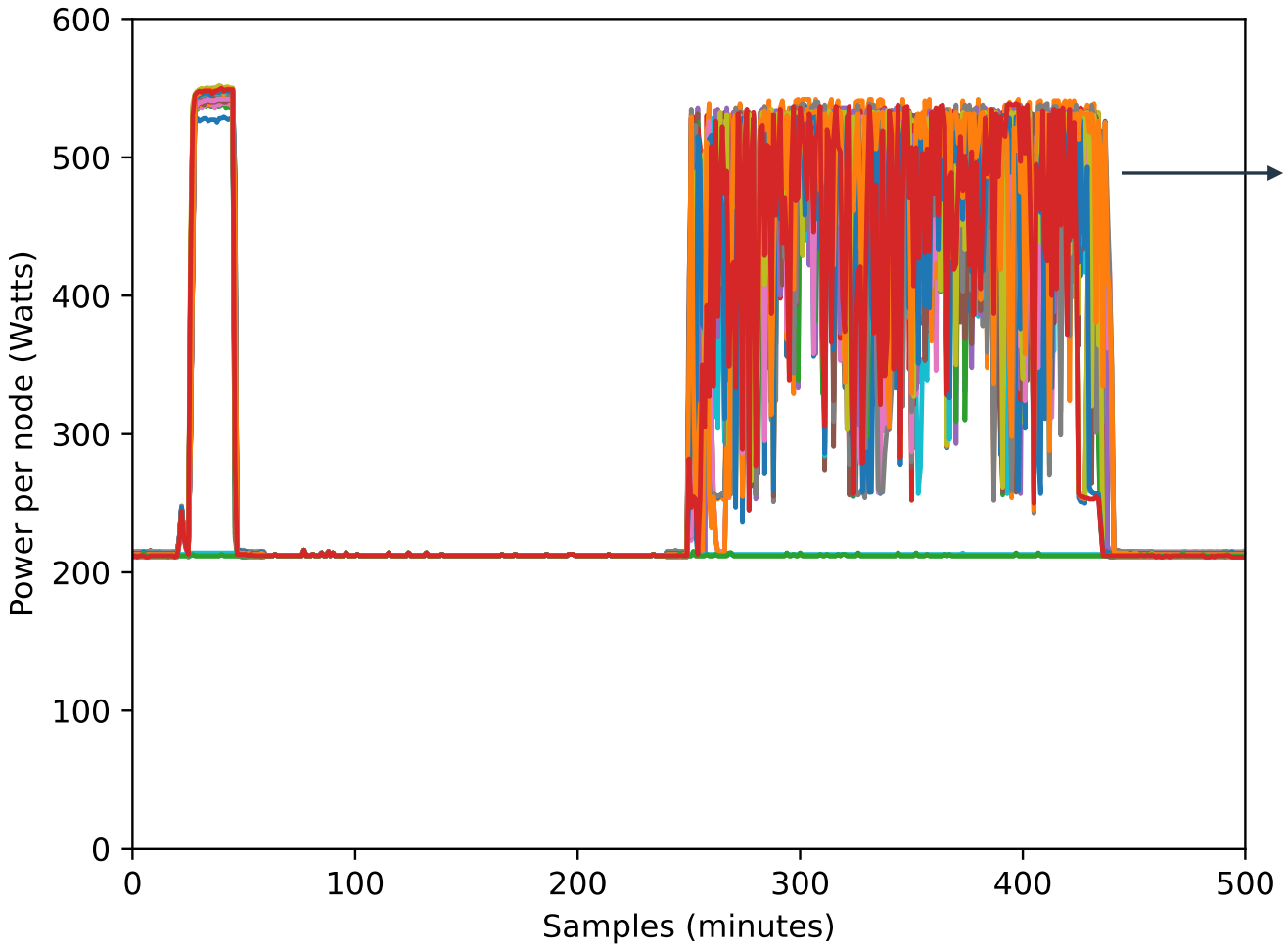
- Node-level power measurement
  - Each line represents power draw for 1 node
  - Full system, 34 nodes in total
  - Idle power draw: 213W
- Two identical aerodynamics simulations with OpenFOAM using 32 nodes
  - On the left: **no** I/O
  - On the right: **excessive** I/O

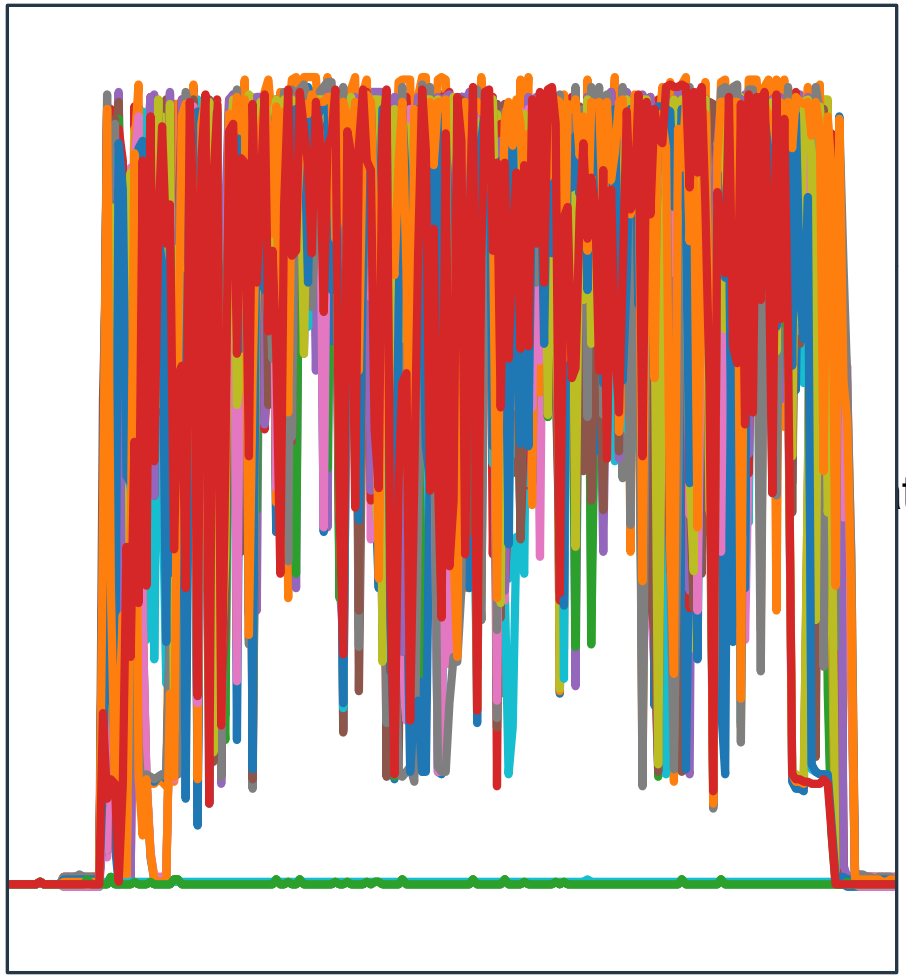
Impact of inefficient I/O configuration



- Node-level power measurement
  - Each line represents power draw for 1 node
  - Full system, 34 nodes in total
  - Idle power draw: 213W
- Two identical aerodynamics simulations with OpenFOAM using 32 nodes
  - On the left: **no** I/O
  - On the right: **excessive** I/O

Impact of inefficient I/O configuration

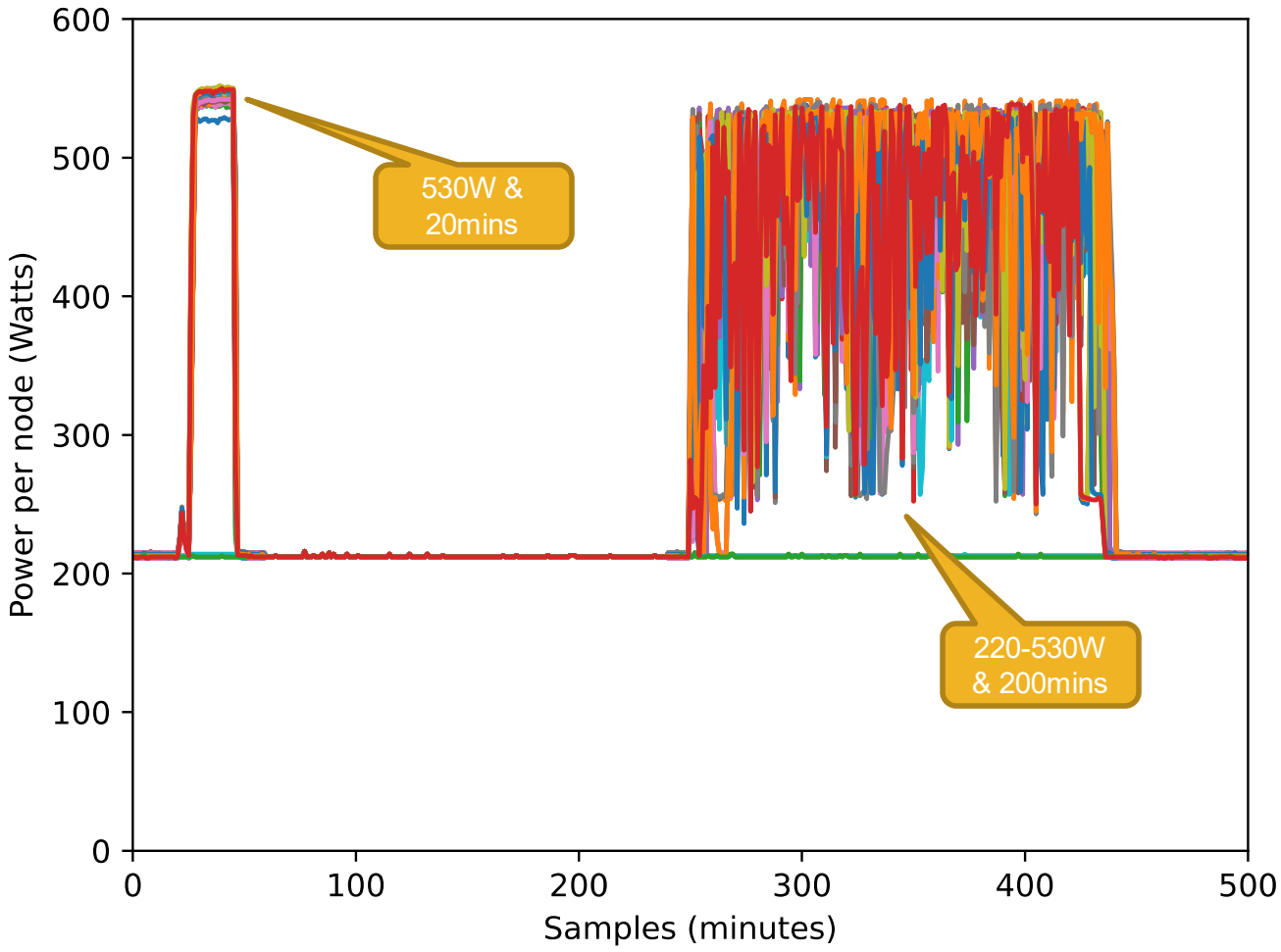




1 node

tions

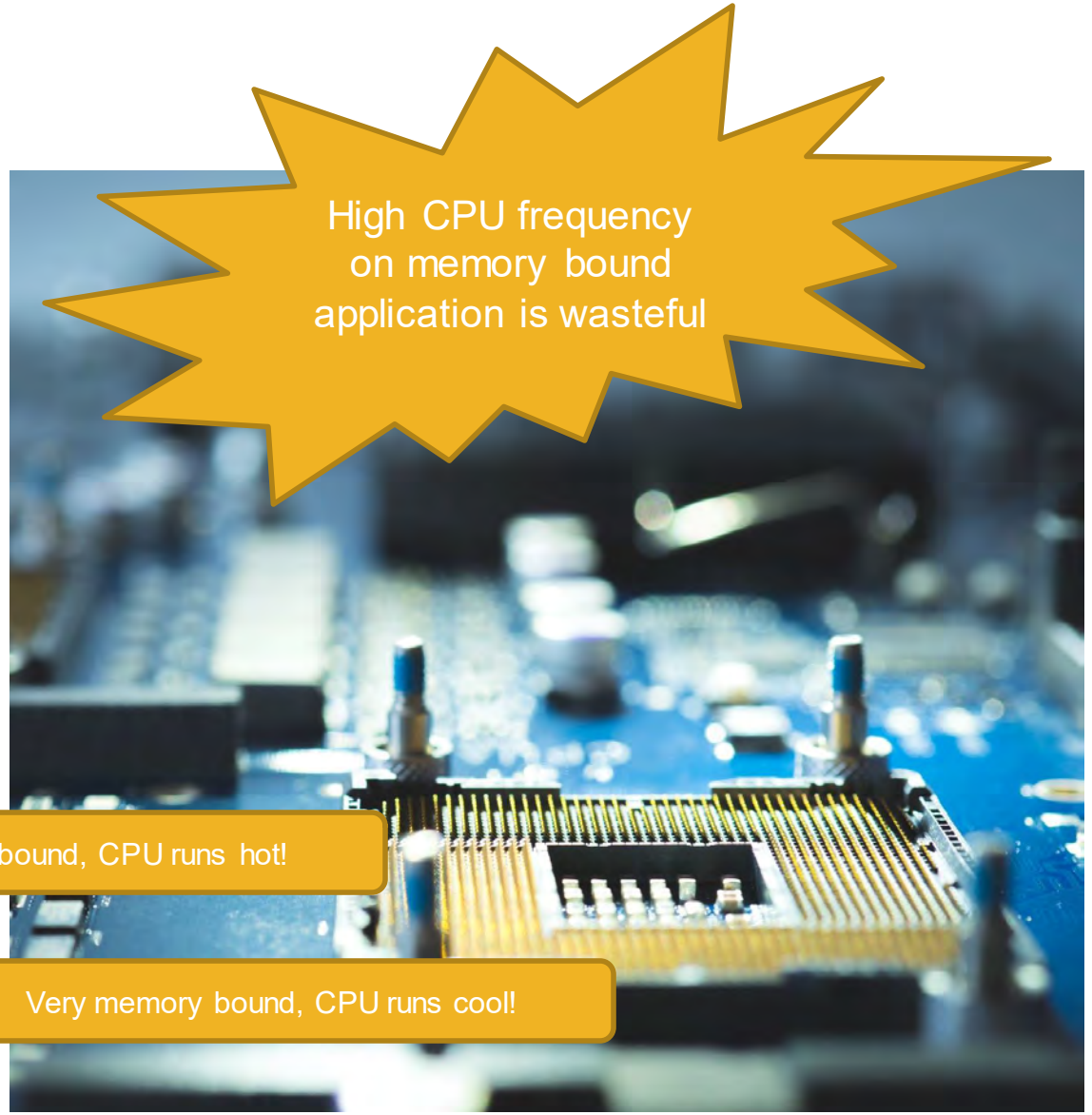
Impact of inefficient I/O configuration



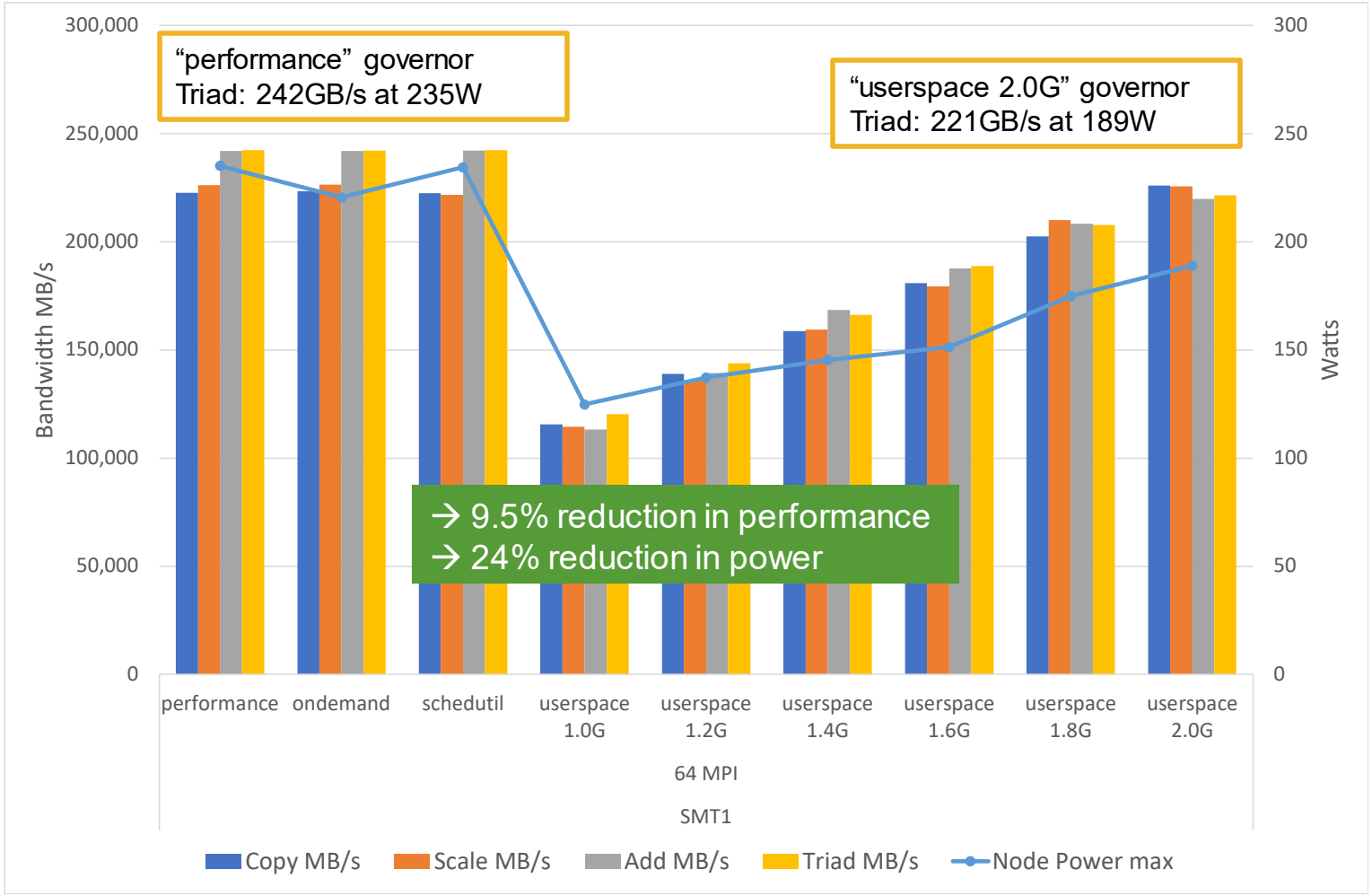
- Node-level power measurement
  - Each line represents power draw for 1 node
  - Full system, 34 nodes in total
  - Idle power draw: 213W
- Two identical aerodynamics simulations with OpenFOAM using 32 nodes
  - On the left: **no** I/O
  - On the right: **excessive** I/O
- Excessive I/O means network contention & frequent stalling

CPU clock frequency

- Frequency ∝ power draw
- Often taken care of by Dynamic Voltage & Frequency Scaling (DVFS)
  - Heat and power management
  - Influenced by governor
- Might not do what you expect…
- Simple test on Cirrus (Intel Broadwell)
  - FIRESTARTER → 2.2GHz
  - HPCG → 2.4GHz
  - OpenFOAM (simpleFoam) → 2.6GHz



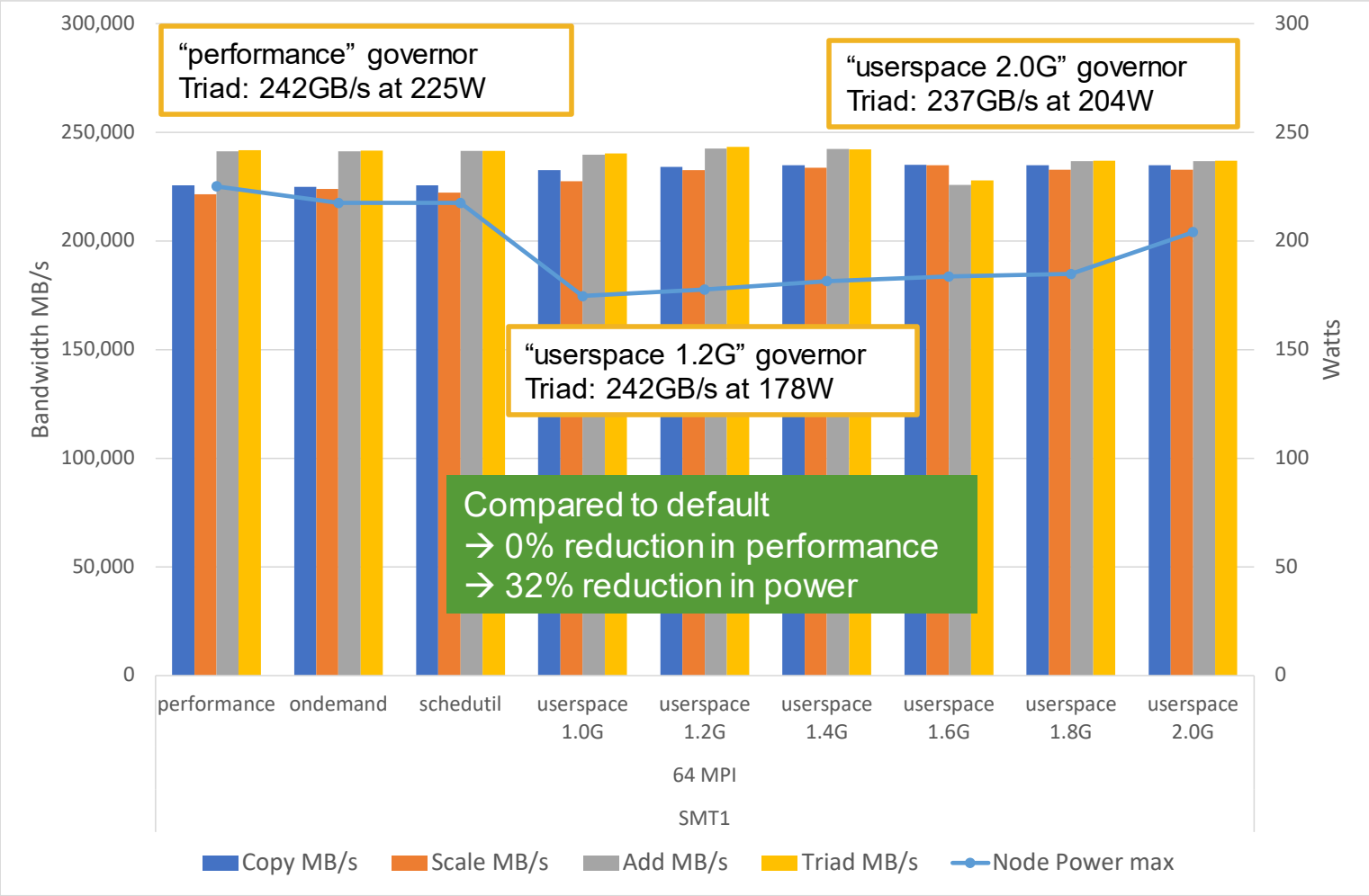
Frequency scaling & STREAM



- Marvell ThunderX2 Arm-based test system
- Influencing CPU frequency using different governors
- CPU can boost up to 2.5GHz
  - Default is 2.2GHz

epcc

Frequency and memory scaling & STREAM



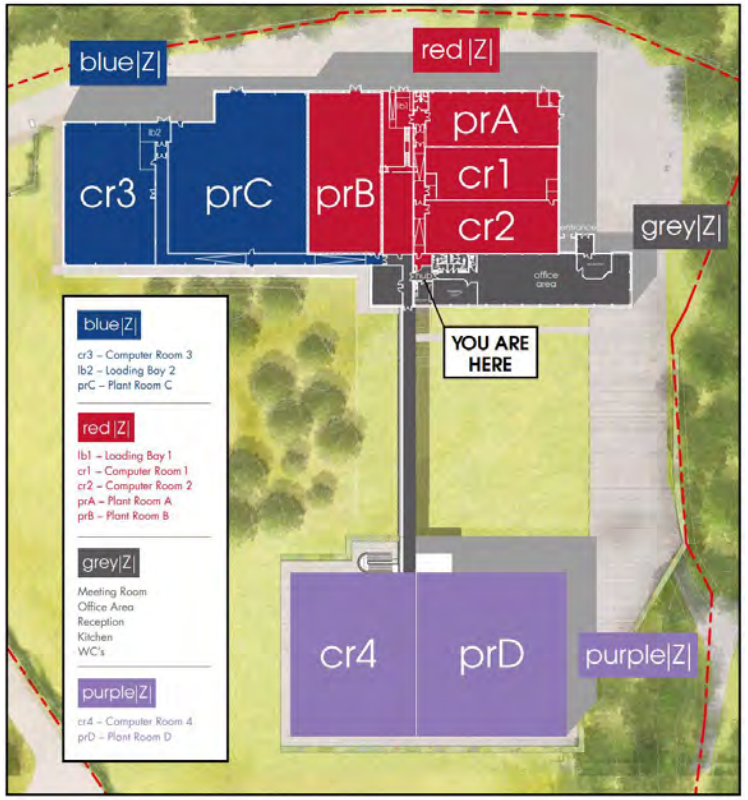
- Marvell ThunderX2 Arm-based test system
- Influencing CPU frequency using different governors
- Influencing memory subsystem frequency by enabling memturbo
  - Increased from 2.2 to 2.3GHz
  - CPU can no longer boost and is limited to 2.2GHz

EPCC's path towards Net Zero

epcc

Long term investment

- Data centre efficiency requires **long term strategic** investment
  - There is of an (inevitable) upfront cost in emissions
- Oldest ACF machine room from 1970s
- Infrastructure must support new developments in power and cooling



epcc

Renewable energy

Procured using 100% **certified** renewable energy framework agreement

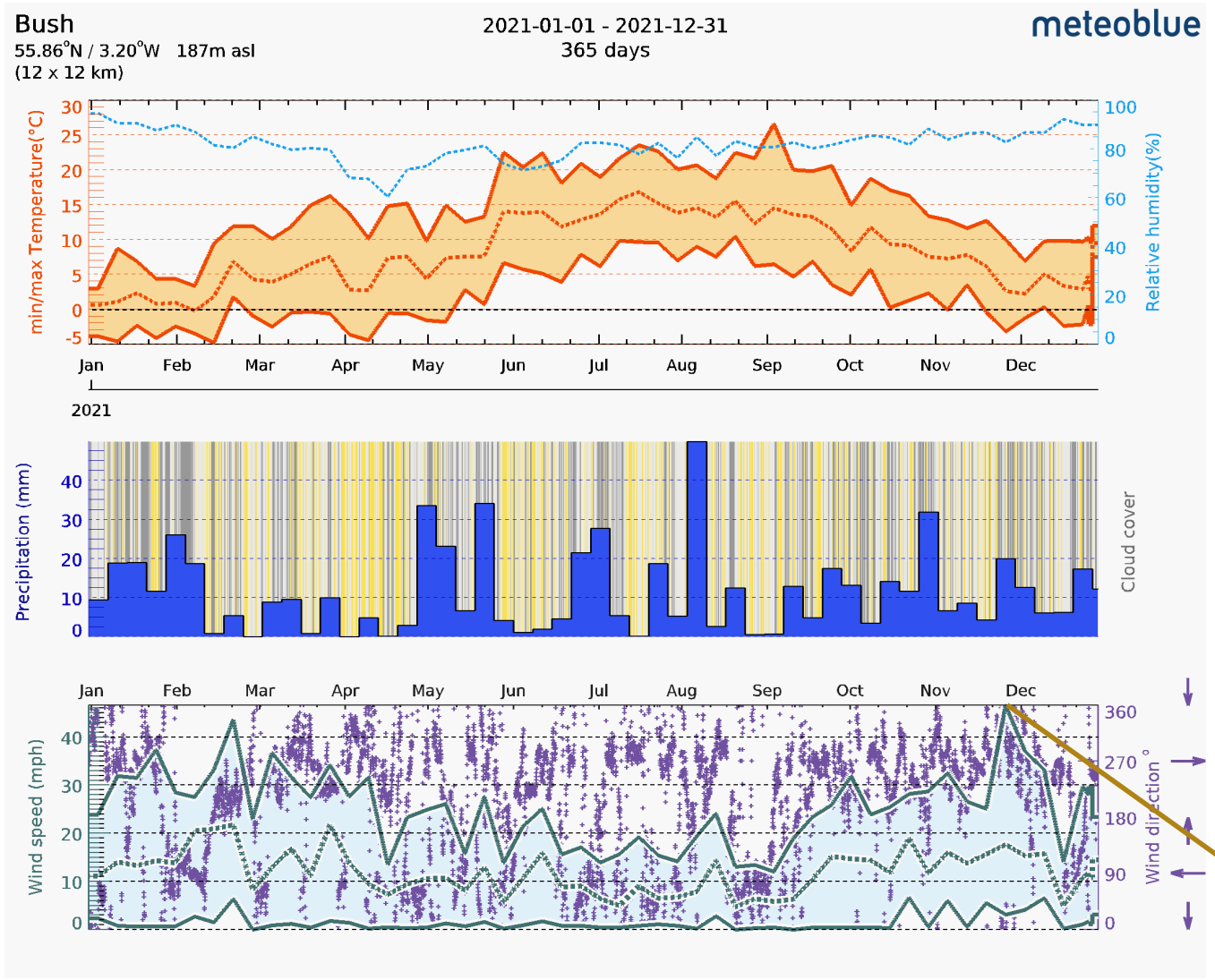
- ✓Renewable Energy Guarantee of Origin (REGO)
- ✓Classed as Net Zero

Both **cost** and **carbon** efficient

So, job done?
Well not quite!

https://www.gov.scot/publications/electricity-framework-agreement/

Weather at the ACF



Free cooling!

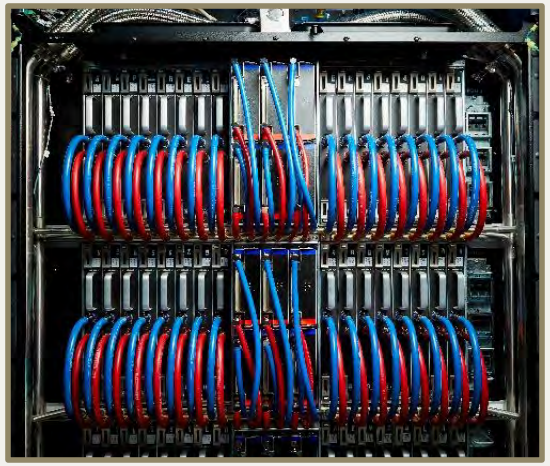
Warm water is pumped to the roof, outside air & fans chill it

Storm Arwen…

https://www.meteoblue.com/en/weather/historyclimate

epcc

Other forms of cooling we use



Direct liquid cooling

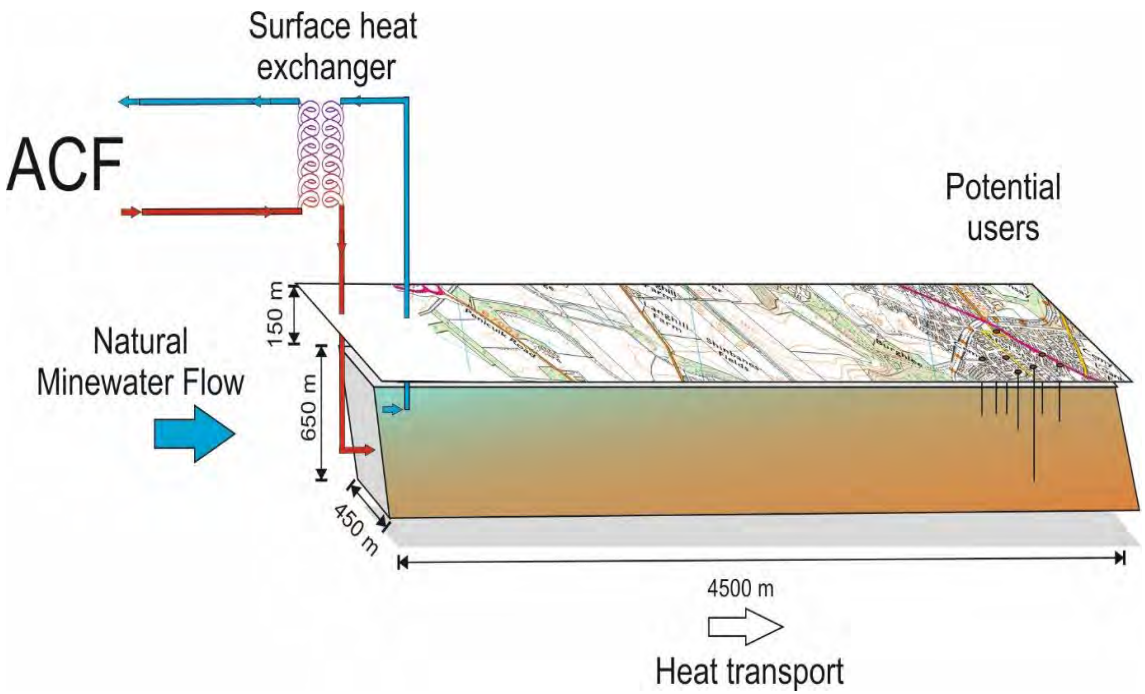
Adiabatic cooling

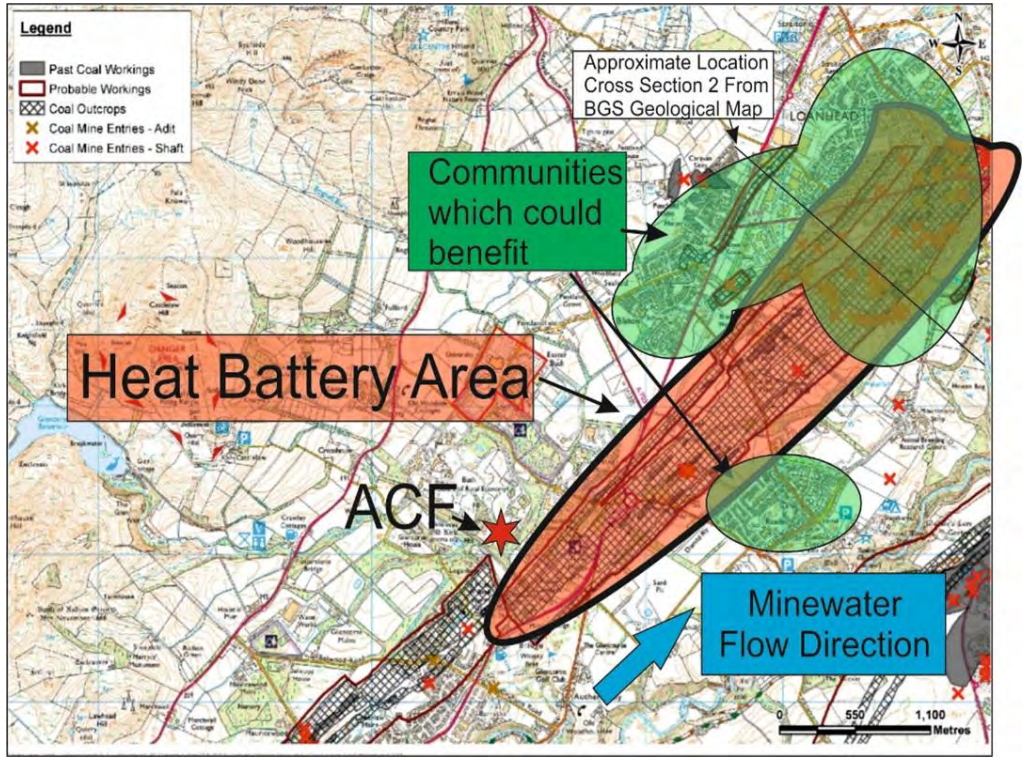
Temperature controlled racks

Geothermal battery feasibility project

Problem: We want to be able to reuse our excess heat, but nowhere nearby can use it

Solution: Move the heat to where it is useful

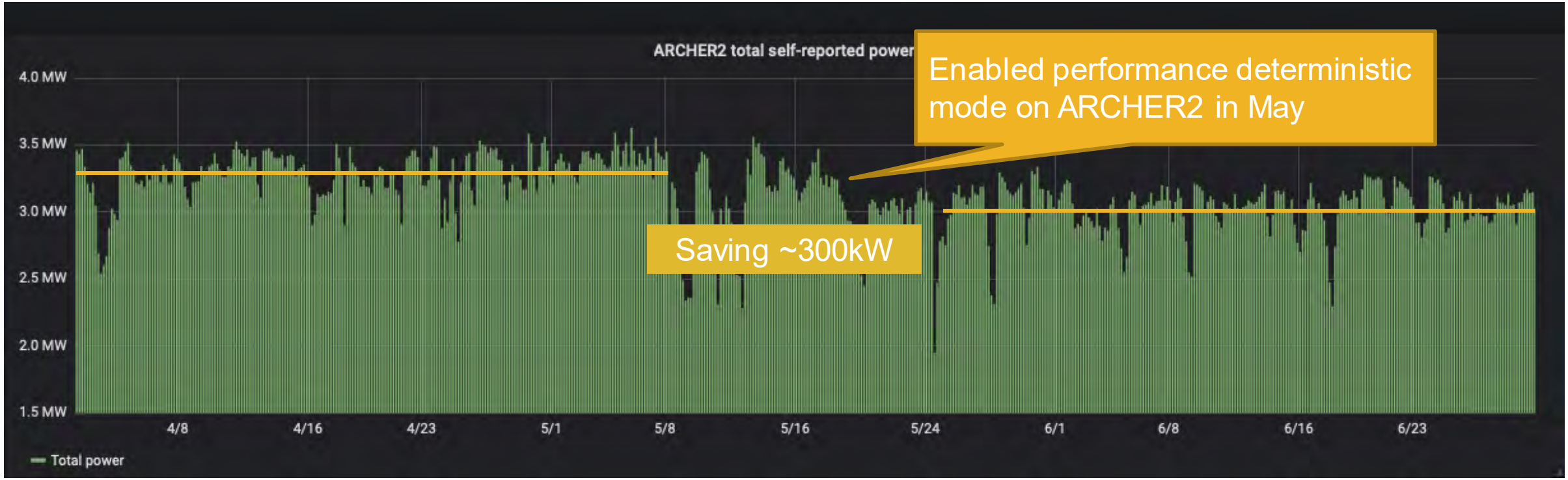




epcc

EPYC modes

- **Power** deterministic
  - Allows the highest possible performance
  - CPU will run as fast as it can for given TDP or power input – might vary
- **Performance** deterministic
  - Will deliver the same predictable performance across CPUs
  - Might result in slightly different power consumption

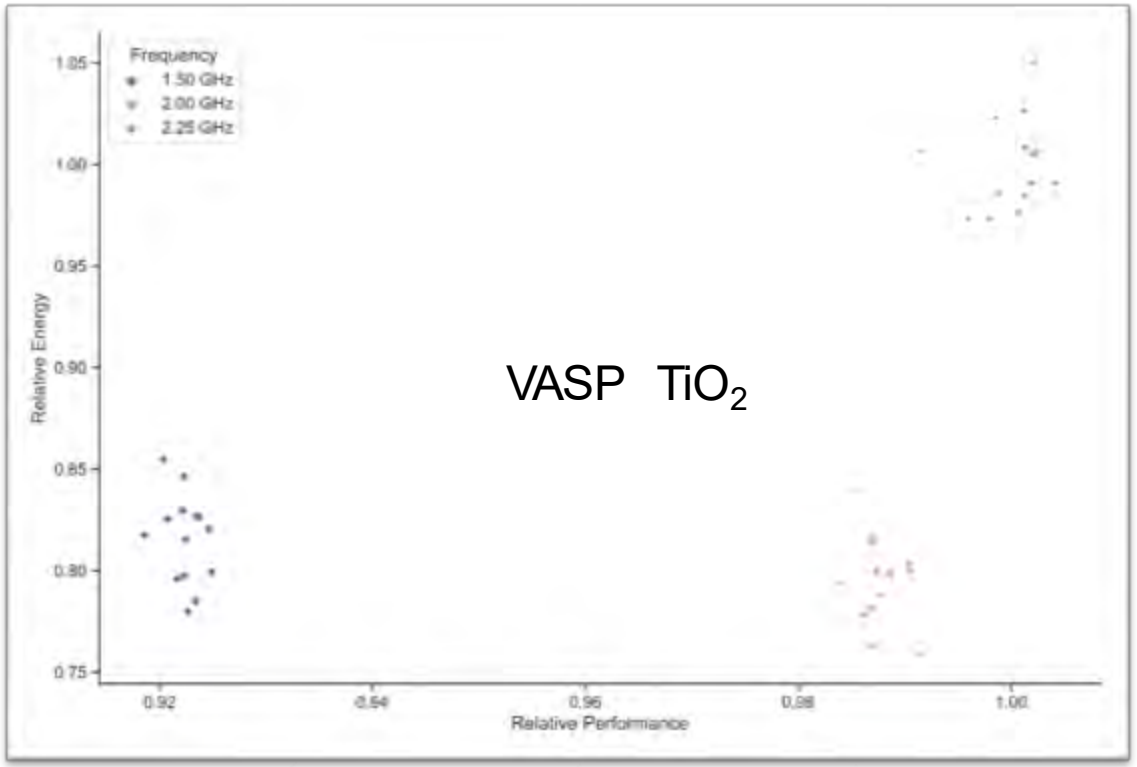


ARCHER2 CPU frequency reduction

Summary of relative energy and performance at 2.00 GHz, compared to 2.25 GHz

Benchmark (single node)	Energy	Performance
VASP (TiO_2)	-20%	-1%
CASTEP (Al Slab)	-13%	-1%
GROMACS (1400k atoms)*	-5%	-15%
OpenSBLI (TGV 512ss)	-20%	-5%
LAMMPS (LJ 8M atoms)**	-4%	-21%
NAMD (STMV 1M atoms)**	-5%	-33%

* Data from Laura Moran, EPCC
** Data from Douglas Shanks, HPE



- Investigating reducing default CPU frequency from 2.25GHz to 2.0GHz
- Allow users to override this default, provide higher frequency defaults for some codes

epcc

EPCC's contributions to Net Zero DRI project

- Taking snapshots of IRIS digital research infrastructure energy consumption over fixed periods of usage
- Developing a carbon model to evaluate the overall carbon usage of IRIS for those periods, taking into account active usage (energy used during the runs) and embodied usage (carbon spent building & providing the DRI)
- HPC-JEEP are presenting in the UKRI Net Zero DRI Project session (Friday morning)

ARCHER2 Net Zero Case Study

- Commissioned by Net Zero DRI project
- Aim is to understand the emissions resulting from the operation of the ARCHER2 service
- Case Study to be published in 2023

- Power, cooling and heat re-use
- Minimising emissions
- System software changes for energy reduction
- Energy utilisation at the application level
- Influencing user behaviour

Education

Efficient software is important → the **developers**' responsibility

Efficient *use* of software is equally important → the **users**' responsibility

Deployment of software is important → the **system providers**' responsibility

- Education is key – enable developers/users/system providers to understand implications of their choices
  - As a University department and national service provider, one of our core missions
  - EPCC's MSc in HPC has included lectures on power & energy efficiency for several years

Final thoughts

HPC systems are **scientific instruments** that are used to find solutions to many of the problems humanity faces

→ to discover new vaccines

→ to design new renewable energy solutions

→ and even to model the climate, in order to more accurately predict climate change and its impact

Significantly reducing scientific throughput is a false economy

Net Zero HPC must be achieved in a fair manner while maintaining, or indeed increasing, the amount of science we do

Net Zero HPC is not just a “noble dream”…

- Very much an inescapable reality
- Provided a perspective **from EPCC’s point of view**
  - All centres/sites will be different, but striving for operationally efficiency must be at the heart
- Renewable energy provision is key
  - Challenge is that (global) demand is continuously growing

UK Research and Innovation

Session Overview

09:30 - 09:50 **Project Overview** (Martin Juckes)
09:50 - 10:00 **HPC-JEEP** (Alastair Basden and Andy Turner)
10:00 - 10:10 **IRISCAST** (Jonathan Hays)
10:10 - 10:20 **ENERGETIC** (Deepayan Bhowmik and Teymoor Ali)
10:20 - 10:30 **CARBON-QUANDRI** (Daniel Schien)
10:30 - 11:00 **Panel Discussion**
(Chair: Ag Stephens; Panel: Martin Juckes, Wim Vanderbauwhede, Justin O'Byrne)

UK Research and Innovation

NET ZERO DIGITAL RESEARCH INFRASTRUCTURE (DRI)

Roadmap for UKRI to reach Net Zero by 2040

Goal: actions to reduce the carbon emissions from data generation, analysis, storage and dissemination.

PLEASE HELP if you are the manager, supervisor or contact of a UKRI-owned/majority-funded "facility". **COMPLETE OUR SURVEY** to help us map the carbon landscape of the UKRI DRI.

UK Research and Innovation

https://bit.ly/netzerodri

https://net-zero-dri.ceda.ac.uk

support@ceda.ac.uk

UK Research and Innovation

UKRI Net Zero Digital Research Infrastructure Scoping Project

https://net-zero-dri.ceda.ac.uk/

Martin Juckes, Charlotte Pascoe, Ag Stephens, Poppy Townsend, Katie Cartmell, Jen Bulpett

CIUK, Manchester, Friday 1st December 2022

UK Research and Innovation

Project Ambition

- Collect **evidence to inform** UKRI Digital Research Infrastructure (DRI) Investment decisions
- Provide UKRI and their community with an **outline roadmap for achieving** carbon neutrality in their DRI by 2040 or sooner
- Enable UKRI to play a **positive and leading role** in the national and global transition to a sustainable economy

UK Research and Innovation

Who we are

Scoping project (£1.8m) – ending in Summer 2023

- Core project team – CEDA/NCAS
- Science Advisory Board (Prof Mary E Black)
- Steering Committee (Anna Angus-Smyth - NERC)
- Project partners – various universities/UKRI councils are undertaking some work

Stephen Mobbs
PI

Martin Juckes
Project Lead

Poppy Townsend
Communications Manager

Jen Bulpett
Senior Project Manager

Charlotte Pascoe
Science Officer

Ag Stephens
Technical Officer

Katie Cartmell
Project Manager

Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Partners

The core team is supported by partners from 20 institutions, bringing a huge range of experience.

What is the DRI?

- The UKRI DRI is the UKRI owned and majority funded Digital Research Infrastructure

- In practice, it does not matter whether a facility is 40%, 60% or 100% funded by UKRI: this project is focused on gathering evidence to support those who want to reduce the net emissions of digital research infrastructure to zero.

UK Research and Innovation

Some Carbon Basics

- The carbon footprint is dominated by power supply and manufacturing.
- Carbon offsets do not increase costs much, but does not work very well, if at all.
- Carbon sequestration appears to work, but could double costs. Sequestration costs could go down as technology increases or go up if/when demand outstrips supply.
- Drastic reductions in emissions are needed; different approaches will be needed for power supply and manufacturing.
- In parallel, we need to ensure that we make best possible use of resources, so we are not paying for power consumption which can be avoided.

Carbon Budget Challenges

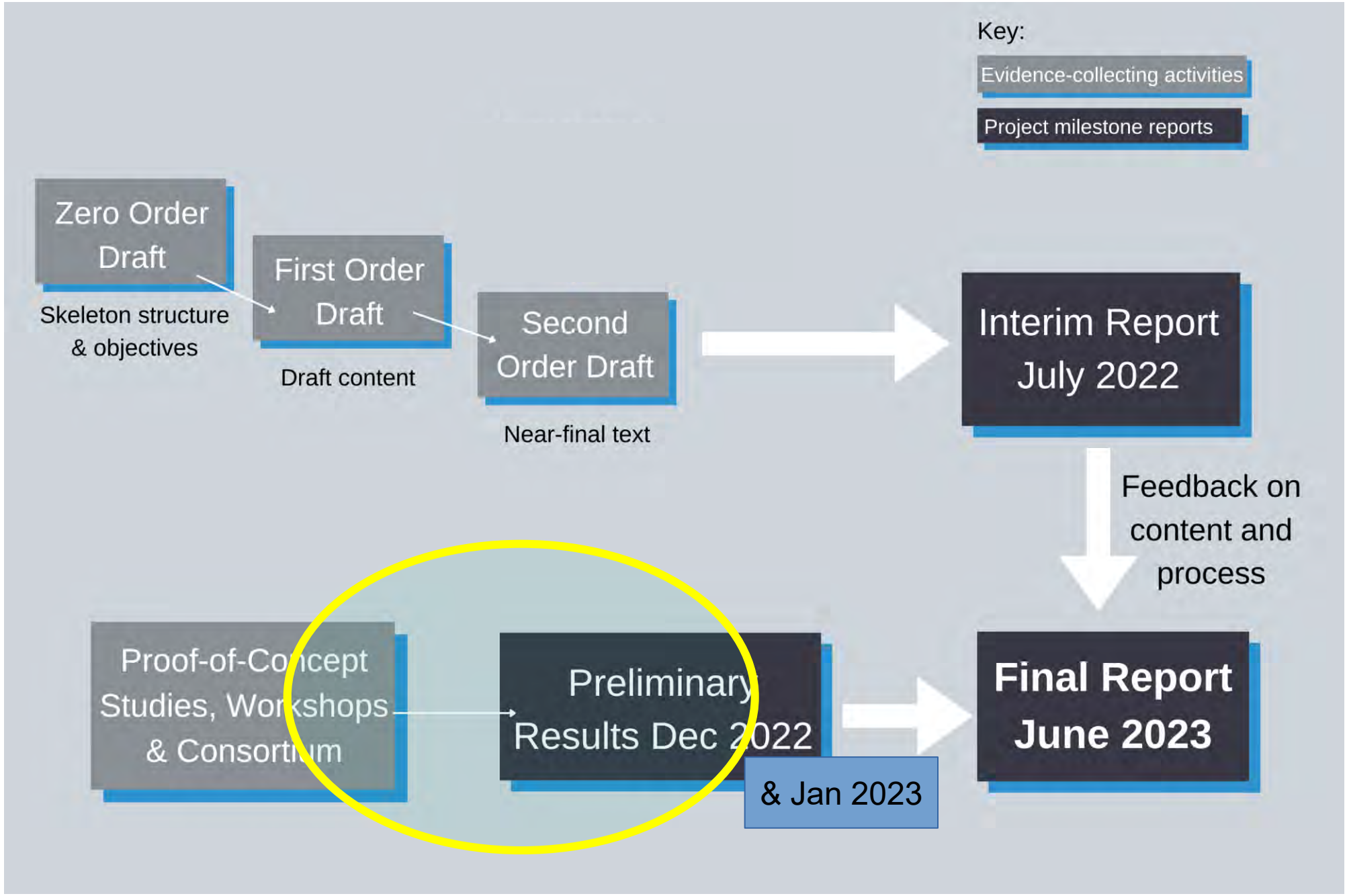
- Multiple metrics: there are 3 measures of the carbon footprint of electricity supply:
  - **National Carbon Intensity** : national annual average
  - **Purchase Carbon Intensity** : reflecting financial flows
  - **Grid Carbon Intensity** : reflecting the actual local flow of power, including high use of fossil fuels at times of peak demand
- Users do not have clear information about the footprint of their work, so little incentive to improve efficiency.
- The carbon footprint of manufacturing is large, but very poorly quantified.
- **The majority of the carbon footprint is tied up in activities which are outside the direct control of the DRI stakeholders, such as institutional electricity supply and procurement rules**

UK Research and Innovation

How are we collecting evidence?

- Consortium work
  - Mapping the existing infrastructure
  - Connecting with experts
- Project partners
  - Sandpit projects
    - 6x proof-of-concept studies
    - 1x workshop series
  - We've partnered with 19 institutions!
- User behaviour & implicit associations survey
- Literature survey
- Assess funding framework
- Engagement with stakeholders
  - Feedback on interim report

Timeline for key project outputs



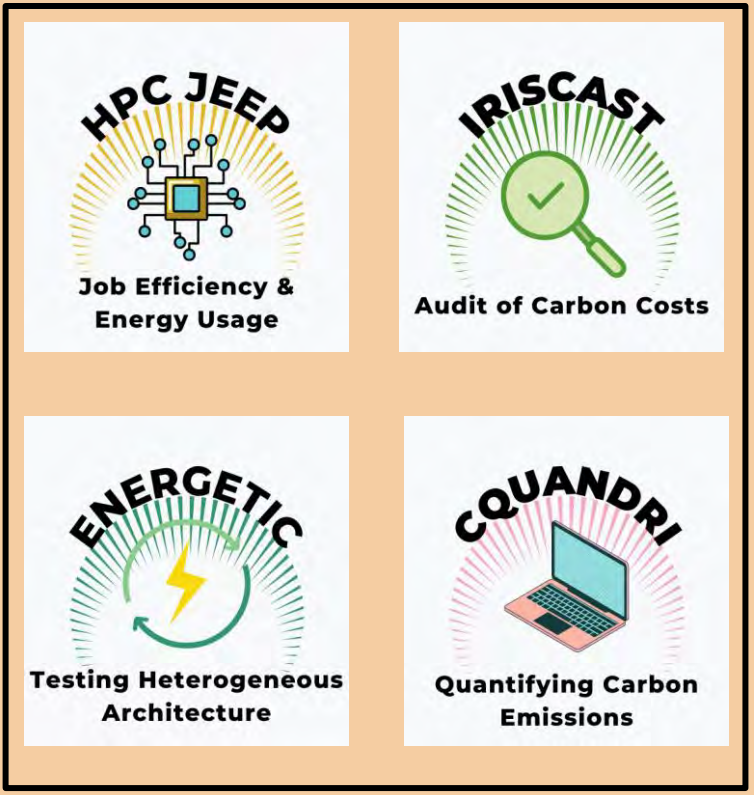
UK Research and Innovation

Sandpit events

Sandpit A - 9th and 11th May: community and organisational challenges
Sandpit B - 23rd and 25th May: technical and operational challenges

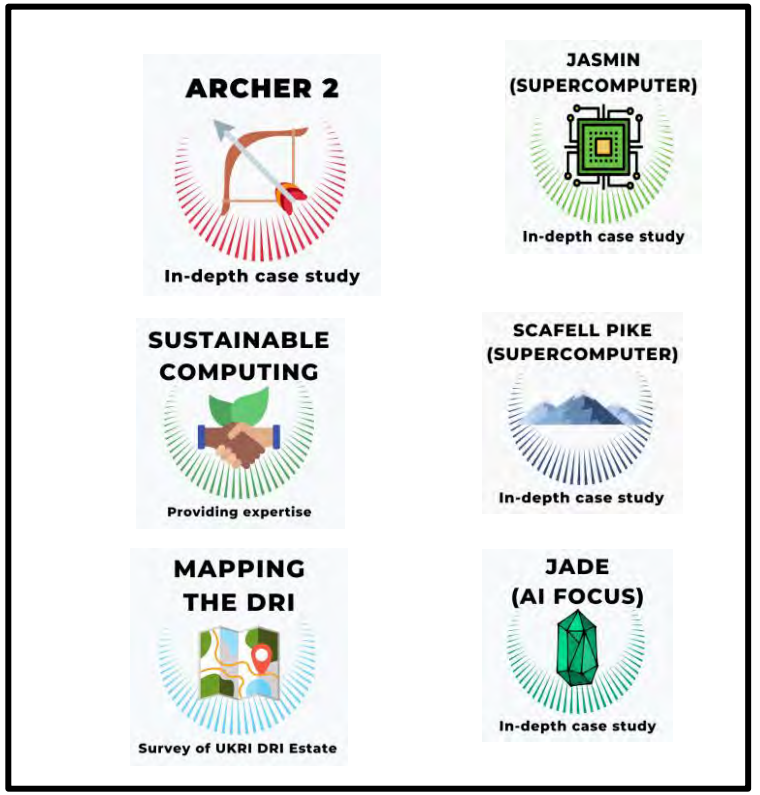
- Each sandpit consisted of two 3-hour online sessions (Monday and Wednesday):
  - setting the scene, meeting each other, exploring ideas.
  - forming teams, defining objectives
- Short proposals were submitted on Thursday and evaluated by a panel on Friday
- 7 projects funded, 4 on technical and operational challenges (presenting today) and 3 on community and organisational challenges.

UK Research and Innovation



The UKRI Net Zero DRI Scoping project contains 9 consortium projects (right) and has funded 7 additional projects (left) through a sandpit event.

These projects will investigate a broad range of technical and social issues related to the Net Zero target.



Machines and Workflows

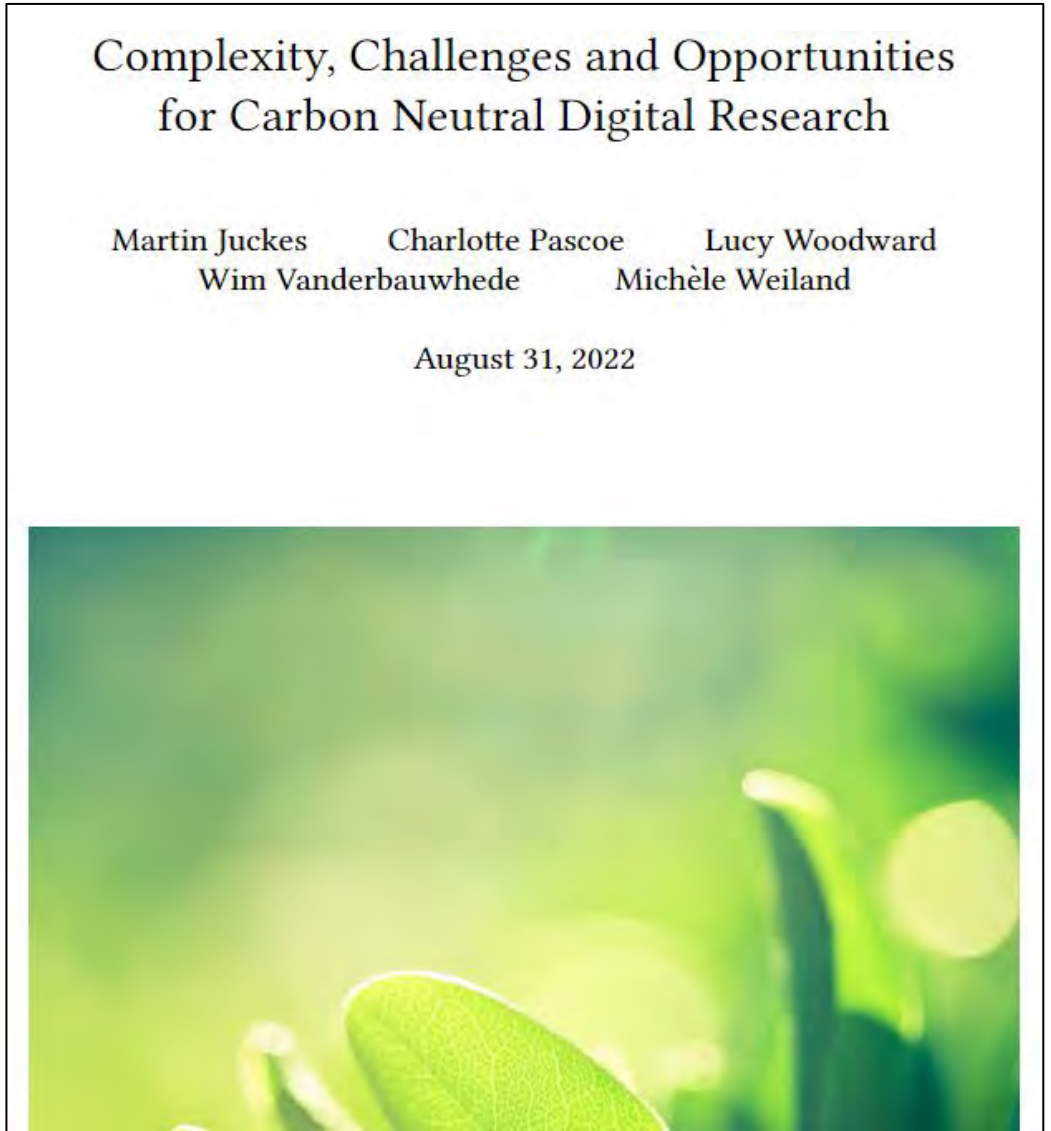
People and Process

The interim report : published August 2022

- Initial findings based on a literature survey and stakeholder engagement

bit.ly/nzdri_interim

UK Research and Innovation



Selected Interim Recommendations

- Build consensus, lead by creating a space for ideas, adopt best practice
- Use multi-year contracts for electricity supply; exploit on-site generation and storage
- Use contracts and other lines of influence to reduce carbon intensity of supply chain
- Invest in people to improve efficiency of resource use
- Develop policies to ensure that efficiency leads to lower carbon footprint rather than all going to higher throughput
- Reduce emissions as much as possible and explore all options for removing carbon from the atmosphere
- **Improve quantification of the immense societal benefit delivered by the UKRI DRI in parallel to improving quantification of carbon footprint**

Coming soon

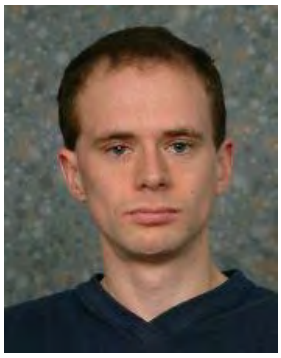
- Sandpit project final reports: January 2023
- Stakeholder workshops: February 2023
- Cross-UKRI Workshop : May 2023
  (Showcasing findings; review recommendations)
- Publication of project conclusions: June 2023

https://net-zero-dri.ceda.ac.uk/
support@ceda.ac.uk

UK Research and Innovation


UK Research
and Innovation
Thank you
Get in touch: support@ceda.ac.uk

Alastair Basden and Andy Turner - HPC-JEEP



Abstract: HPC-JEEP is analysing the compute node energy use data available from the ARCHER2 and DiRAC COSMA HPC services to understand what meaning can be extracted from this type of data to help users, service providers and funders analyse the energy/emissions and how they can contribute towards net zero goals. In particular, we are looking at how we can analyse energy use by project, research area or software across the whole services; how you might implement a charging scheme with a component of energy use; and what energy/emissions metrics can be reported back to users, service providers and funders based on compute node energy use data.

Bios: Alastair is a HPC manager for the DiRAC Memory Intensive service at Durham University, and a member of the DiRAC technical directorate.

Andy is a Principal Architect at EPCC and provides technical leadership for the ARCHER2 Computational Science and Engineering (CSE) service.

HPC JEEP

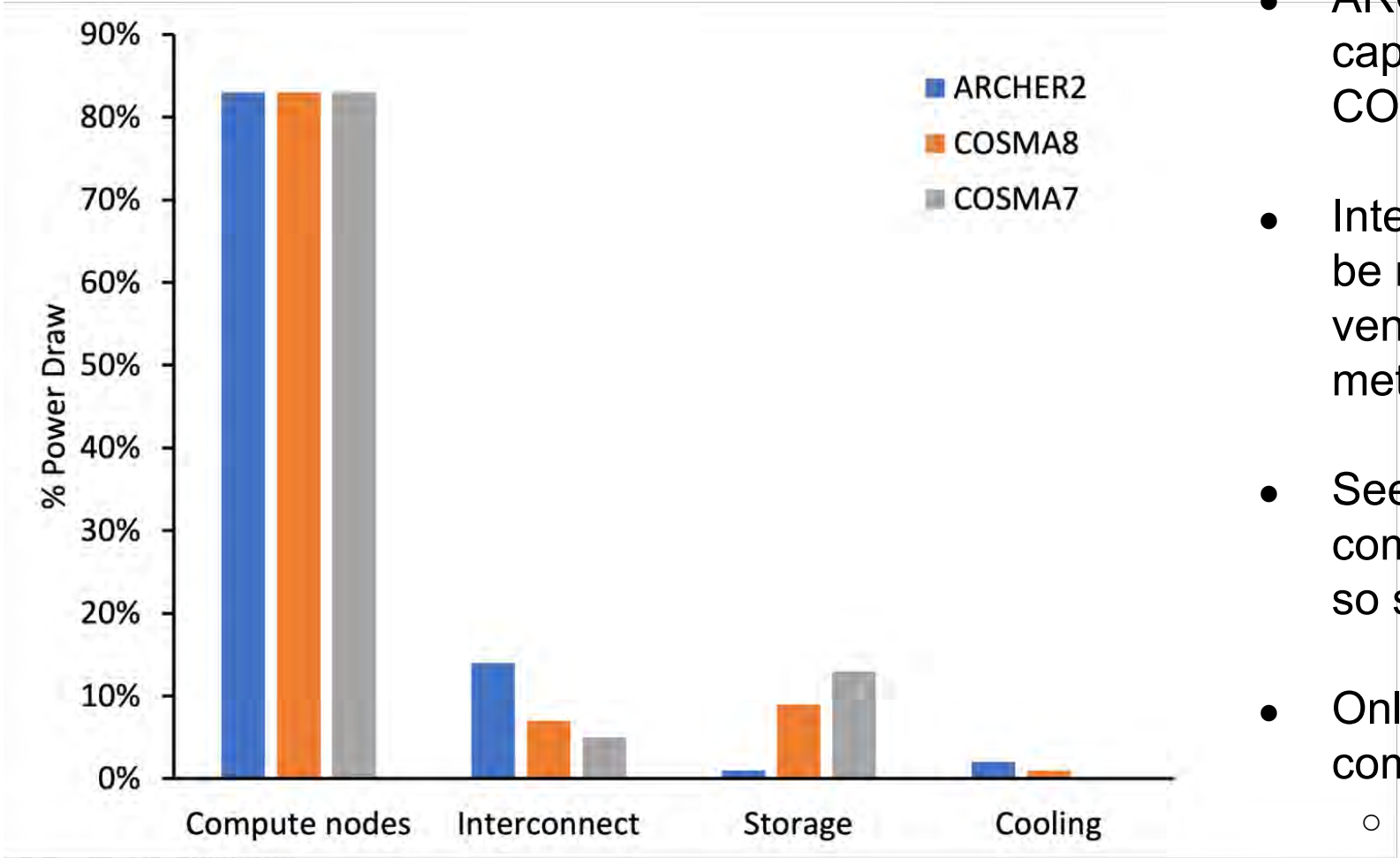
HPC Job Efficiency and Energy Profiling

CIUK
December 2022
Andy Turner (EPCC)
Alastair Basden (Durham/DiRAC)

HPC-JEEP scoping project aims

- Understand what per-job energy data we currently have from HPC systems and what types of analyses this data can support to help transition towards net zero
    - Report: https://doi.org/10.5281/zenodo.7128628
- Understand if the energy use data can potentially support introduction of energy-based charging
- Propose energy (and, potentially, emissions) metrics that can be provided back to HPC service stakeholders to help them transition towards net zero

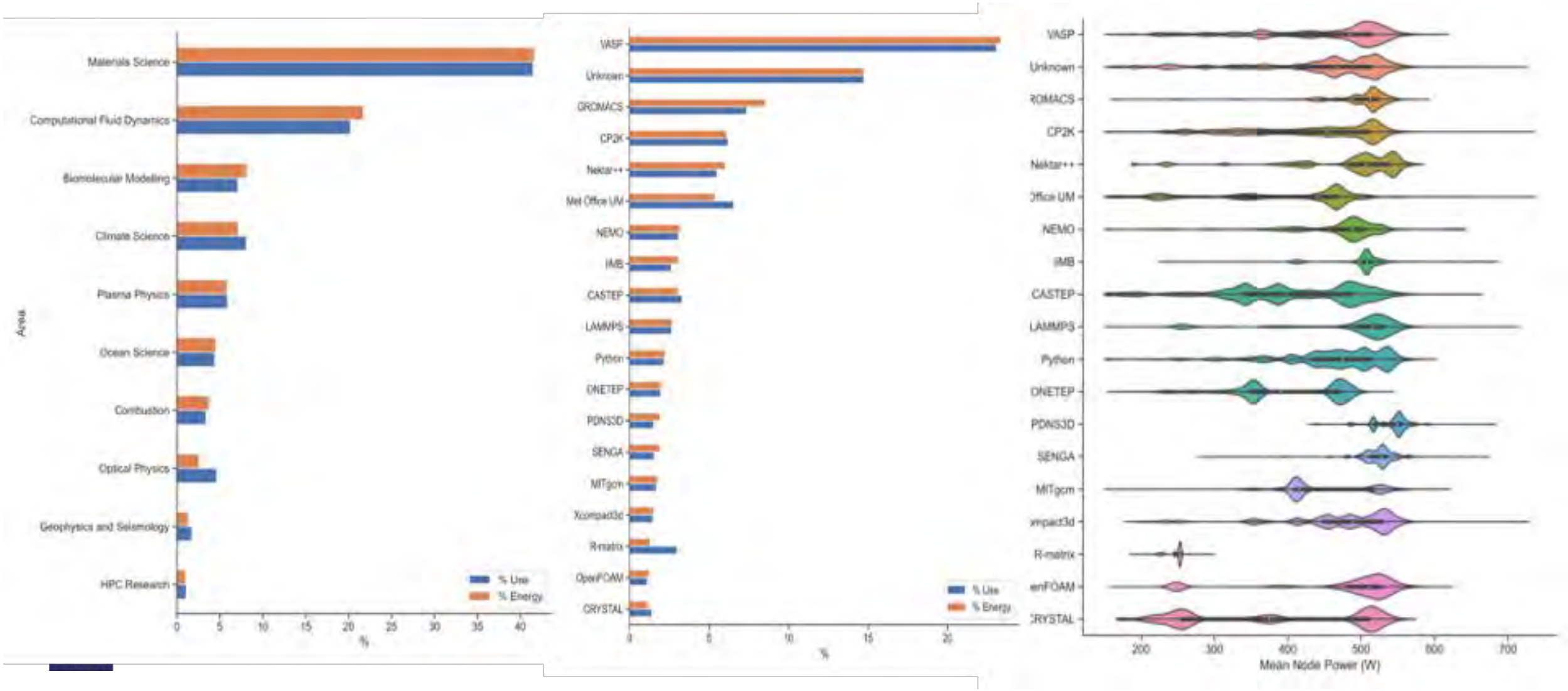
Power use on ARCHER2/COSMA systems



- ARCHER2 has lower storage capacity per node than COSMA
- Interconnect differences could be real or due to different vendor measurement methodologies
- Seems coincidental that the compute node numbers are all so similar
- Only includes “in-cabinet” components
  - Cooling is cabinet CDU, rather than plant rooms

Analysing ARCHER2 energy data

Methodology and tools at:
https://doi.org/10.5281/zenodo.7128628



Note: this is only compute node

Energy-based charging on ARCHER2

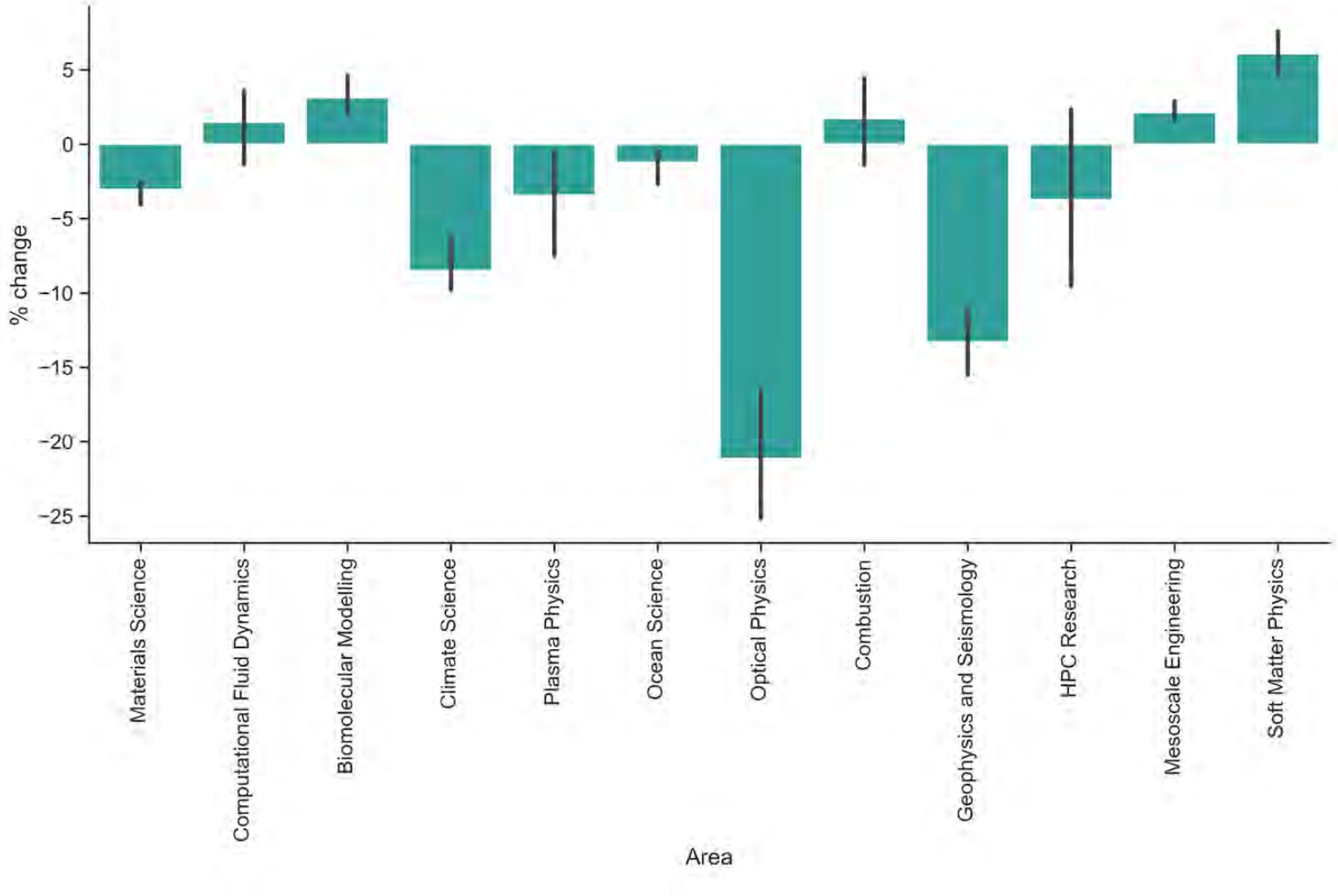
Charging based on:

50% Residency - how many nodes you have for how long

50% Energy - how much energy the job uses

Comparison is to 100% residency charge (nodeh) for 3 month period. Bars indicate range of monthly variation.

Overall reduction in total charge by 3% - corresponds to an overall 3% allocation boost unless allocations are updated.



Embodied energy - embodied CO_2

- Building an HPC system embeds CO_2 produced during production
  - HPC systems are often produced in countries with high carbon intensity
- COSMA7 compute:
  - Dell C6420 servers: 1,240 $kgCO_2$ (according to Dell)
  - In production for ~4 years (so far)
  - 452 nodes, total energy consumption ~3200 MWh (7 MWh/server)
  - Average CO_2 intensity in North East over previous 13 months ~38 gCO_2/kWh
    - 269 $kgCO_2$ per server (over the 4 year lifetime), 67 $kgCO_2$ per year:
      - 18 years for production CO_2 to equal embodied CO_2
    - Embodied % for a 4 year lifetime is ~80%
    - This will only increase as the UK national grid greens
  - Note: Calculation different depending on CO_2 intensity.
    - E.g East Midlands, 280 gCO_2/kWh: 500 $kgCO_2$/year, 2.5 year payback
    - Embodied % for a 4 year lifetime is ~40%

Embodied CO_2 notes

- A per-region approach is not necessarily valid
  - UK has a national grid
  - Average national CO_2 intensity over past ~3 months is ~180g / kWh
  - So, 320kg CO_2/year from COSMA7 nodes
    - 4 years operation for embodied CO_2 to equal compute production CO_2
      - Will increase in future years (assuming embodied CO_2 doesn't change)
- How long should we be running systems for?
  - 4 years means CO_2 is ~50% embodied
  - 8 years seems reasonable (though obviously, many factors)
    - 33% of CO_2 produced will be due to the embodied part
    - Probably longer than we currently do!
- Important to push suppliers for lower embodied energy

User Reporting

- Quarterly emails sent to COSMA users and project PIs
  - Total energy used by their jobs for each user
    - Compute node
    - Estimate of fraction of storage/fabric
    - Carbon intensity value over that period
    - Mass of CO_2 generated
    - Some context (flights, miles driven, household usage, etc)
  - Total energy used by each project
    - And a list of largest users
- In future, UKRI may charge by kWh rather than core-hour
  - Helps to advise on how much to apply for
  - Benefit to making codes more efficient

UK Research and Innovation

Summary

- Providing users with a summary of their compute CO_2
- Providing UKRI with recommendations for future systems
  - Both procurement and operation

IRISCAST: IRIS Carbon Audit Snaphot

J. Hays – IRIS Science Director
IRISCAST Project PI

CIUK 2022 – UKRI NetZero Scoping Project
2nd December 2022

UK Research and Innovation

eInfrastructure for Research and Innovation for STFC

IRIS is a cooperative community bringing together (mainly) STFC computing interests

Formed bottom up by science communities and compute providers

Works closely with STFC but run by the community

UK Research and Innovation

IRIS-CAST – The **C**arbon costing for computing **A**udit **S**napsho**T**

Good robust decisions need good robust information

Challenges/questions

Estimating the carbon costs for scientific computing across a broad heterogeneous landscape

Identifying the key drivers

Identifying the hurdles and barriers

Communicating the costs to drive change

Working coherently across different communities

Actions and Objectives

Work together coherently across different facilities with different remits, tooling, and capabilities.

Learn by doing!

Document the gaps, the barriers and the issues, drive requirements for future work and decision making

Communicate across our communities and build a foundation for future action

IRIS-CAST – The **C**arbon costing for computing **A**udit **S**napsho**T**

Good robust decisions need good robust information

IRISCAST is a 6 month project funded within the UKRI Net Zero Scoping Project

Project Team

Alison Packer (STFC)
Anish Mudaraddi (STFC)
Derek Ross (STFC)
Dan Traynor (QMUL)
Jon Hays (QMUL)
Alex Owen (QMUL)
Dan Whitehouse (Imperial)
Adrian Jackson (Edinburgh)
Alastair Basden (Durham)
Nic Walton (Cambridge)
Alex Ogden (Cambridge)

IRIS-CAST – The **C**arbon costing for computing **A**udit **S**napsho**T**

Good robust decisions need good robust information

Facilities

	STFC SCARF
QMUL GridPP Tier 2	DiRAC (Durham)
Imperial GridPP Tier 2	Cambridge IRIS
STFC SCD Cloud	HPC/Cloud

UK Research and Innovation

IRIS-CAST – The **C**arbon costing for computing **A**udit **S**napsho**T**

Good robust decisions need good robust information

Inventory

Data collection

Analysis

Community Engagement

UK Research and Innovation

Good robust decisions need good robust information

DONE

Inventory

- Define the scope of the audit
- Build a comprehensive list of all equipment covered by the audit
- Needed to build carbon model including embodied costs

Good robust decisions need good robust information

Data Collection

DONE(ish)

- Collect data over a 24 hour period covering differing operating conditions
  - Rack, Node, and Job level logging
- Store data in central repository

Good robust decisions need good robust information

In progress

Analysis

- Integrate the different datasets into coherent curated data set
- Refine carbon model
- Extract insights, observations, and conclusions

Good robust decisions need good robust information

In progress

Community Engagement

- Talk at CIUK
- Produce draft report
- Publish curated data set and definition of the carbon modelling
- Engage with our communities through an IRIS Workshop – 6th, 7th January in Cambridge

UK Research and Innovation

IRIS-CAST – The **C**arbon costing for computing **A**udit **S**napsho**T**

Good robust decisions need good robust information

Next steps…

DONE

Inventory

DONE(ish)

Data collection

In progress

Analysis

In progress

Community Engagement

Deepan Bhowmik and Teymoor Ali - ENERGETIC

Abstract: Current, leading-edge HPC systems are often heterogeneous, comprised of combinations of multiple compute units and accelerators, including (but not limited to) CPUs, GPUs and FPGAs. HPC is a significant contributor to energy usage. However, the energy-to-solution varies between these architectures. In terms of minimising energy consumption, this choice of possible architectures presents a set of challenges to HPC system maintainers and algorithm developers, such as, a) which configuration of architectures provides the lowest energy consumption? or b) which combination of architectures should a code target in order to minimise its energy consumption.

The ENERGETIC project is conducting research to answer these questions through energy measurement benchmarks on prototypical algorithms. In this talk, we shall present our initial findings and energy profiles obtained from both existing HPCs as well standalone computing systems.

Bios: Dr Deepayan Bhowmik is a Senior Lecturer in Data Science in the School of Computing at Newcastle University, UK. He is conducting research in fundamental signal and image processing, applications and their system implementations. His research interests include heterogeneous computer architecture with CPU, GPU and FPGAs, embedded and low power vision systems, computer vision, and other image processing applications. Dr Bhowmik received research fundings from various UK research councils, Royal Academy of Engineering, EU and industries.

Teymoor Ali is a research associate in School of Computing at Newcastle University, UK. His research interests include image processing, heterogenous architectures, image sensor characterisation methods and high-level synthesis tools.

Energy-aware Heterogeneous Computing at Scale (ENERGETIC)

Teymoor Ali & Deepayan Bhowmik

Newcastle University

Motivation/Aim

- Current HPCs consist of various combinations of accelerators CPUs, GPUs and FPGAs.
- Little data on the energy efficiency of codes or algorithms across different architectures
  - No established framework or methodologies.
  - Little use of existing tools.
- Project Aim: whether the use of heterogeneous architecture could significantly reduce the energy-to-solution.

Benchmark Algorithms

- Selected HPC Challenge benchmarks
  - Single Precision General Matrix Multiplication (SGEMM)
  - 2D Fast Fourier Transform (FFT)
  - STREAM (Main Memory Bandwidth)

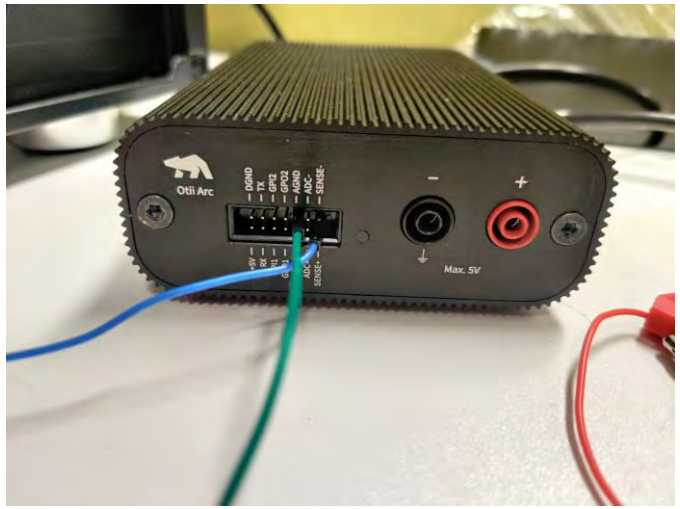
- And also deep learning based computer vision (CNN)
  - Still under processing

Benchmark systems

- Standalone Heterogenous System
  - CPU: i9-11900KF
  - GPU: Nvidia A2000
  - FPGA: Xilinx Alveo U50

- High Performance Clusters:
  - EPCC FPGA Test Bed
    - FPGA: Xilinx Alveo U280
  - Myriad
    - CPU: Xeon Gold 6240 CPU
    - GPU: NVIDIA A100

Measurement Approach - Datalogger



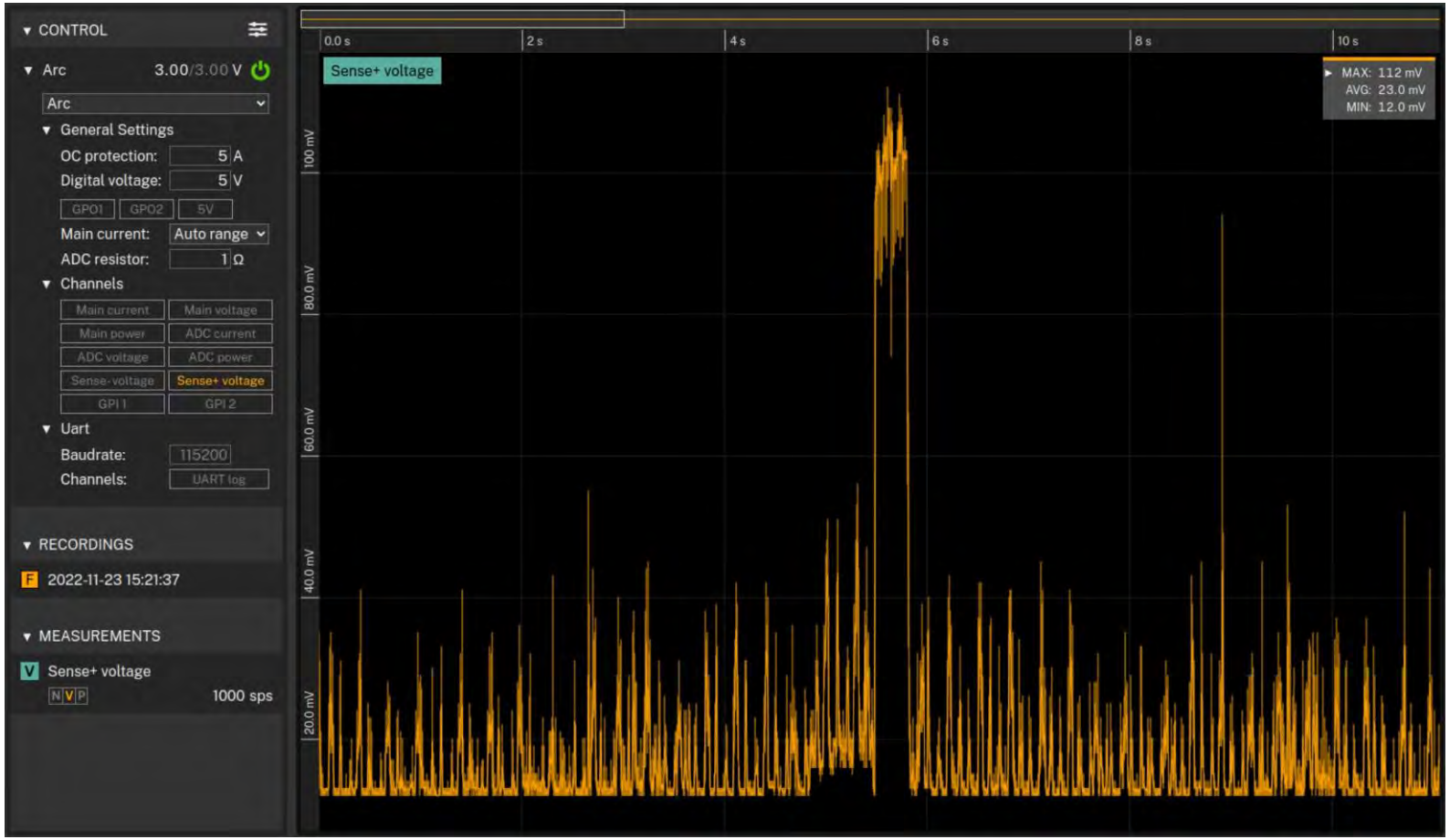
Hardware:

- Current Clamp
- Otii arc 3 datalogger
- Multimeter

Power Measurement Software:

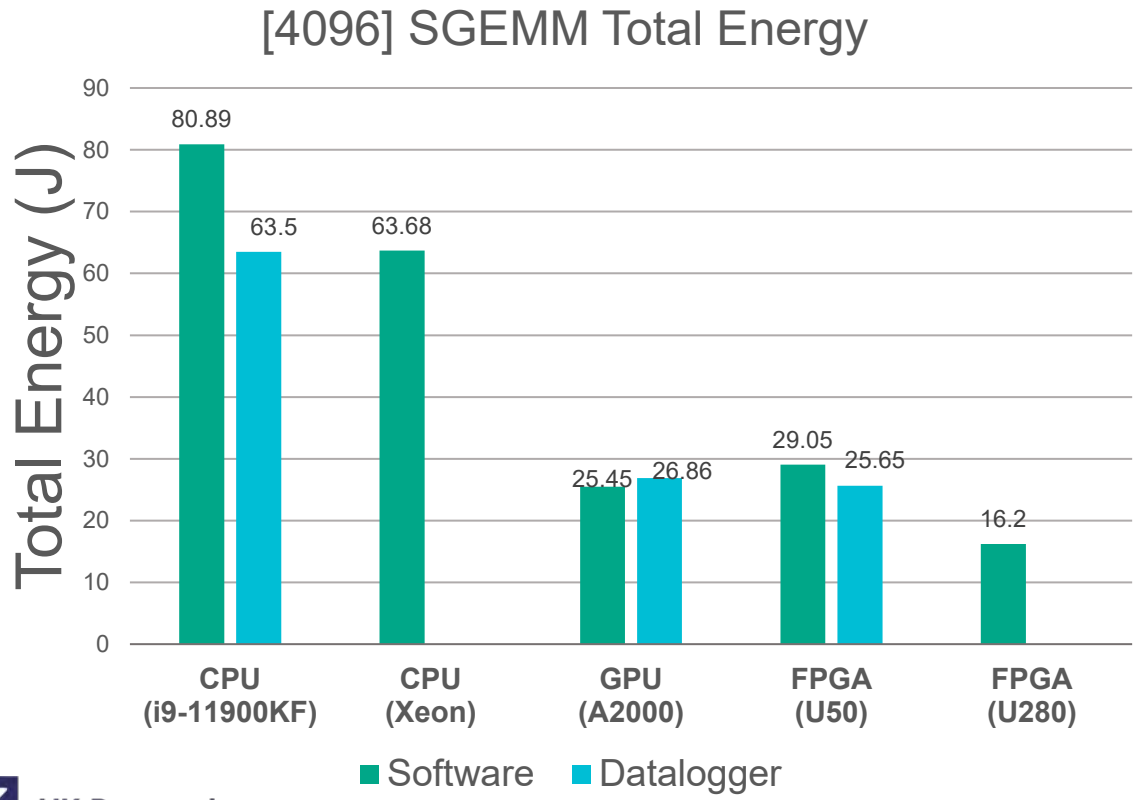
- CPU (RAPL)
- GPU (NVML)
- FPGA (XBUTIL)

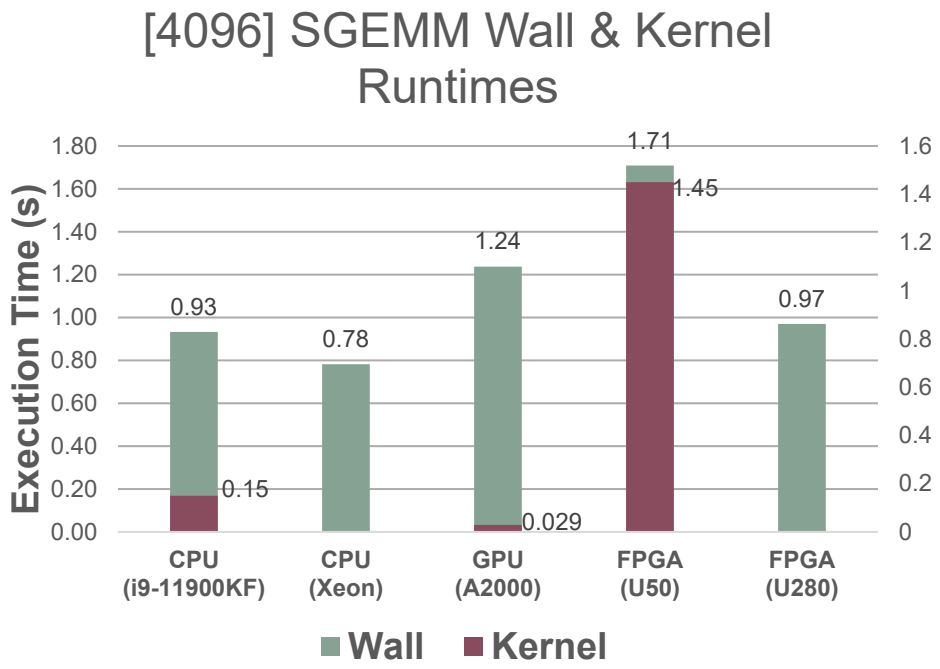
Datalogger Software



UK Research and Innovation

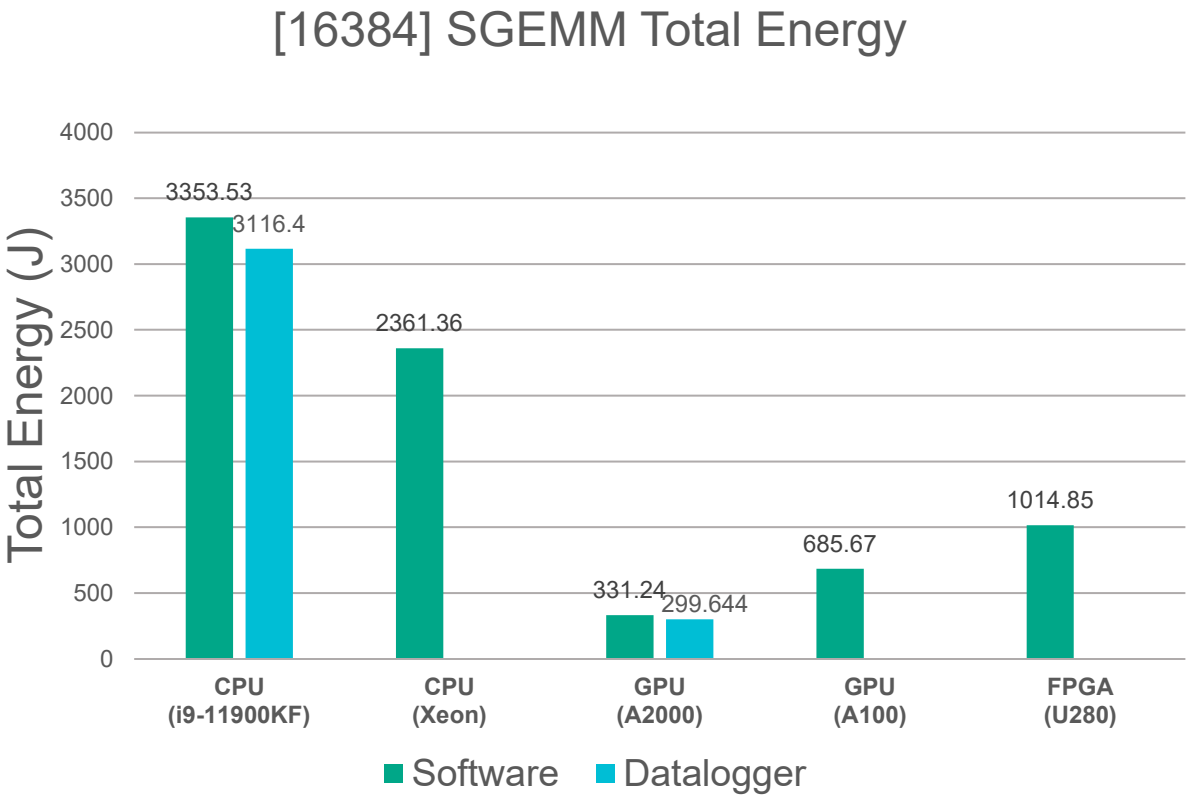
Results: SGEMM [4096]

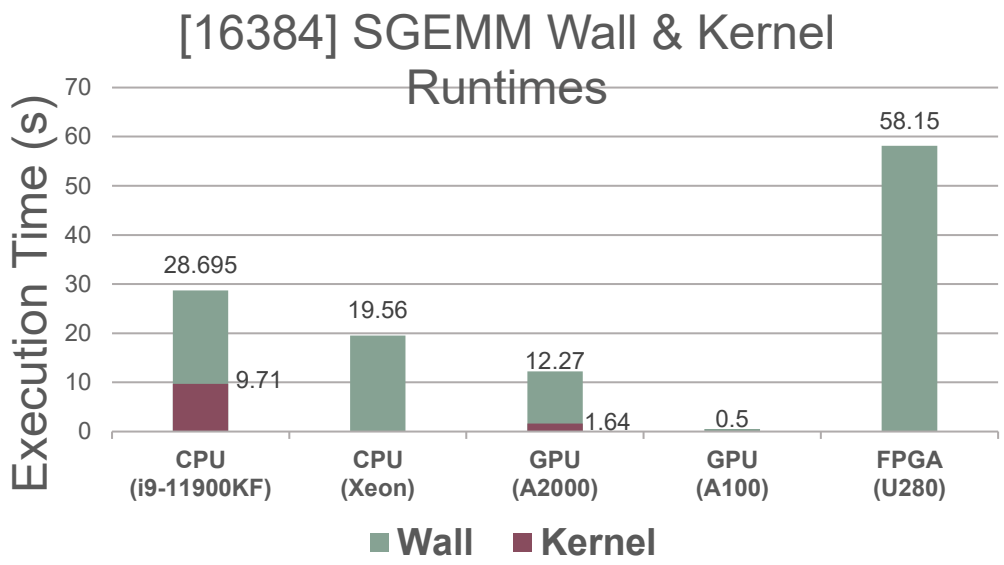




UK Research and Innovation

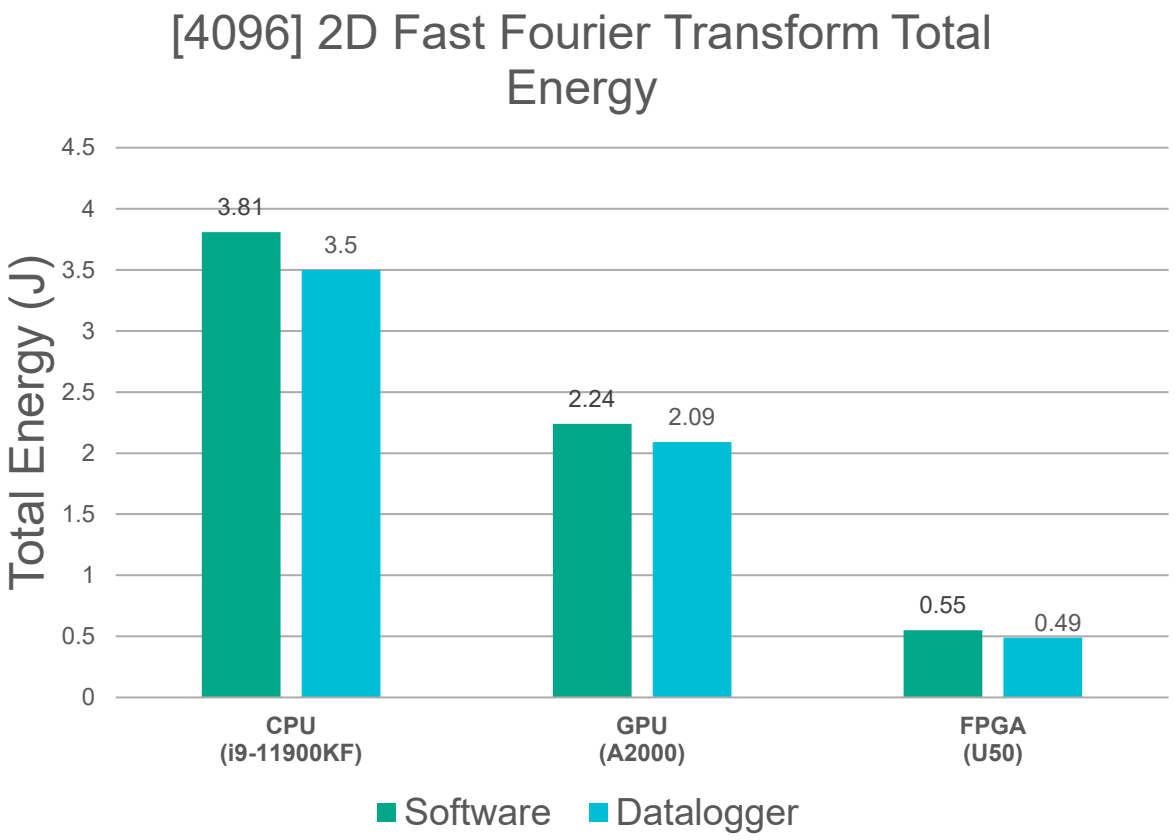
Results: SGEMM[16384]

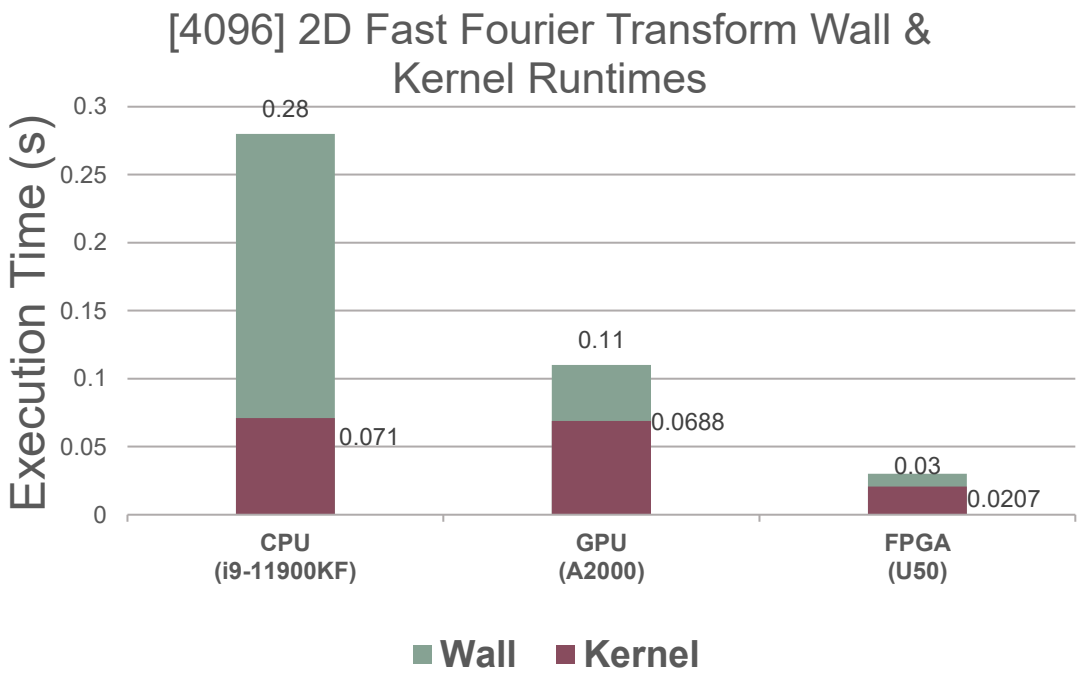




UK Research and Innovation

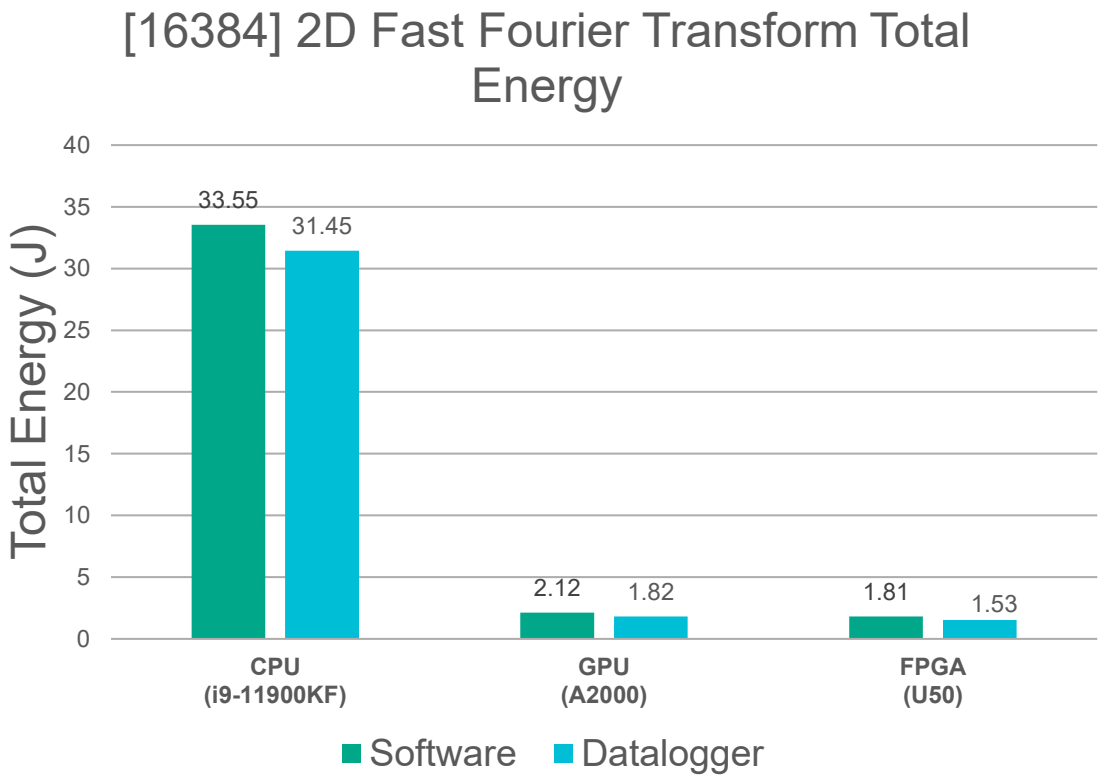
Results:2D FFT [4096]

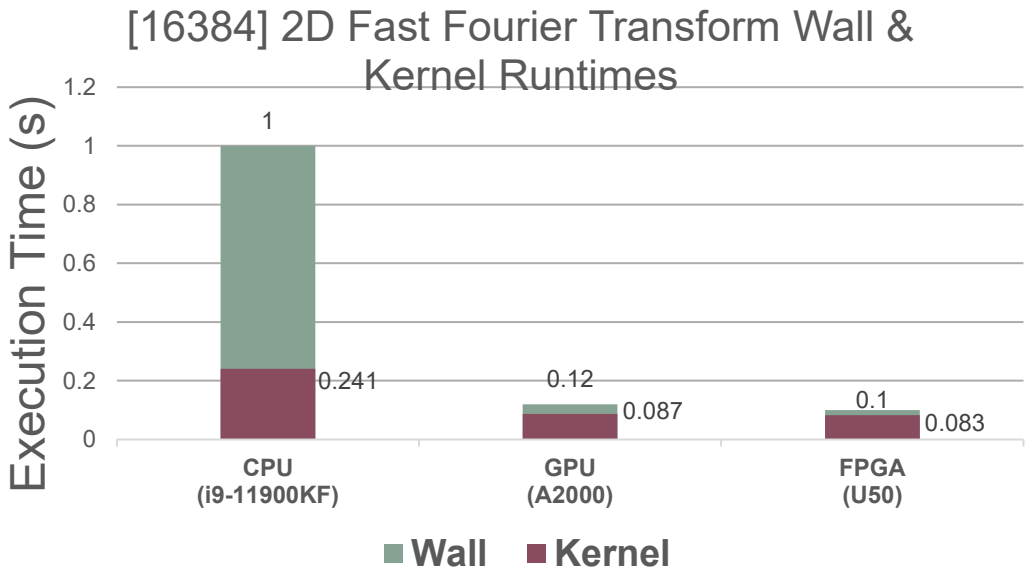




UK Research and Innovation

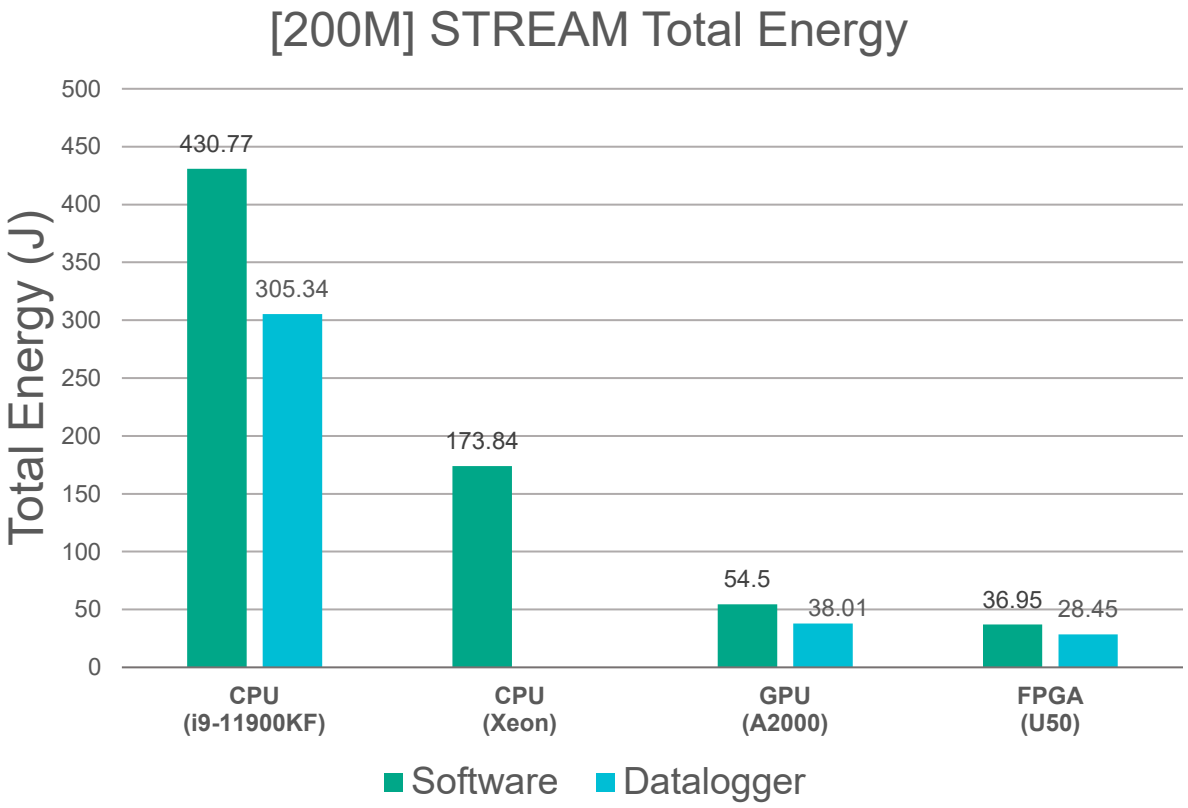
Results:2D FFT [16384]

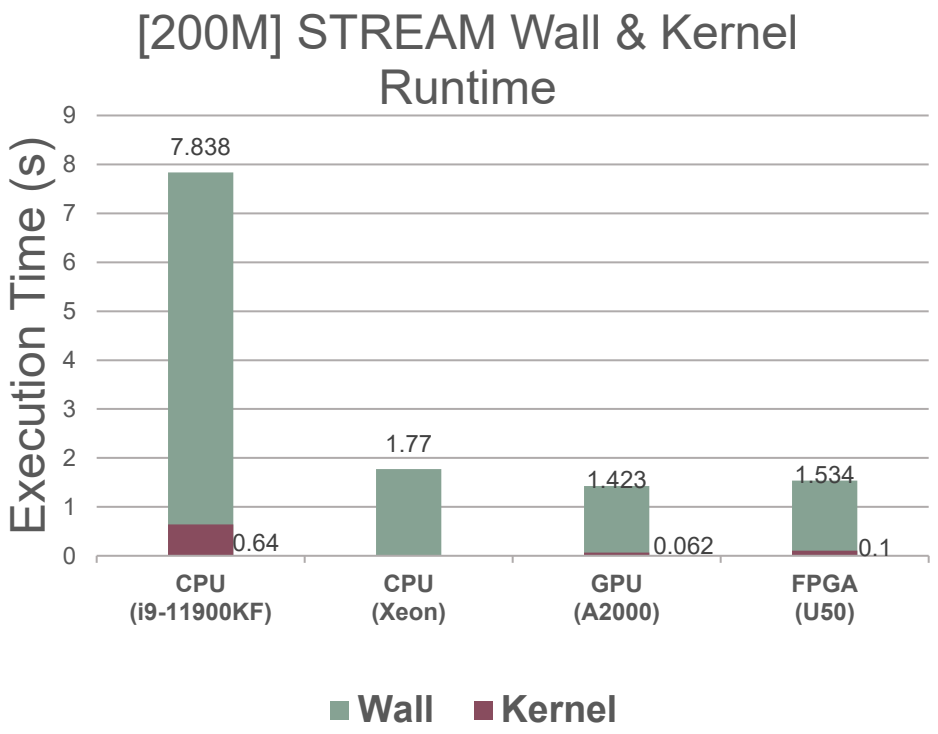




UK Research and Innovation

Results: STREAM





UK Research and Innovation

Conclusions

- Particular Algorithms are more energy efficient on one architecture over the another,
  - exploiting heterogeneity might be an answer to lower energy to solution.
- Significant time is spent optimising FPGA ports compared to both CPU and GPU.
- Greater In-depth architecture knowledge needed for FPGA's over CPU/GPU.

Hardware Clock Details

	SGEMM	FFT	STREAM
CPU: i9-11900KF	3.50 GHz	3.50 GHz	3.50 GHz
CPU: Xeon Gold 6240	2.60 GHz	2.60 GHz	2.60 GHz
GPU: Nvidia A2000	1200 Mhz	1200 Mhz	1200 Mhz
GPU: Nvidia A100	1095 Mhz	1095 Mhz	1095 Mhz
FPGA: Xilinx Alveo U50	Data: 300 Mhz	Data: 300 Mhz	Data: 300 Mhz
FPGA: Alveo U280	Kernel: 300 Mhz	Kernel: 300 Mhz	Kernel: 300 Mhz

Daniel Schien - CARBON-QUANDRI

UKRI Net Zero Digital Research Infrastructure Project

Abstract:

Bio: Daniel is a Senior Lecturer in Computer Science at the University of Bristol. His research focuses on improving our understanding of the environmental impact from information and communication technologies (ICT) and the reduction of such impact. He has pioneered new methods and tools to assess the carbon footprint of digital media which have been applied by major international media companies.

University of BRISTOL

Carbon QuanDRI

Daniel Schien, University of Bristol
Noa Zilberman, University of Oxford
David Greenwood, Newcastle University
Alastair Dewhurst, STFC

UK Research and Innovation

https://net-zero-dri.ceda.ac.uk/cquandri/

/ˈkwɒnd(ə)ri/ - a state of perplexity or uncertainty over what to do in a difficult situation.

NetZero DRI *Services*

- Goal: Carbon-intelligent provision from a service-based perspective
- Why: New management capabilities for NetZero goals
  - Exploit dynamic marginal variability of grid carbon intensity (spatial, temporal and volume)
  - Enable efficiency within providers and consumers through transparency on a service level
- Challenge: Metrics and Methods for service-based assessments are currently missing
  - What flow do we need to measure (physical flow data, I.e. data, electricity) service flow (jobs, API requests, sessions) and value (financial)
  - And how does it translate to carbon over *various time scales* and LCA life cycle phases (data from procurement, operation, decommission)

Evidence for UKRI Net-Zero Strategy

- Case Study - HTC Compute Service
  - Metrics: Compute, Net and Disk I/O, Archival Storage Volume
- Model of Site Electricity Footprint based on HTC Metrics
  - Cooling
  - Storage
  - Network
  - Compute
- Carbon Footprint
- Marginal Carbon Intensity Model to Evaluate Carbon Reductions from Demand Response Mechanisms

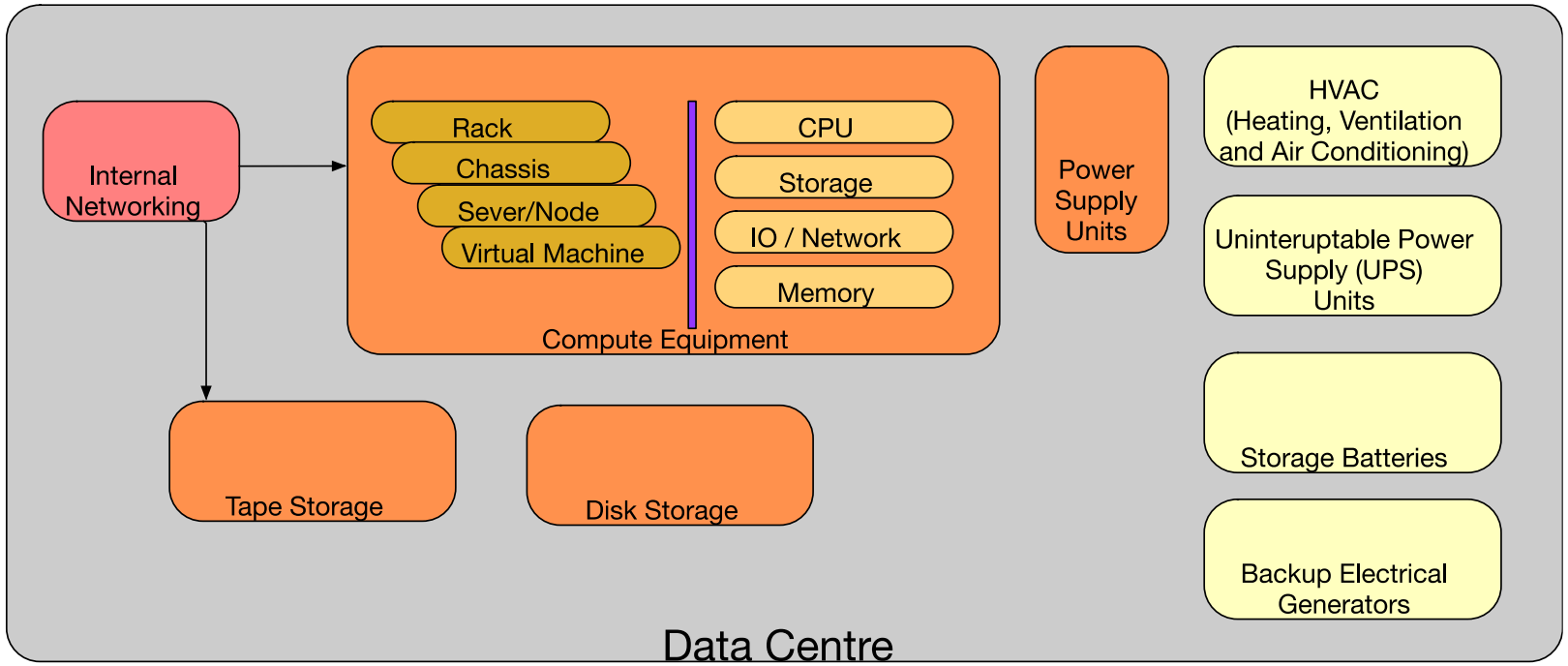
UK Research and Innovation

Approach

- Measurement trace that captures the sustainability aspects in the operation of a DRI service
- Location-based Marginal Carbon intensity model
- Combine to Carbon model of compute to run scenarios

Measurement Trace

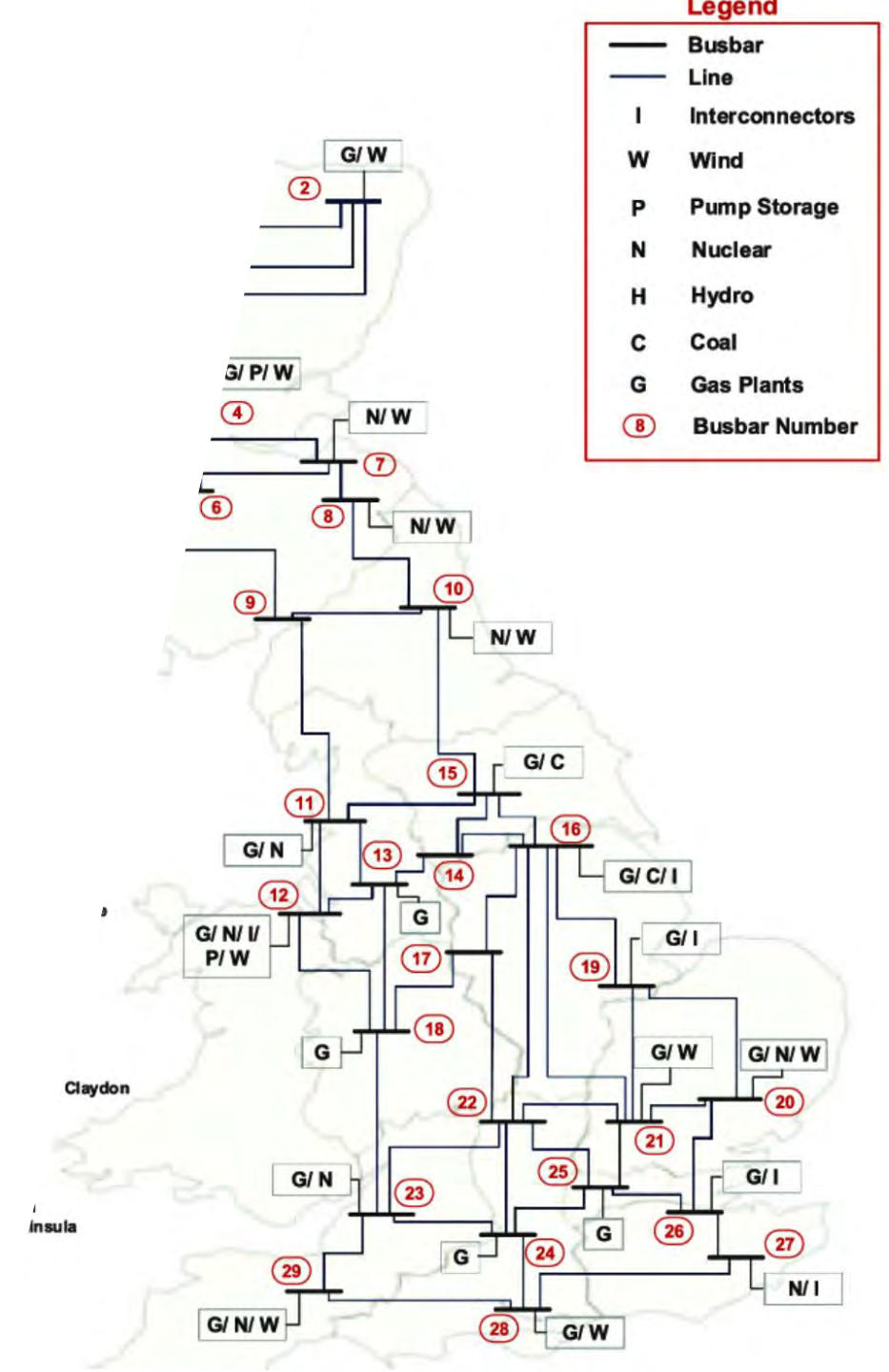
- Collecting, Site, Rack and Node power consumption
- Correlate with service data (jobs)



UK Research and Innovation

Energy Network Carbon Intensity

- 29 bus representation of the GB transmission system
- Optimal power flow model to estimate generation mix at each busbar
- Calculate
  - System Average Carbon Intensity
  - Average Nodal Carbon Intensity
  - Locational Marginal Carbon Intensity



UK Research and Innovation

Carbon Model

- Enable carbon budgets for users
- Trade-off between embodied and use phase carbon
- Evaluate benefit from creating flexible capacity
  - Batteries, overcapacity

Thank You

UK Research and Innovation

UKRI Net Zero Digital Research Infrastructure Project

Wim Vanderbauwhede, Justin O'Byrne, Martin Juckes - Panel Discussion

Bios: Professor Wim Vanderbauwhede is the lead of the Low Carbon and Sustainable Computing activity at the School of Computing Science of the University of Glasgow. He received a PhD in Electrotechnical Engineering from the University of Gent, Belgium in 1996. He has been a lecturer in the School of Computing Science at the University of Glasgow since 2004. His research has resulted in over 150 refereed conference and journal papers as well as several books and book chapters. Before returning to academic research, Prof. Vanderbauwhede worked in the electronics industry as a Design Engineer and Technology R&D Engineer.

Justin works in STFC as an Associate Director within the Programmes Directorate, and also as UKRI's Acting Co-Director for Digital Research Infrastructure (DRI). The DRI programme has £130m to allocate over the next few years and we are extremely conscious of the choices that face us on the net zero front.

Panel Discussion

Chair: Ag Stephens (STFC CEDA)

Panel: Martin Juckes (STFC CEDA)

Wim Vanderbauwhede (Uni of Glasgow)

Justin O'Byrne (UKRI)

UK Research and Innovation

Ilektra Christidi (Senior Research Software Developer, UCL Advanced Research Computing Centre)

Coupling the Time-Warp algorithm with a Kinetic Monte Carlo framework for exact distributed simulations of heterogeneous catalysts

Abstract: Kinetic Monte-Carlo (KMC) simulations have been instrumental in multiscale catalysis studies, enabling the elucidation of the complex dynamics of heterogeneous catalysts and the prediction of macroscopic performance metrics, such as activity and selectivity. However, the accessible length- and time-scales have been a limiting factor in such simulations. I will present a recently established approach for exact, distributed, lattice-based simulations of catalytic kinetics, which couples the Time-Warp algorithm, an optimistic approach to Parallel Discrete Event Simulation, with the Graph-Theoretical KMC framework in the "Zacros" software package, enabling the handling of complex events within large lattices.

The performance of this MPI-based parallel KMC implementation broadly depends on the amount of available memory. Therefore, its performance as a function of parameters that control its memory usage will be presented, as well as its weak and strong scaling benchmark results. Performance improvements of 1-4 orders of magnitude were observed, depending on the simulated chemical system, memory available, parameter choice, and number of MPI processes used. The results of a simulation of a system with 16M sites that exhibits large-scale pattern formation, which was not feasible with the serial algorithm, will also be shown.

Bio: Dr Ilektra Christidi is a particle physicist by calling and research software engineer by profession. She received her PhD from Stony Brook University in NY, for research on the rarest physical process ever observed, a rare decay of charged Kaon particles that provides insight into the matter-antimatter asymmetry of the universe. As a postdoctoral researcher with the ATLAS experiment of the LHC at CERN, she studied background processes to the detection of the Higgs particle that involve lepton pairs, the identification of high-momentum b-jets and measurement of their production rate, the commissioning and data quality assurance of the Muon Spectrometer sub-detector, and the software-level trigger of the Inner Tracker sub-detector.

She moved to software engineering first in industry, developing algorithms and libraries for processing geophysics data for oil and gas searches, and moved to UCL as a research software developer in its central Research IT Services team in 2016. While being a generalist RSE, she mostly focuses on computational projects, often involving HPC. She enjoys working with researchers all over the university to develop robust, sustainable, and performant software to deliver research results, and has so far been involved in various projects, including parallelising simulation software for surface catalysis, single-core optimisation of a particle physics event generator, development of data analysis and visualisation software for yeast genome studies, benchmarking of image analysis codes for radio astronomy, as well as the ExCALIBUR benchmarking project, to name a few.

Coupling the Time-Warp algorithm with a KMC framework for exact distributed simulations of heterogeneous catalysts

Dr Ilektra Christidi

Senior Research Software Developer

UCL Advanced Research Computing Centre

Computing Insight UK, 1-2 December 2022

Outline

- The *Zacros* Kinetic Monte Carlo package
- Time-Warp algorithm
- Performance
    - Parameter tuning
    - Scaling
    - Large system

The *Zacros* team

A long-standing team effort between UCL Chemical Engineering and RITS/ARC


Prof. Michail Stamatakis


Raz Benson


Giannis Savva


Ilektra Christidi


David Stansby


Roland Guichard


Srikanth Ravipati


Miguel Pineda


James Hetherington

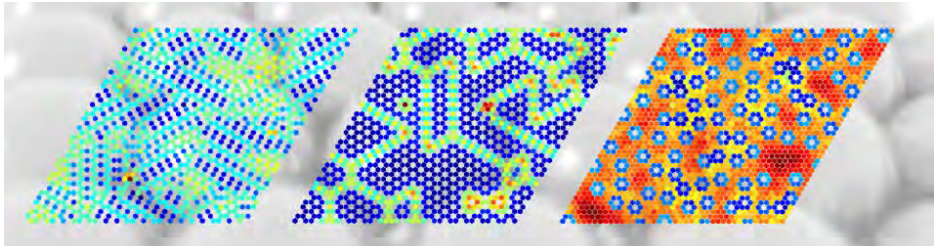

Jens NIelsen


Mayeul D'Avezac

Zacros

Zacros is a Kinetic Monte Carlo package for simulating molecular phenomena on catalytic surfaces.

- **A lattice with sites**
- **A queue of processes (adsorption, desorption, diffusion, reactions)**
  - **Ordered in time based on probability, calculated stochastically based on the energetics of each site**
- **Every KMC step, the most imminent process is executed, lattice occupancies updated, energetics recalculated, and the queue updated**



Limitations

- Large lattices, long KMC times → collective phenomena, eg. pattern formation.
- Need to parallelise, but KMC is serial in nature (which process happens now depends on processes in the past)
- Processes are local, therefore domain decomposition should be possible → need a way to handle processes and interactions across domains
    - The ***Time-Warp*** algorithm

The parallelisation scheme

An *optimistic Parallel Discrete Event Simulation*: keep progressing the KMC time until told otherwise.

- **Minimal synchronization required**

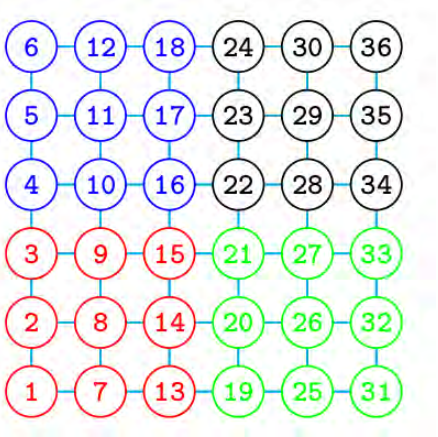
- **Decompose the lattice into subdomains, assign each to an MPI process**
- **Every subdomain progresses independently and holds its own local data structures**
- **Conflicts at the boundaries are communicated via point-to-point *messages* between neighbors asynchronously**
- **Causality violations resolved via a *rollback* procedure**
- **Global communications at fixed wall-clock time intervals to keep track of the *Global Virtual Time* (GVT)**

The Time-Warp algorithm

Domain decomposition

Events that need to be communicated between MPI processes:

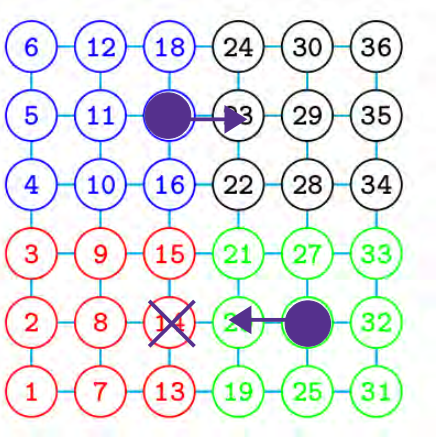
- **Particles across borders**
- **Interactions across borders**



Domain decomposition

Events that need to be communicated between MPI processes:

- **Particles across borders**
- **Interactions across borders**

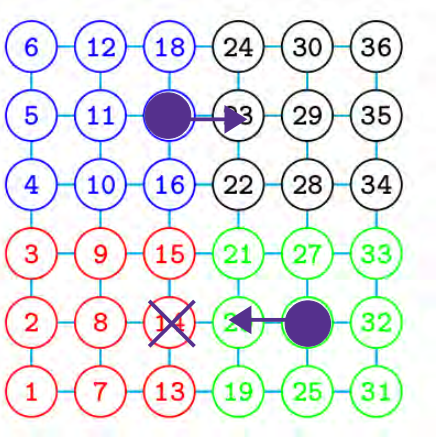


Domain decomposition

Events that need to be communicated between MPI processes:

- **Particles across borders**
- **Interactions across borders**

=> Halos have to be large enough to capture those events



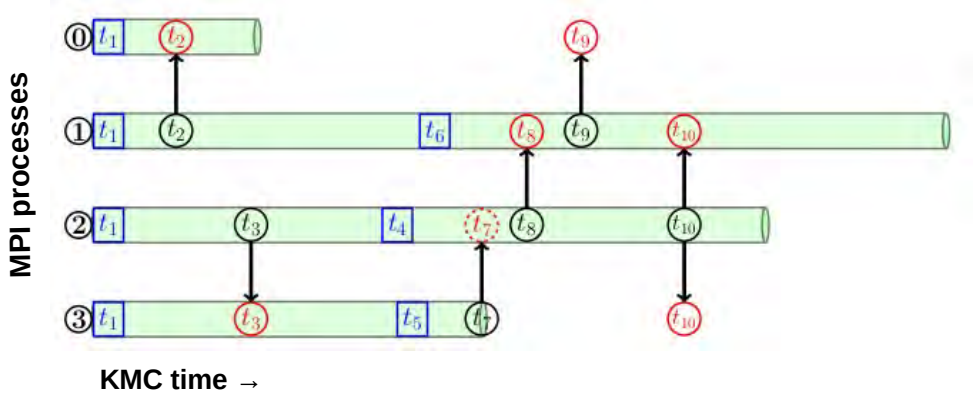
The Time-Warp algorithm

- A *message* between neighboring MPI processes signals the receiver to add the messaged event to its own process queue at the right time.
- At every KMC step, the most imminent process can be a local one or a messaged one.
- If the received message is in the past of the local KMC time, it's a *causality violation*, and a *rollback* has to happen

Data structures:
- Message queue
- State queue

The Time-Warp algorithm

- A *message* between neighboring MPI processes signals the receiver to add the messaged event to its own process queue at the right time.
- At every KMC step, the most imminent process can be a local one or a messaged one.
- If the received message is in the past of the local KMC time, it's a *causality violation*, and a *rollback* has to happen

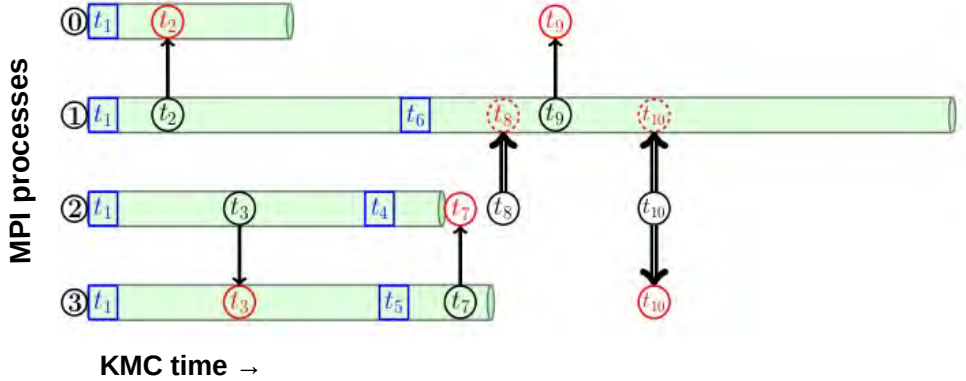


Data structures:
- Message queue
- State queue

t_i – state snapshot at time t_i
t_j – message sent at time t_j
t_j – message received at time t_j
t_j – anti-message sent at time t_j

Rollback

- A *message* between neighboring MPI processes signals the receiver to add the messaged event to its own process queue at the right time.
- At every KMC step, the most imminent process can be a local one or a messaged one.
- If the received message is in the past of the local KMC time, it's a *causality violation*, and a *rollback* has to happen
  - Re-instate a system state previous to the message timestamp



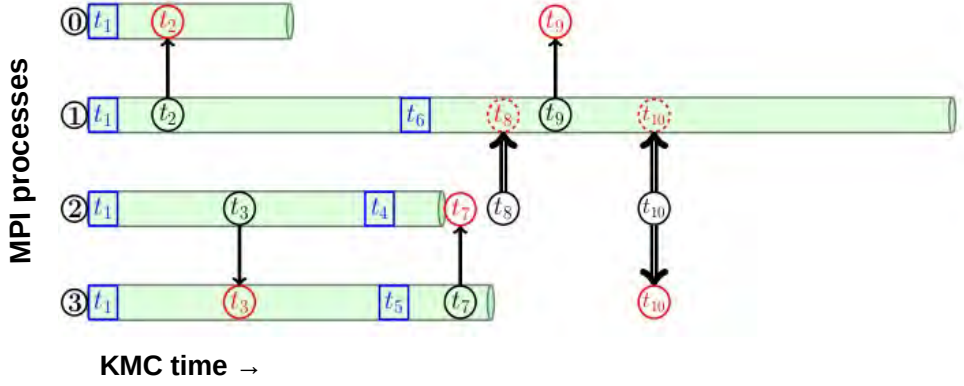
Data structures:
- Message queue
- State queue

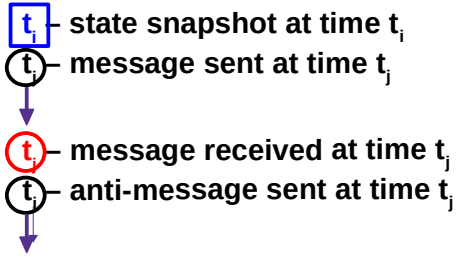
t_i – state snapshot at time t_i
t_j – message sent at time t_j
t_j – message received at time t_j
t_j – anti-message sent at time t_j

Rollback

- A *message* between neighboring MPI processes signals the receiver to add the messaged event to its own process queue at the right time.
- At every KMC step, the most imminent process can be a local one or a messaged one.
- If the received message is in the past of the local KMC time, it's a *causality violation*, and a *rollback* has to happen
  - Re-instate a system state previous to the message timestamp
  - Undo any messages sent to other MPI processes after that time, by sending them *anti-messages*

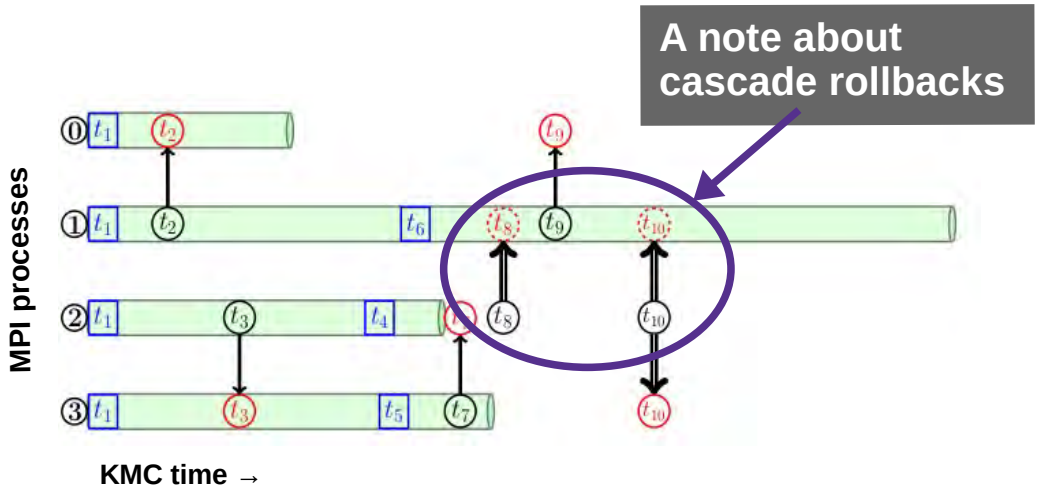
Data structures:
- Message queue
- State queue





Rollback

- A *message* between neighboring MPI processes signals the receiver to add the messaged event to its own process queue at the right time.
- At every KMC step, the most imminent process can be a local one or a messaged one.
- If the received message is in the past of the local KMC time, it's a *causality violation*, and a *rollback* has to happen
  - Re-instate a system state previous to the message timestamp
  - Undo any messages sent to other MPI processes after that time, by sending them *anti-messages*



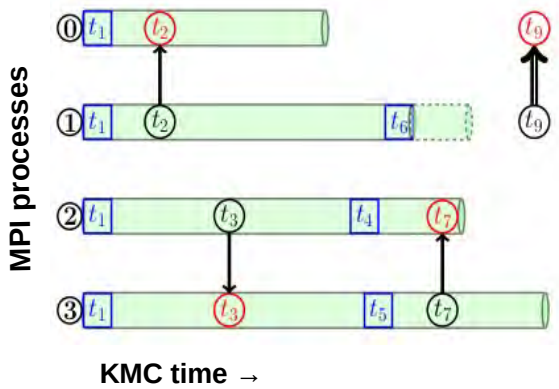
Data structures:
- Message queue
- State queue

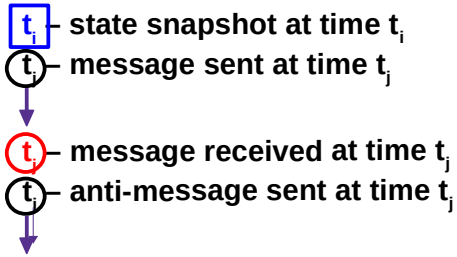
t_i – state snapshot at time t_i
t_j – message sent at time t_j
t_j – message received at time t_j
t_j – anti-message sent at time t_j

Rollback

- A *message* between neighboring MPI processes signals the receiver to add the messaged event to its own process queue at the right time.
- At every KMC step, the most imminent process can be a local one or a messaged one.
- If the received message is in the past of the local KMC time, it's a *causality violation*, and a *rollback* has to happen:
  - Re-instate a system state previous to the message timestamp
  - Undo any messages sent to other MPI processes after that time, by sending them *anti-messages*
  - Re-simulate the history, taking into account the received event

Data structures:
- Message queue
- State queue





Global communication

Need to know the *Global Virtual Time* (GVT) – minimum of KMC times of all MPI processes and the timestamps of any messages in transit

- **Clean up state and message queues from old entries that will not be used anymore (t<GVT)**
- **Decide when the run is over – GVT≥KMCtime(max)**

Collective communication happens at a regular wall-clock time interval.

Algorithm parameters

Time Warp is a memory hungry algorithm – KMC state snapshot queue can get large. Therefore, performance depends on

- **Memory available**
- ***Snapshot-taking interval*** **(δ_{snap}): how often a state snapshot is taken, in # of events**
    - **The more snapshots in the queue, the more efficient the rollbacks**
- ***GVT interval*** **($\Delta\tau_{GVT}$): how often the GVT is calculated, in s**
    - **Memory is cleaned up for re-use after every GVT calculation**

Optimum will be when the benefit of having enough snapshots in the state queue to perform efficient rollbacks, offsets the cost of taking more snapshots and/or global communications

→ Experimented with different state queues for efficient snapshooting:

- ***linked list*** **vs** ***vector***

Systems used for benchmarking

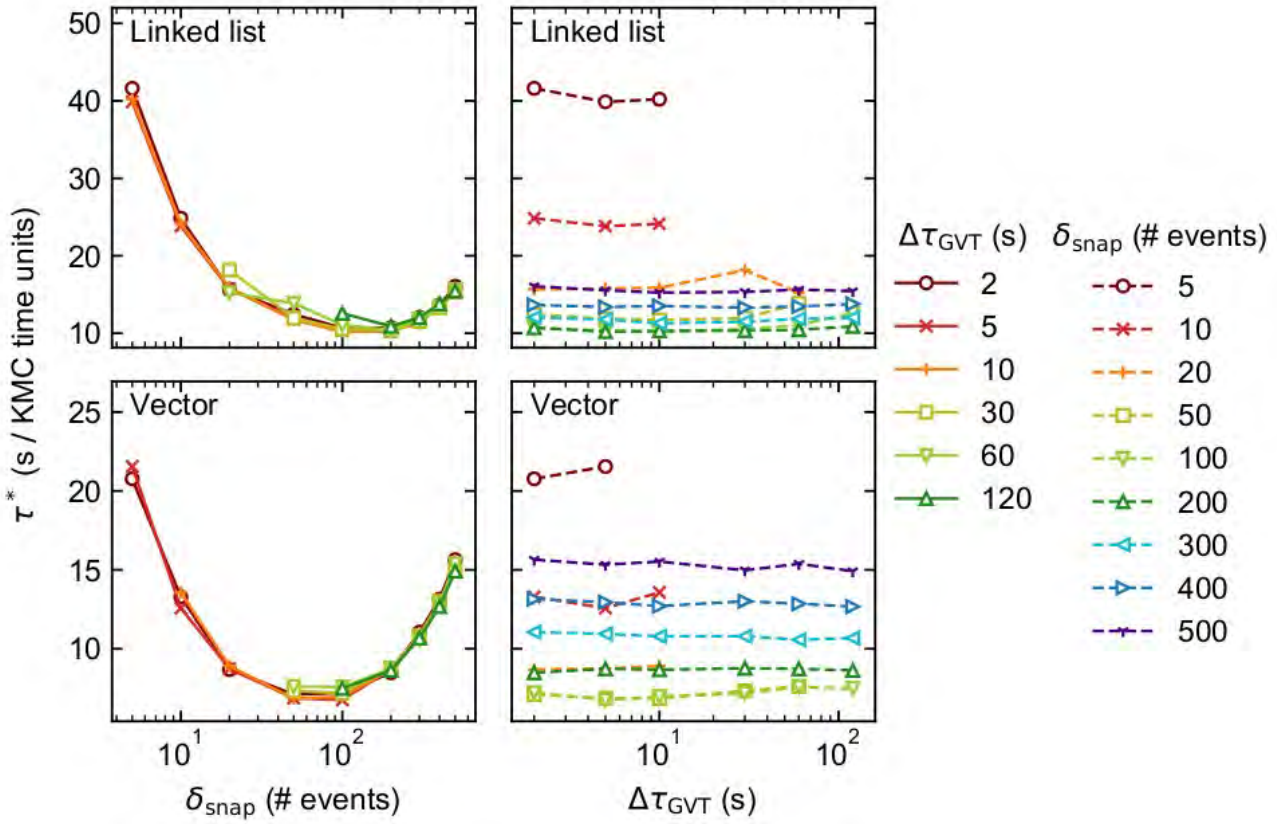
- System 1: adsorption, desorption, diffusion
  - Strong coupling between domains, due to particles crossing boundaries
- System 2: adsorption, desorption, nearest-neighbor lateral interactions
  - Weak coupling between domains, due to energetic clusters crossing boundaries
- System 3: several reactions (complex CO oxidation mechanism) and full energetic cluster expansion
  - Realistic system, with the strongest coupling between domains and large halos

Parameter studies

System 1

200x200 sites

4 MPI processes

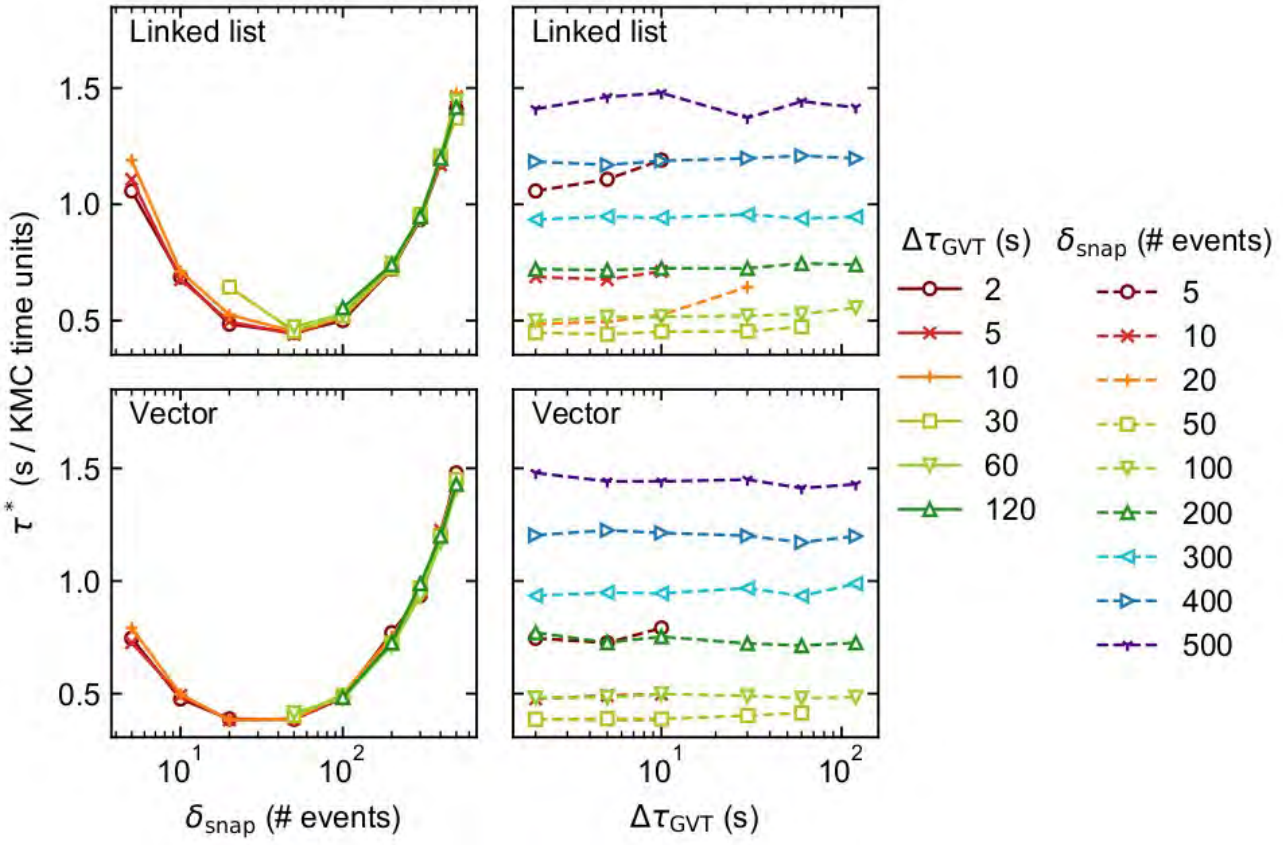


Parameter studies

System 2

200x200 sites

4 MPI processes



Parameter studies - conclusions

- **There is an optimum choice of δ_{snap} and $\Delta\tau_{GVT}$, though it varies by chemical system and number of MPI processes**
- **δ_{snap} affects performance much more than $\Delta\tau_{GVT}$**
- **Type of queue makes a difference**

→ Guidance to users to perform preliminary studies before production runs, to choose optimal parameters

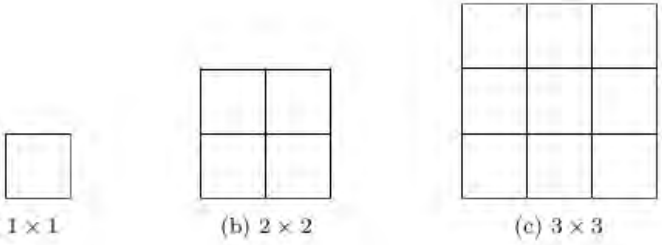
G. Savva, et al (2022). *Large-scale benchmarks of the Time-Warp/Graph-Theoretical Kinetic Monte Carlo approach for distributed on-lattice simulations of catalytic kinetics*. Under review.

Scaling studies

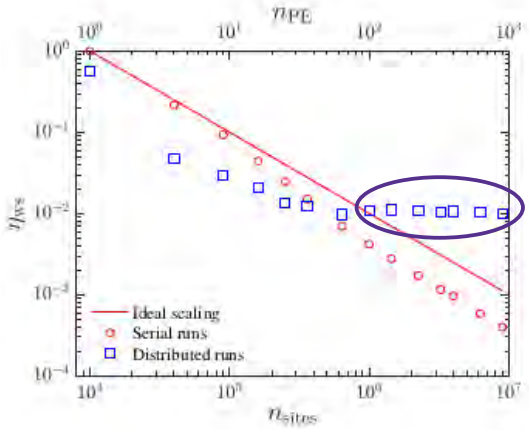
Weak scaling

$$\eta = \frac{t^*\left(n_{\text{sites}}\right)}{t^*\left(n_{\text{sites}}^{min}\right)}$$

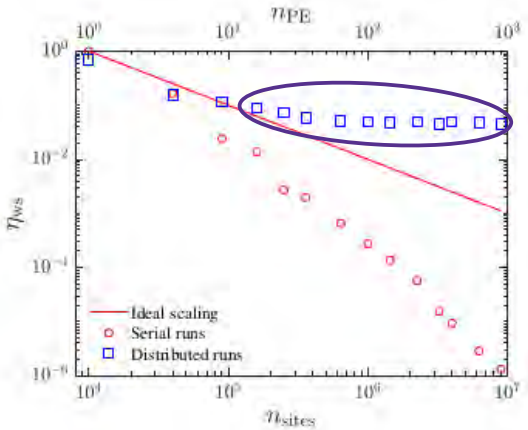
$$\eta_{\text{ws}} = \frac{t^*\left(n_{\text{sites}} : n_{\text{PE}}\right)}{t^*\left(n_{\text{sites}}^{min}\right)}$$



(a) 1×1 (b) 2×2 (c) 3×3



(d) System 1

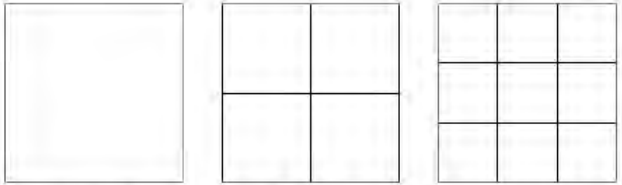


(e) System 2

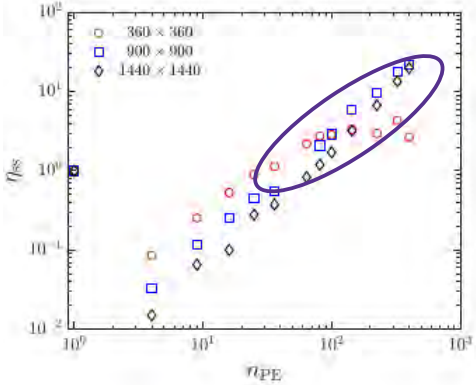
Scaling studies

Strong scaling

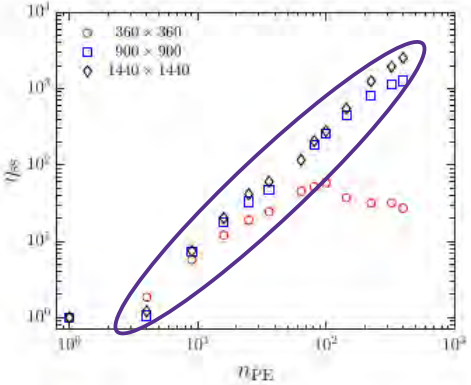
$$\eta_{\mathrm{ss}} = \frac{t^*\left(n_{\mathrm{sites}} : n_{\mathrm{PE}}\right)}{t^*\left(n_{\mathrm{sites}}\right)}$$



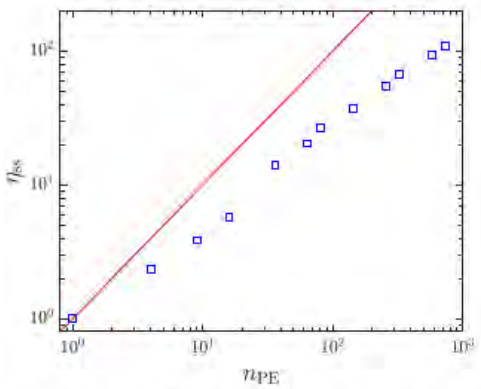
(a) 1×1 (b) 2×2 (c) 3×3



(d) System 1



(e) System 2



(a) System 3 - performance benchmarks

Scaling studies - conclusions

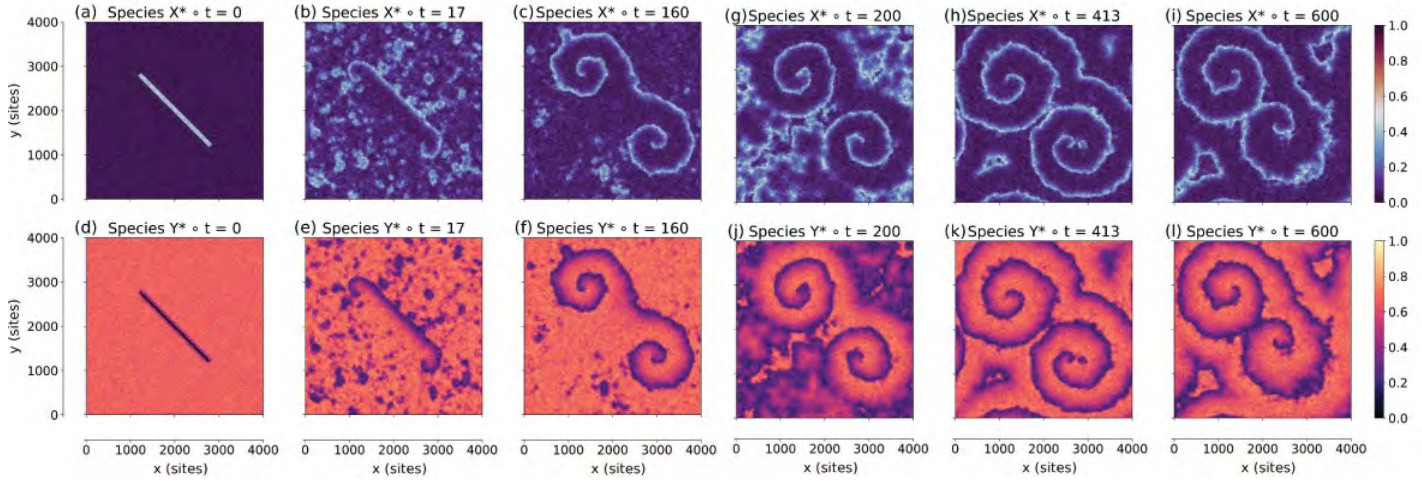
- **The Time Warp algorithm as implemented in Zacros can have significant overheads and memory requirements**
- **For large enough lattices (i.e. minimizing the relative size of halos) it scales well and out-performs serial KMC**

S. Ravipati, et al. (2022). *Coupling the Time-Warp algorithm with the Graph-Theoretical Kinetic Monte Carlo framework for distributed simulations of heterogeneous catalysts*. Comput. Phys. Commun. 270, 108148. (doi: 10.1016/j.cpc.2021.108148)

Performance on large system

Simulated a lattice-based variant of the Brusselator system of 16M sites on Thomas@UCL

- **25x25 = 625 MPI processes → x16 speedup**
- **40x40 = 1600 MPI processes → x36 speedup**
- **Total runtime: 620 KMC s == 38 days (~1.5yr of serial run)**
- **More than 1.6 trillion events*** **(*without the rollbacks)**



G. Savva, et al. (2022). *Exact Distributed Kinetic Monte Carlo Simulations for On-Lattice Chemical Kinetics: Lessons Learnt from Medium- and Large-Scale Benchmarks*. Under review.

Conclusions

- The implementation of Time-Warp into *Zacros* has allowed larger lattices to be simulated for longer KMC times, which was not possible with the serial algorithm
- Optimal parameters found and methodology developed
- Good scaling, depending on simulated system and number of MPI processes
- Large simulation performed successfully
- Further work on memory usage optimisation ongoing

Zacros

Advanced lattice-KMC made easy

For more info https://zacros.org/

Supported by ARCHER-eCSE10-08, ARCHER2-eCSE01-13, Leverhulme Trust RPG-2017-361, EU Horizon 2020 GA: 814416

Facilities used: Thomas@UCL, ARCHER2

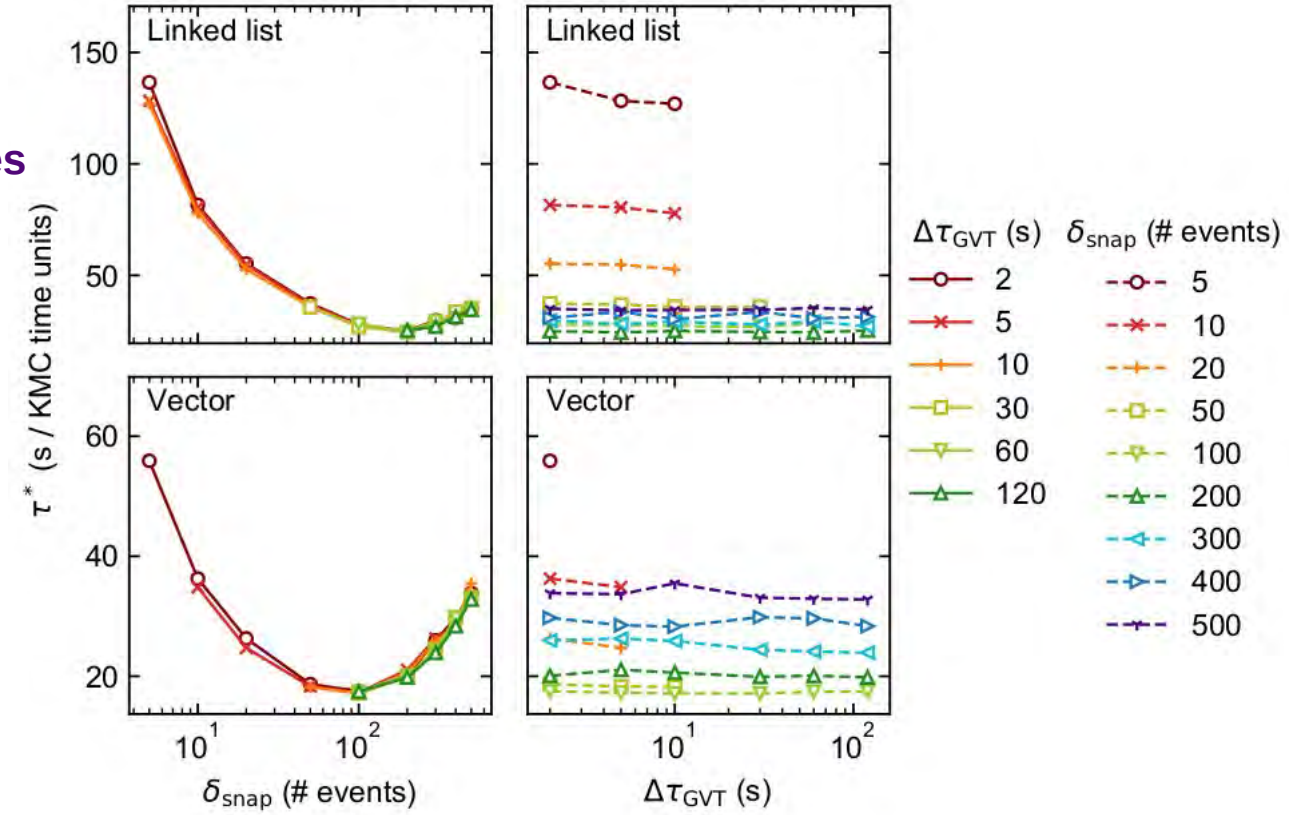
Extras

Parameter studies

System 1

1200x1200 sites

144 MPI processes

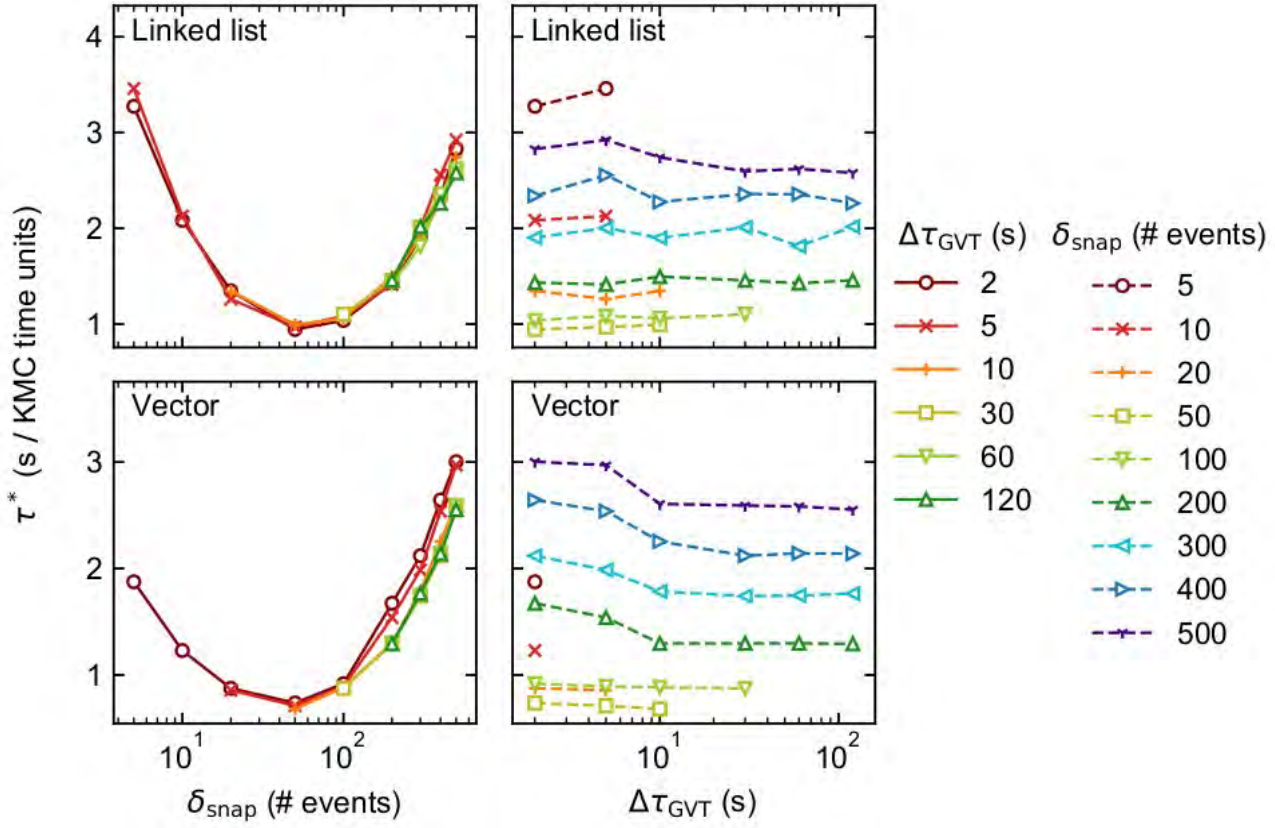


Parameter studies

System 2

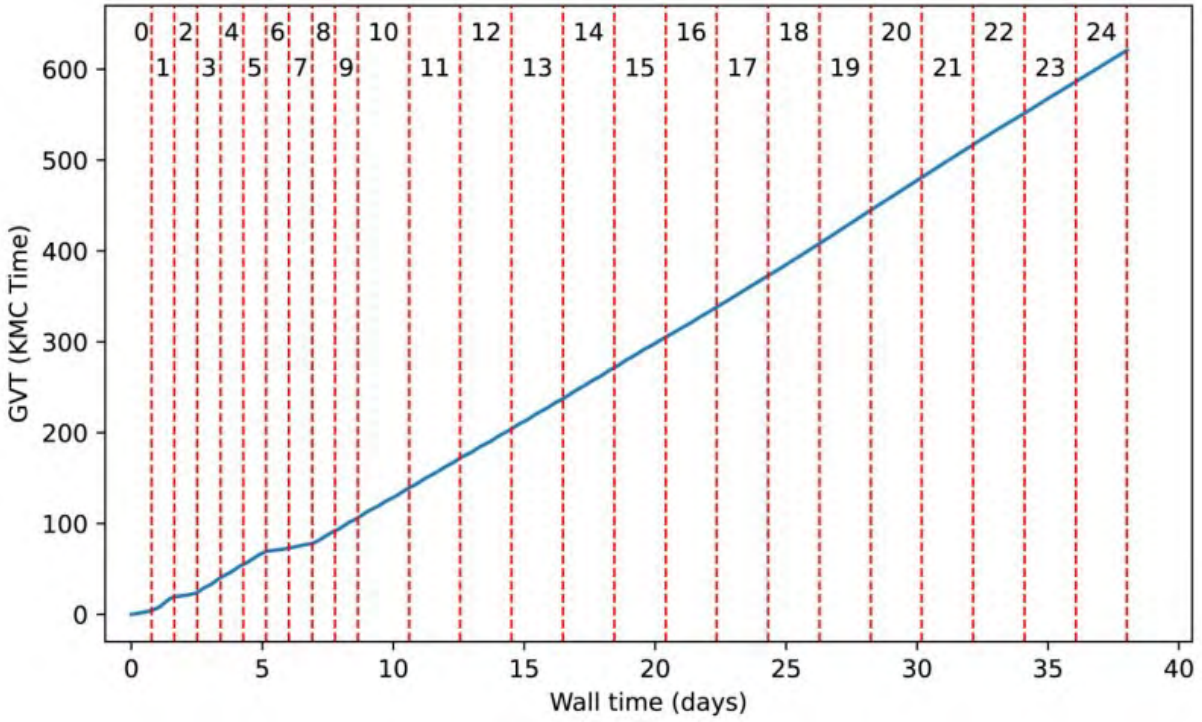
1200x1200 sites

144 MPI processes



Performance on large system

GVT progression in the 625 MPI processes run

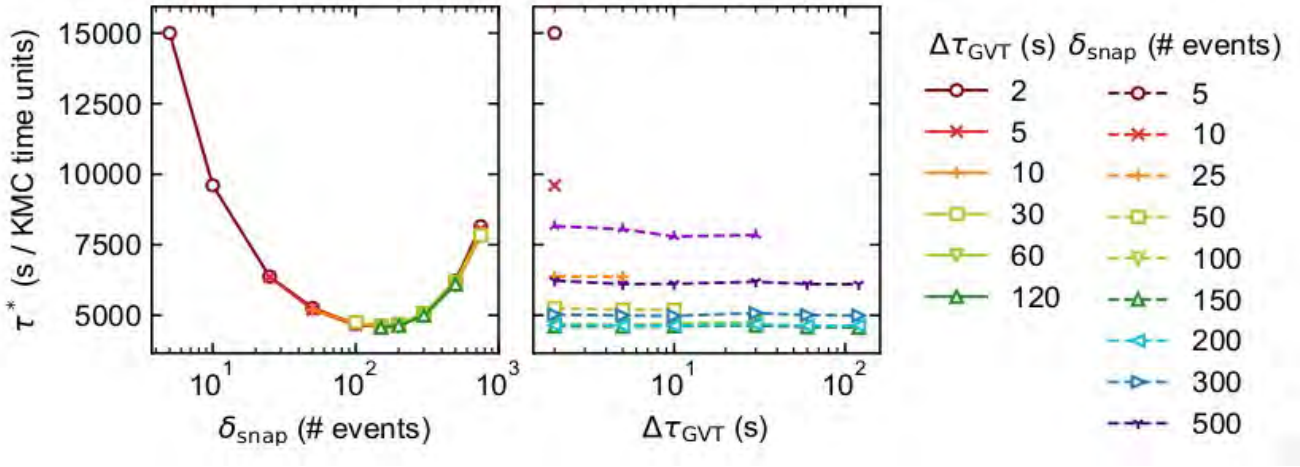


Parameter studies

Brusselator

1200x1200 sites

625 MPI processes



State queue size management

A KMC state snapshot is taken every certain number of events, provided by the user (*snapshot-taking interval*).

A fixed, user-defined amount of memory is available for the state queue throughout the run.

When memory gets filled up, the state queue is *sparsified* (every other stored state is deleted) to free up space, and the snapshot-taking interval is increased (to delay it being filled up again).

Rollbacks are less efficient when the state queue is sparsified, but at least we don’t run out of memory.

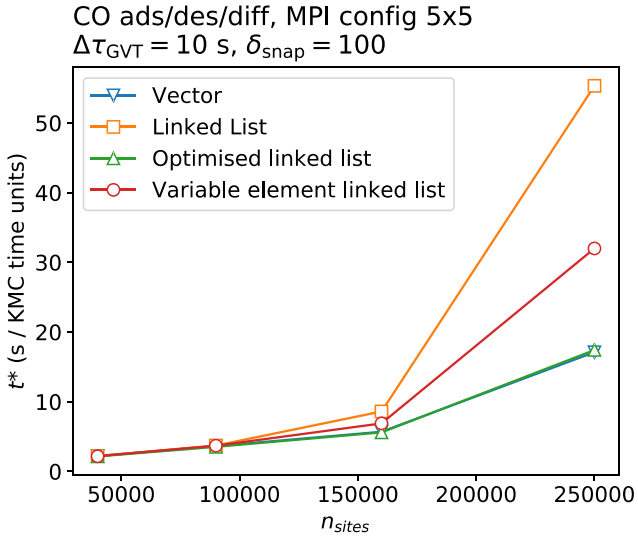
=> Even with sub-optimal parameters chosen, the simulation will not crash, it will just not be as performant

State queue data types

- Doubly-linked list
  - *Pros*: flexible and straightforward to incorporate in algorithm
  - *Cons*: inefficient due to frequent allocations/deallocations of large objects (state snapshots) – every time we take snapshot, rollback, or cleanup
- Vector
  - *Pros*: all allocations happen in the beginning → efficient
  - *Cons*: cannot accommodate variable sized snapshots (so not appropriate for dynamic arrays)
- Optimised doubly-linked list
  - *Pros*: all of doubly-linked list, plus same efficiency as vector – instead of deallocating snapshots at rollback or cleanup, move them to end of queue and re-use them
  - *Cons*: cannot accommodate variable sized snapshots
- Variable-element doubly-linked list
  - *Pros*: same as optimised doubly-linked list, but can also accommodate variable sized snapshots
  - *Cons*: not as efficient as vector/optimised doubly-linked list (some allocations/ deallocations still happen until final size of dynamic arrays is reached)

Further optimisations

- **Investigate memory usage optimisation: implement dynamic arrays**
  - **Preliminary results: memory utilisation for the process queue went up by 25% (from 65% to 80%)**
- **Optimise KMC state snapshot queue: want the flexibility of linked list, but the performance of vector queue**
  - **“optimised linked list” and a variant of it for dynamic arrays inside each state**
  - **Preliminary results: optimised linked list performs as good as vector, variable element linked list performs ok**



Elizabetta Boella (Lancaster University & Cockcroft Institute)

ECsim: a massively parallel Particle-In-Cell code for plasma physics with OpenACC support

Abstract: The Particle-In-Cell (PIC) method is a computational technique used to explore the physics of plasmas at a microscopic level. The plasma is described through a statistical distribution of positive and negative charges sampled via computational particles. These computational particles interact via electromagnetic fields that they produce. These fields are obtained solving Maxwell's equations on a fixed grid, where source terms are computed by interpolating the particles to the grid.

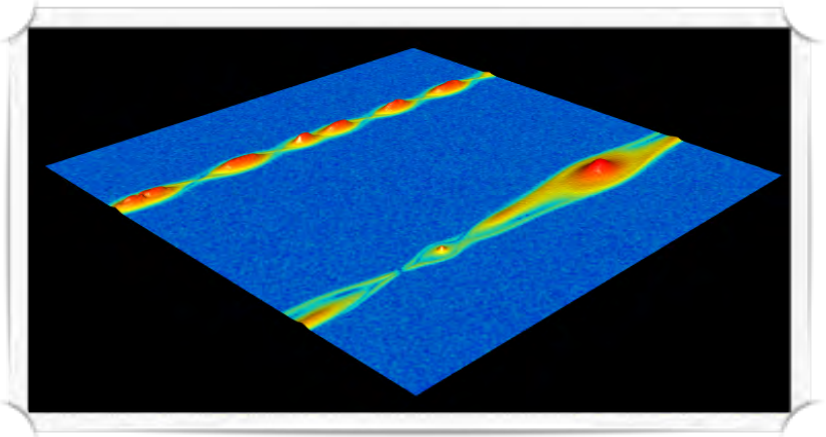
In this talk, we describe our massively parallel PIC code ECsim. We discuss the inclusion of OpenACC directives in the code to port particle kernels to GPUs. For typical numerical parameters used in our simulations, we show that the version of the code that leverages GPUs runs up to 5 time faster than the CPU version. We compare code performance on different generations of NVIDIA GPUs. Finally, we report on scaling tests obtained on different supercomputers.

Bio: Dr. Elisabetta Boella is a lecturer in Physics at Lancaster University & the Cockcroft Institute of Accelerator Science and Technology. She has a long-time experience in the development of kinetic codes for exploring the microphysics of space and laboratory plasmas. She has also extensive expertise in High-Performance-Computing. She is one of the main developers of the plasma code ECsim. Her most recent effort regarding the code concerns the tentative to off-load some of the code calculations to GPU via OpenACC. She is a volunteer for Women in High Performance Computing.

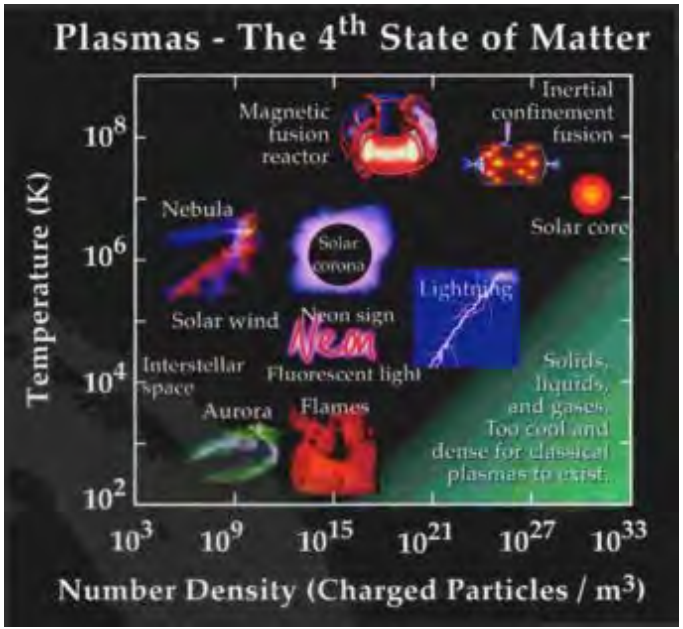
ECsim:
a massively parallel Particle-In-Cell code for plasma physics with OpenACC support

E. Boella[1], M. E. Innocenti[2], M. Bettencourt[3],
M. K. Chimeh[3], G. Lapenta[4], P. Parodi[4],
N. Shukla[5] and F. Spiga[3]

[1] Lancaster University
[2] Ruhr-Universität Bochum
[3] NVIDIA
[4] KU Leuven
[5] CINECA

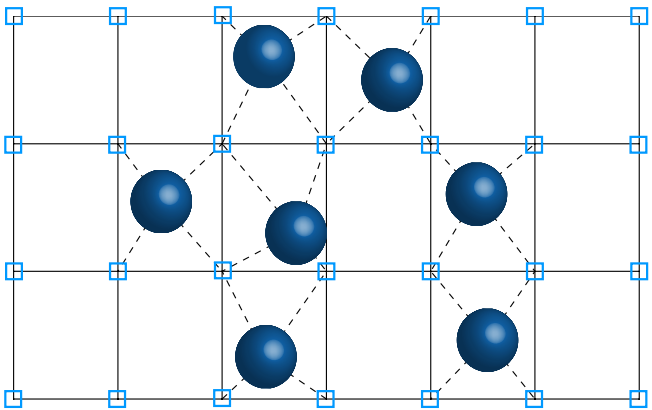


The Particle-In-Cell algorithm models the plasma microphysics



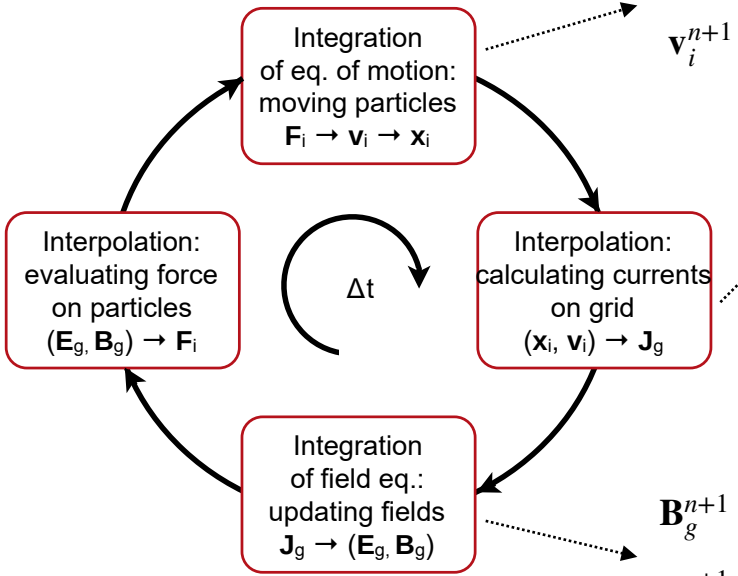
Particle-In-Cell ↔ Particle-Mesh

N_p computational particles, N_g grid cells



Dawson, Rev. Mod. Phys. 55, 403, (1983).

ECsim adopts an implicit discretisation in time for particle and field equations



$$\mathbf{x}_i^{n+1/2} = \mathbf{x}_i^{n-1/2} + \Delta t\, \mathbf{v}_i^n$$

$$\mathbf{v}_i^{n+1} = \mathbf{v}_i^n + \Delta t \frac{q_i}{m_i}\left[\mathbf{E}^{n+\theta}(\mathbf{x}_i^{n+1/2}) + \frac{\overline{\mathbf{v}}_i \times \mathbf{B}^n(\mathbf{x}_i^{n+1/2})}{c}\right]$$

$$\mathbf{J}_g^{n+\theta} = \hat{\mathbf{J}}_g^n + \beta \sum_i q_i \alpha_i \mathbf{E}^{n+\theta}(\mathbf{x}_i^{n+1/2}) S(\mathbf{x}_i^{n+1/2} - \mathbf{x}_g) =$$

$$= \hat{\mathbf{J}}_g^n + \beta \sum_i \sum_{g'} q_i \alpha_i \mathbf{E}_{g'}^{n+\theta} S(\mathbf{x}_i^{n+1/2} - \mathbf{x}_{g'}) S(\mathbf{x}_i^{n+1/2} - \mathbf{x}_g) =$$

$$= \hat{\mathbf{J}}_g^n + \beta \sum_{g'} M_{gg'} \mathbf{E}_{g'}^{n+\theta}$$

$$\mathbf{B}_g^{n+1} = \mathbf{B}_g^n - c\Delta t \nabla_g \times \mathbf{E}^{n+\theta}$$

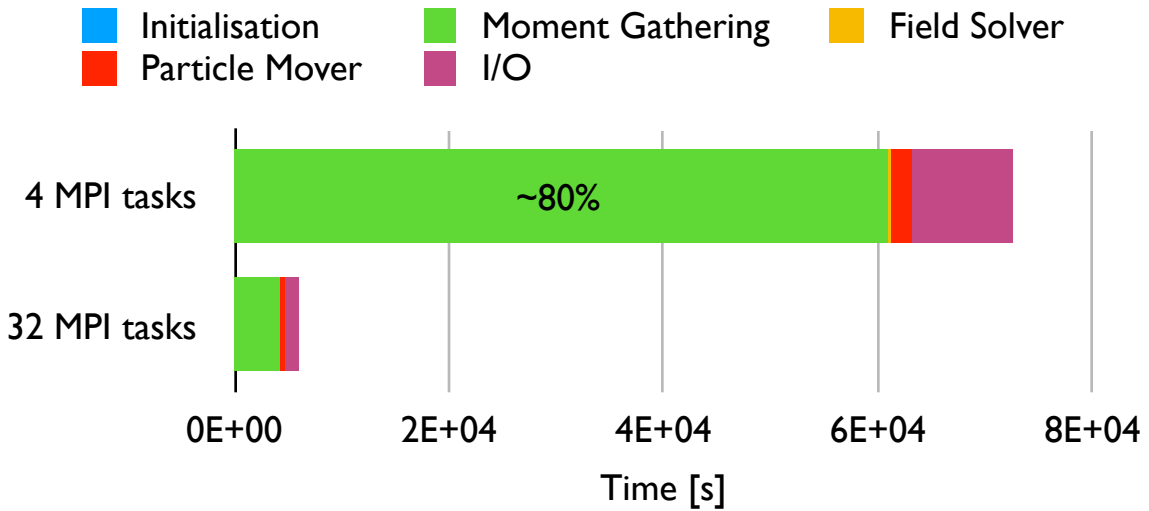
$$\mathbf{E}_g^{n+1} = \mathbf{E}_g^n + c\Delta t \left(\nabla_g \times \mathbf{B}^{n+\theta} - \frac{4\pi}{c} \mathbf{J}_g^{n+\theta}\right)$$

Lapenta et al., J. Plasma Phys. 83, 705830205 (2017). Lapenta, J. Comput. Phys. 334, 349 (2017).
Gonzalez-Herrero et al., Comp. Phys. Commun. 229, 162 (2018).

The moment gathering is the most time consuming portion of the code

ECsim

- Written in C/C++
- Parallelised with MPI
- I/O via HDF5 and H5hut
- Uses PETSc to solve fields
- Built via CMake
- Now includes OpenACC directives



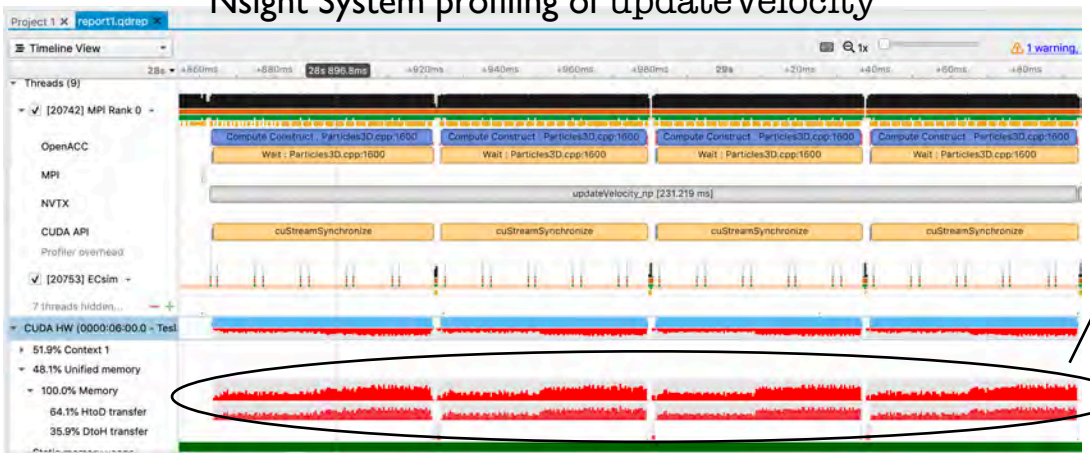
128 x 128 cells, 6400 ppc, 596 iterations
Simulations performed on Marconi100 @CINECA (Italy)
IBM Power9 32 cores/node and 4 NVIDIA V100 GPUs/node

Porting particle mover on GPU is straightforward

- updateVelocity
- updatePosition
- fixPosition

```
#pragma acc parallel loop

for (long long rest = 0; rest < nop; rest++) {

...

}
```

Nsight System profiling of updateVelocity



To solve this

```
cudaMemPrefetchAsync
(x, sizeof(double)*nop, 0, 0);
```

destination device GPU

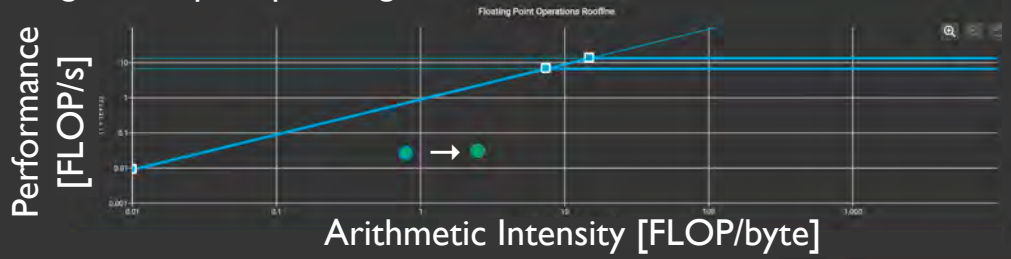
Moment gathering requires atomic operations to avoid race condition

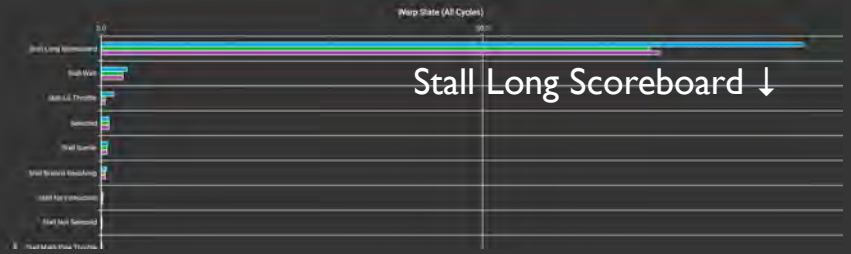
- computeMoments (Most time consuming routine of the code)

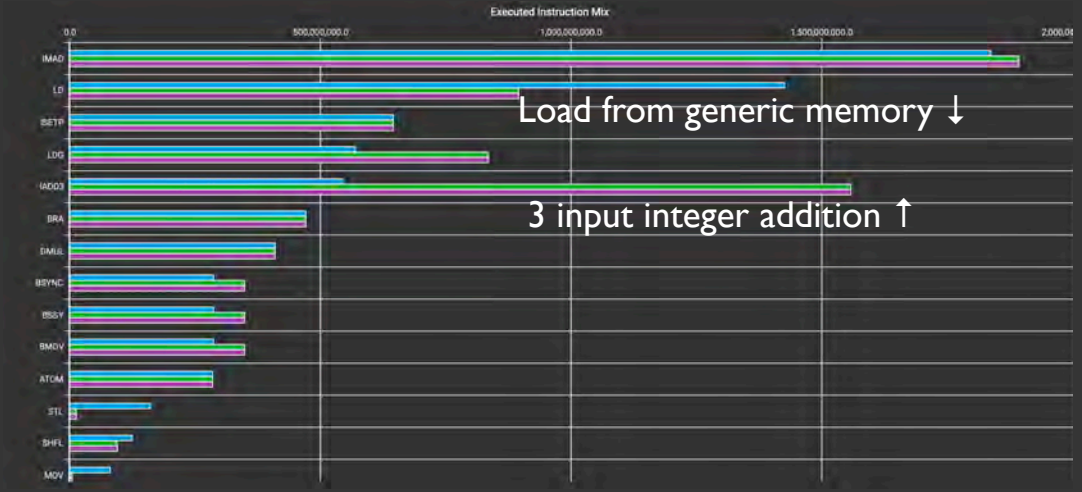
```
void EMfields3D::addRho(double weight[][2][2], int X, int Y, int Z, int is) {
 for (int i = 0; i < 2; i++)
  for (int j = 0; j < 2; j++)
   for (int k = 0; k < 2; k++) {
    const double temp = weight[i][j][k];

    #pragma acc atomic update
    rhons[is][X - i][Y - j][Z - k] += temp * invVOLn[X - i][Y - j][Z - k];
   }
}
```

We managed to improve computeMoments by increasing data locality

Nsight Compute profiling of computeMoments



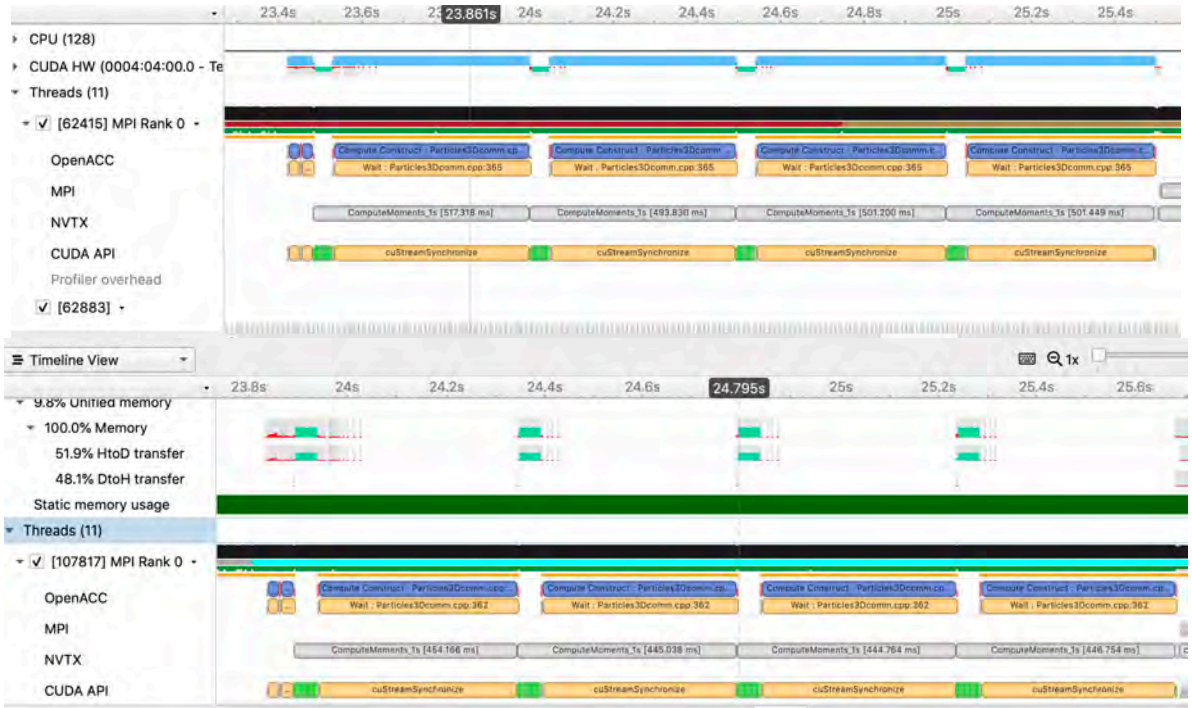




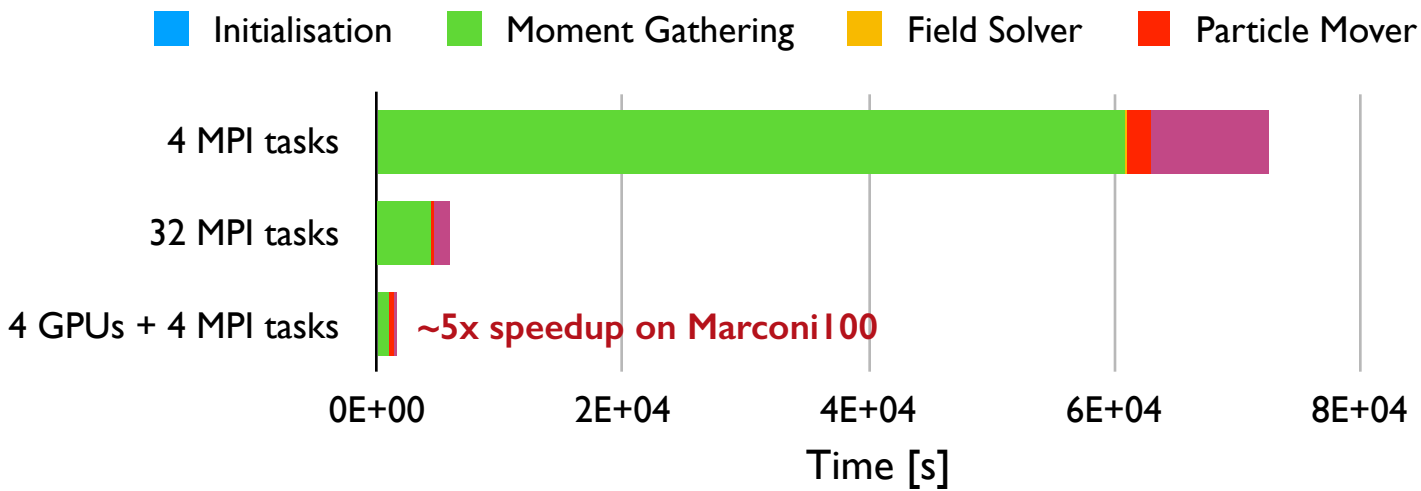
- In our simulation test, we reduced the total execution time from 1511.82 s to 1496.33 s
- We tried to go further and unrolled loops in the routine to increase data locality, but we did not get any time improvement because we increased too much the number of registers

A fine tuning optimisation led to a 12% speed up of the computeMoments kernel

Nsight System profiling of computeMoments



By offloading to GPUs the particle kernels, we achieved a 5x speedup



128 x 128 cells, 6400 ppc, 596 iterations

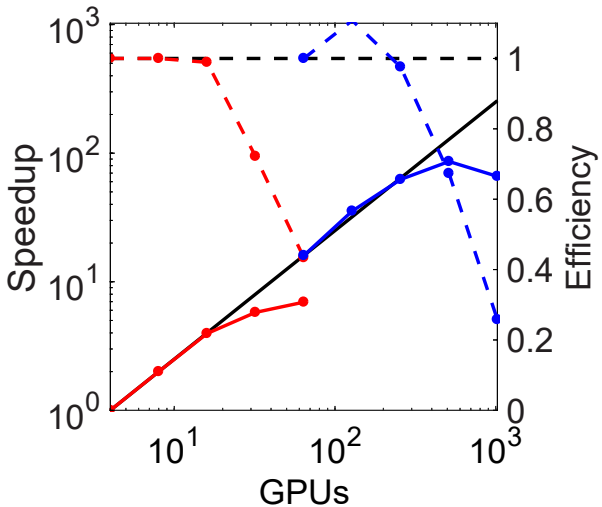
Simulations performed on Marconi100 @CINECA (Italy)

IBM Power9 32 cores/node and 4 NVIDIA V100 GPUs/node

Weak scaling shows an efficiency of 80% up to 1024 GPUs on Marconi100

Strong scaling

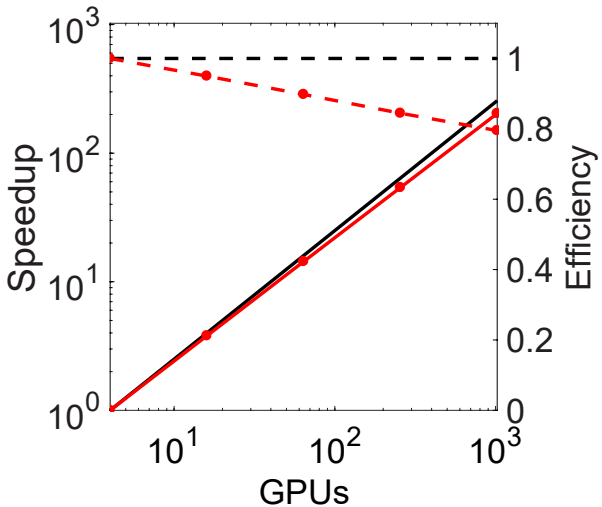
64 x 64 x 32 cells with 1296 ppc



72% efficiency up to 32 GPUs

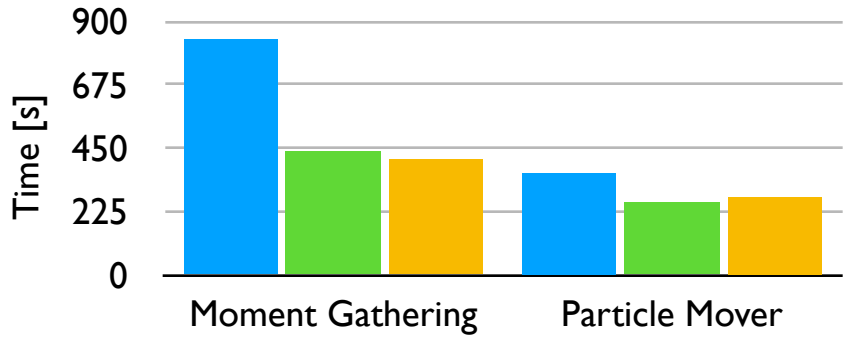
Weak scaling

128 x 128 to 2048 x 2048 cells with 6400 ppc



80% efficiency up to 1024 GPUs

On A100, Moment Gathering becomes twice as fast as on V100



Marcon100: IBM Power9 32 cores/node and 4 NVIDIA V100 GPUs/node

Juwels Booster: AMD EPYC 7402 48 cores/node and 4 NVIDIA A100 GPUs/node

MeluXina: AMD EPYC 7452 32 cores/node and 4 NVIDIA A100 GPUs/node

Summary and perspectives

The most consuming portion of ECsim on CPU is the moment gathering where particles are deposited onto the grid (~80% of the execution time).

By offloading only particle routines to GPU, a speedup of 5x was achieved.

ECsim shows an efficiency of 80% in weak scaling test up to 1024 GPUs.

Next step: porting the field solver to GPU.

This work was partially performed in the framework of the CSCS and CINECA OpenACC Hackathons. We gratefully acknowledge access to Marconi100 via ISCRA and HPCEuropa3, Juwels Booster via GCS and PRACE and MeluXina via EuroHPC.

The Jacky Pallas Memorial Presentation

Dr Djenifer Kappel (Centre for Neuropsychiatric Genetics and Genomics - Cardiff University)

The genomic basis for precision medicine in treatment-resistant schizophrenia

Abstract: Mental illness is currently the main worldwide driver of health problems and disability, and one of the main challenges in adequately addressing and treating mental illnesses is the fact that the most commonly prescribed drugs are not equally effective for everyone. Genetic differences between individuals are known to contribute to how one responds to pharmacological treatment, but few guidelines in implementing this knowledge exist. My research seeks to explore how to use genomic information to make psychiatric treatments, particularly antipsychotics, more beneficial for everyone who needs them. As part of Cardiff University's CLOZUK project, I have accessed genomic and clinical data from thousands of individuals with schizophrenia that take an antipsychotic called clozapine. Clozapine is particularly effective in treating this condition, however due to a range of potentially severe adverse effects, it is currently only employed when other treatments have failed and not everyone eligible gets access to it. To identify avenues for a safer and more efficient way to use clozapine, my work has leveraged clinical treatment records on over 4000 CLOZUK samples. Some insights from this dataset involve the discovery of genetic variants associated to clozapine metabolism, and the establishment of a metric of genetic predisposition to schizophrenia (a construct called a “polygenic risk score”) as a marker of individuals receiving higher doses of the drug than commonly prescribed. These results suggest that the genomics-aware healthcare of the near future might realise the personalisation of medication doses, a process which can currently take months and ultimately relies on trial-and-error procedures, even in conditions as complex as psychiatric disorders.

Bio: Dr. Djenifer Kappel is a Brazilian Early Career Researcher currently working as a Post-doctoral Research Associate at Cardiff University with Dr. Antonio Pardiñas. She initially graduated in Biomedical Science at Federal University of Rio Grande do Sul in Brazil, later obtaining a Master’s and a PhD in Human Genetics (2020) at the same institution. Over the last 10 years she’s been interested in the biological underpinnings of mental illness and the use of bioinformatics and statistical genetics in their discovery. Her current research is based on understanding how genetics can predispose us to psychiatric disorders and impact on their treatment and management.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD
Centre for Neuropsychiatric Genetics and Genomics
Canolfan Geneteg a Genomeg Niwroseiciatrig
The Academy of Medical Sciences

THE GENOMIC BASIS FOR PRECISION MEDICINE IN TREATMENT-RESISTANT SCHIZOPHRENIA

Dr. Djenifer B. Kappel

If the 20th century was the century of physics, the 21st will be the century of biology.

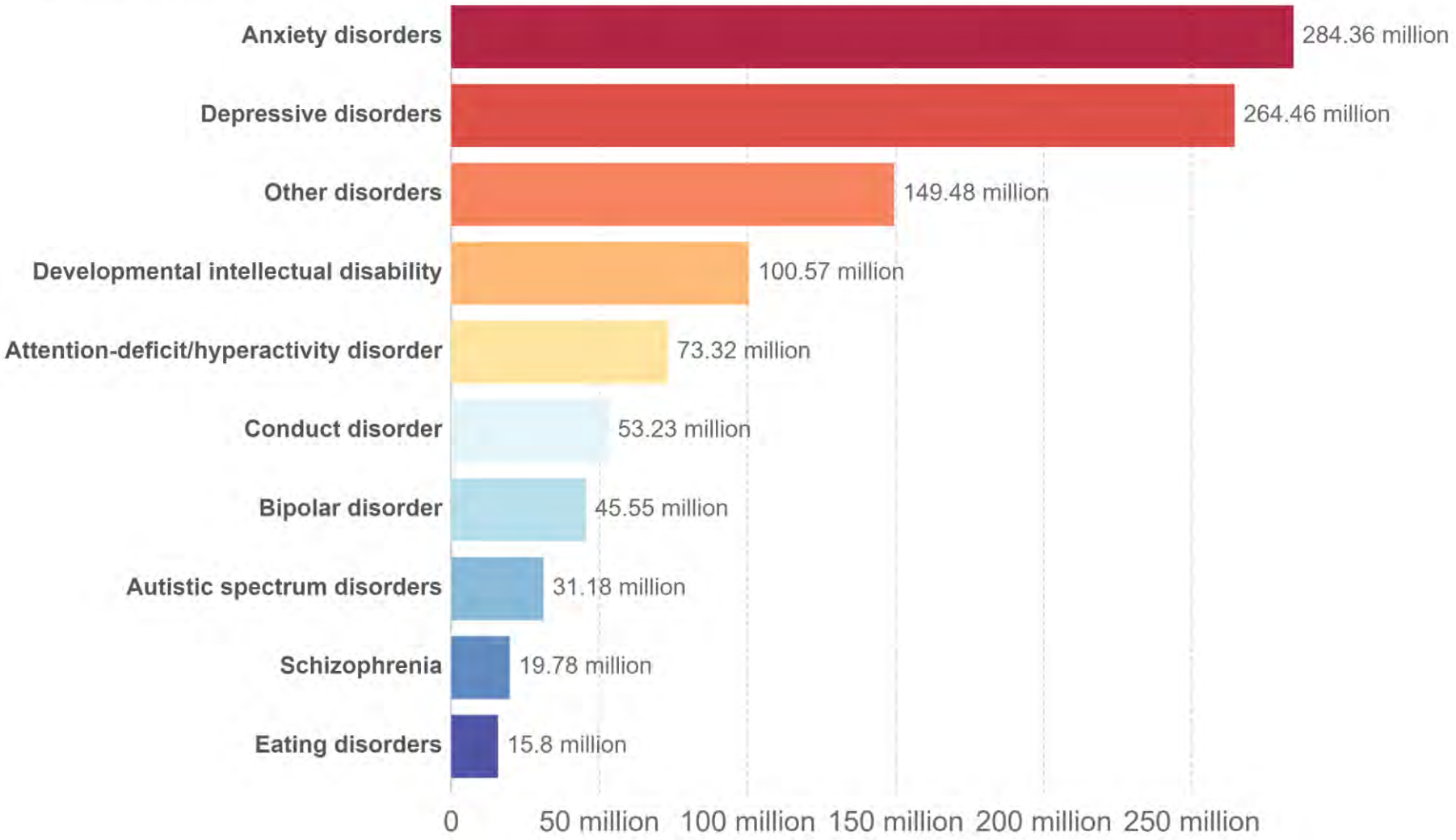
- Craig Venter

Number with a mental or neurodevelopmental disorder by type, World, 2017

Substance use disorders are not included. Figures attempt to provide a true estimate (going beyond reported diagnosis) of prevalence based on medical, epidemiological data, surveys and meta-regression modelling.

Our World in Data

#breakthestigma

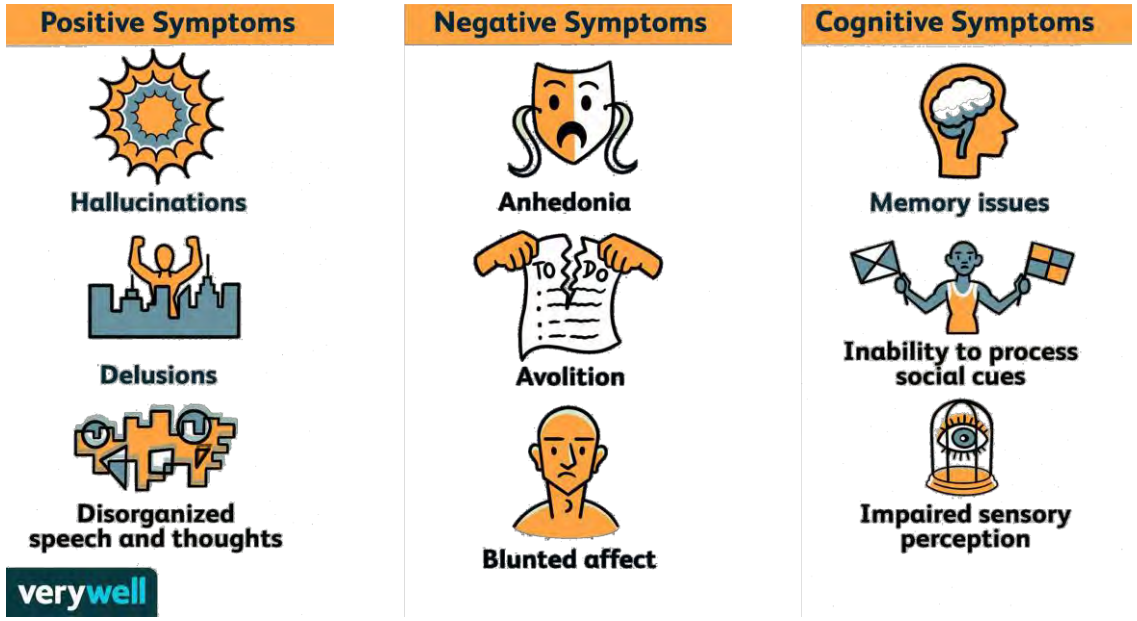


Source: IHME, Global Burden of Disease

Centre for Neuropsychiatric Genetics and Genomics
Canolfan Geneteg a Genomeg Niwroseiciatrig
The Academy of Medical Sciences

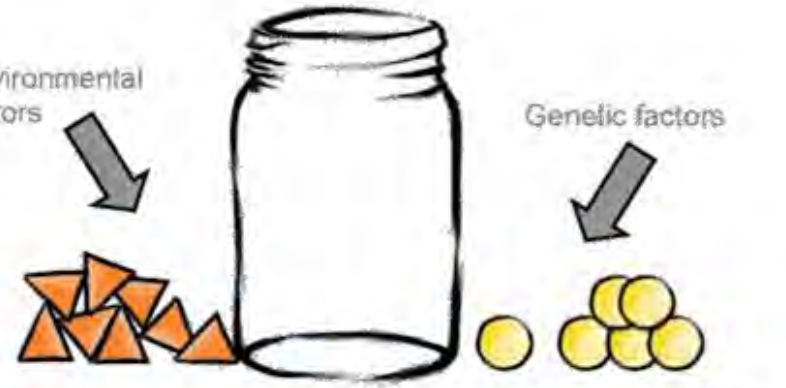
WHAT IS SCHIZOPHRENIA?

Getty Images



Dr. Jehannine Austin

Mental illness jars



TPGG

Development of mental illness

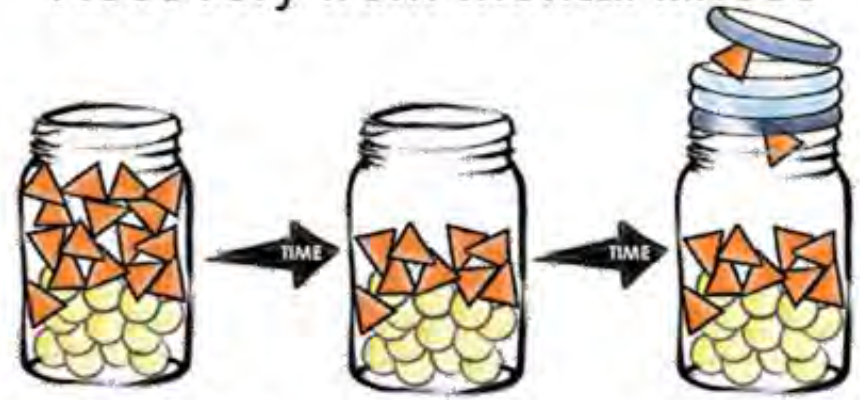
Unaffected

More vulnerable, but not experiencing episode of illness

Experiencing active episode of illness

TPGG

Recovery from mental illness



Experiencing an active episode of illness

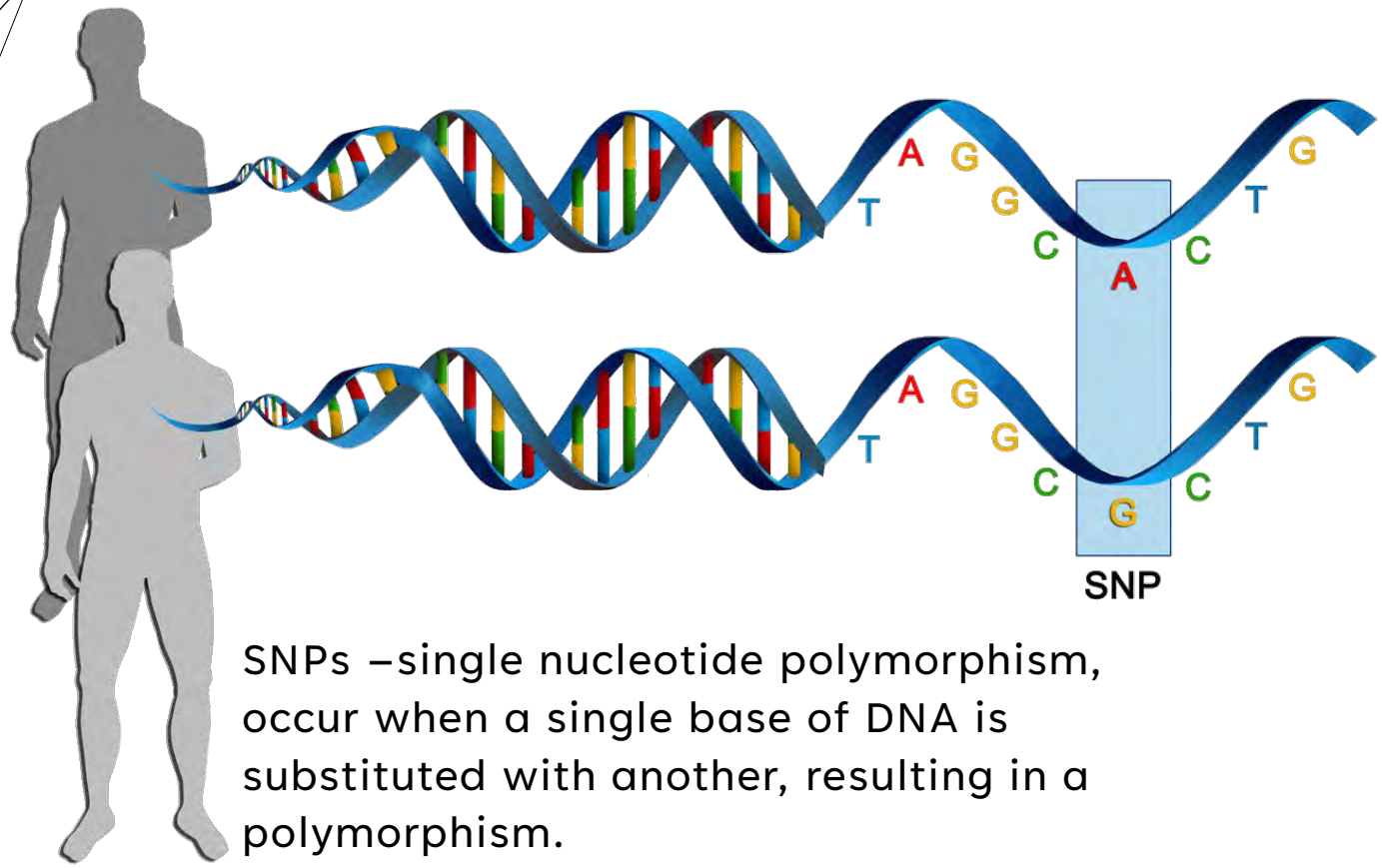
Vulnerable, but not actively ill

Less likely to relapse

Protective factors

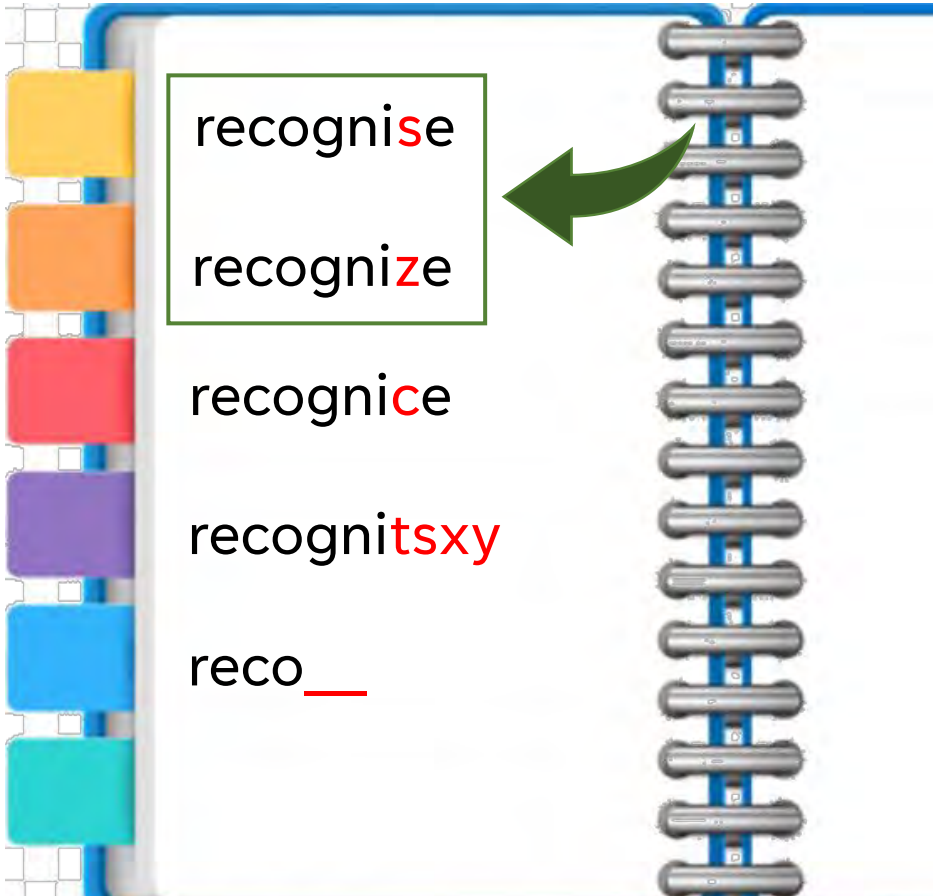
TPGG

GENETIC VARIATION IS ASSOCIATED WITH RISK TO PSYCHIATRIC CONDITIONS

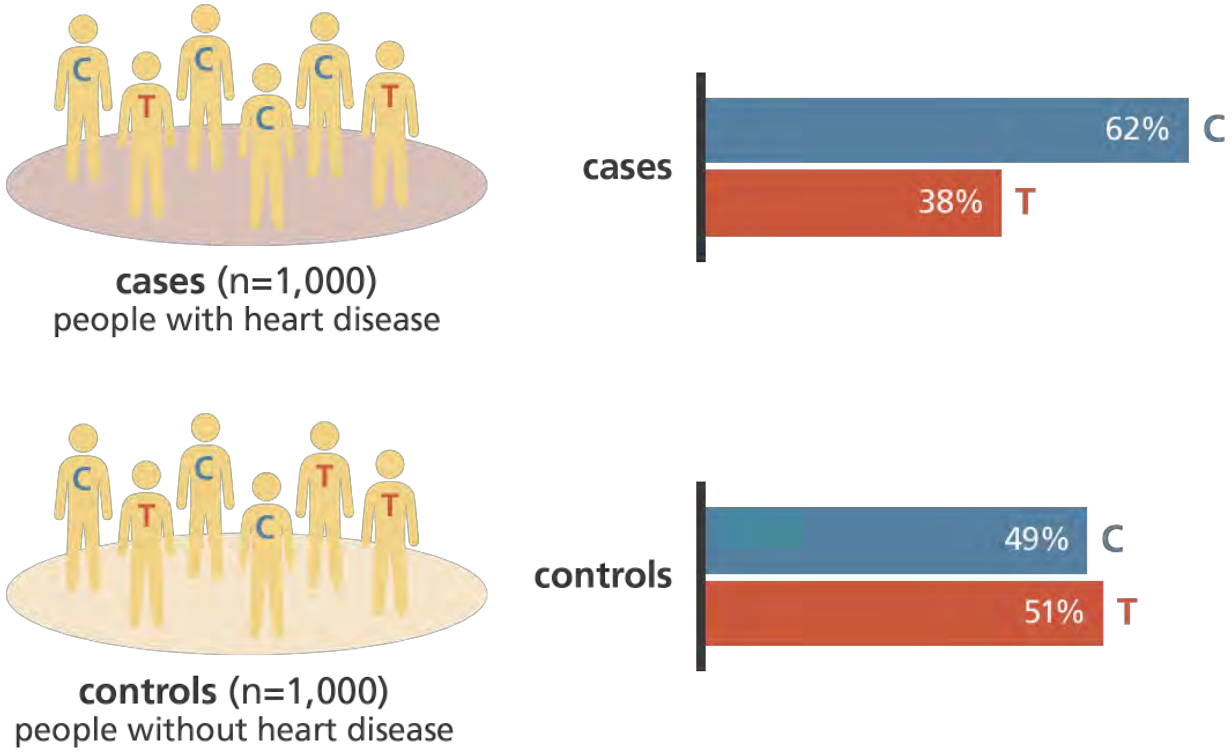


SNPs –single nucleotide polymorphism, occur when a single base of DNA is substituted with another, resulting in a polymorphism.

Each person carries 4-5 million SNPs



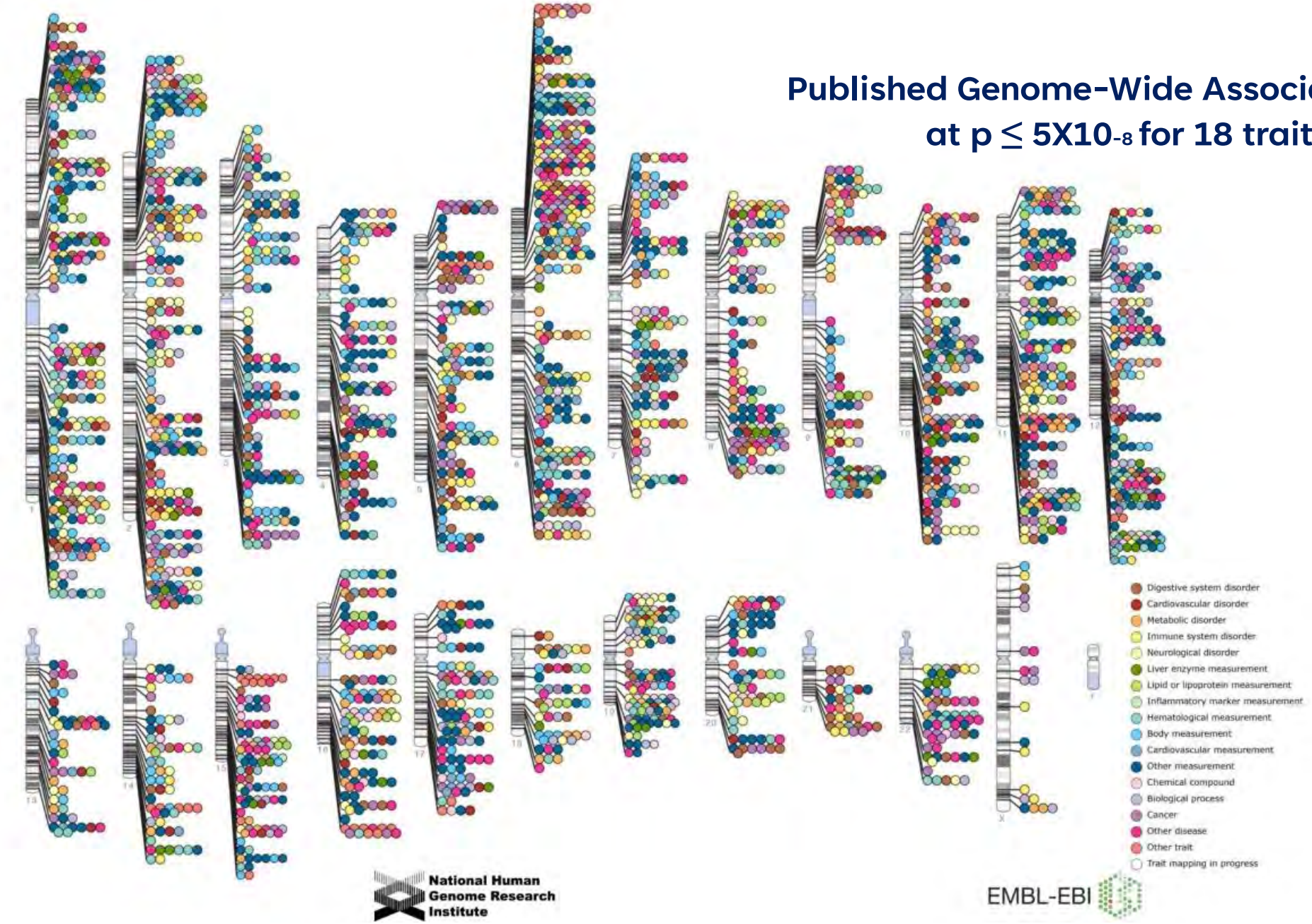
GENOME-WIDE ASSOCIATION STUDIES (GWAS)



- In GWAS we usually analyze single nucleotide polymorphisms (SNPs)
- GWAS is an association study design that aims to capture as much common SNP variation across the genome as possible

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD
Centre for Neuropsychiatric Genetics and Genomics
Canolfan Geneteg a Genomeg Niwroseiciatrig
The Academy of Medical Sciences

Published Genome-Wide Associations through 2012 at $p \leq 5X10^{-8}$ for 18 trait categories



National Human Genome Research Institute

EMBL-EBI

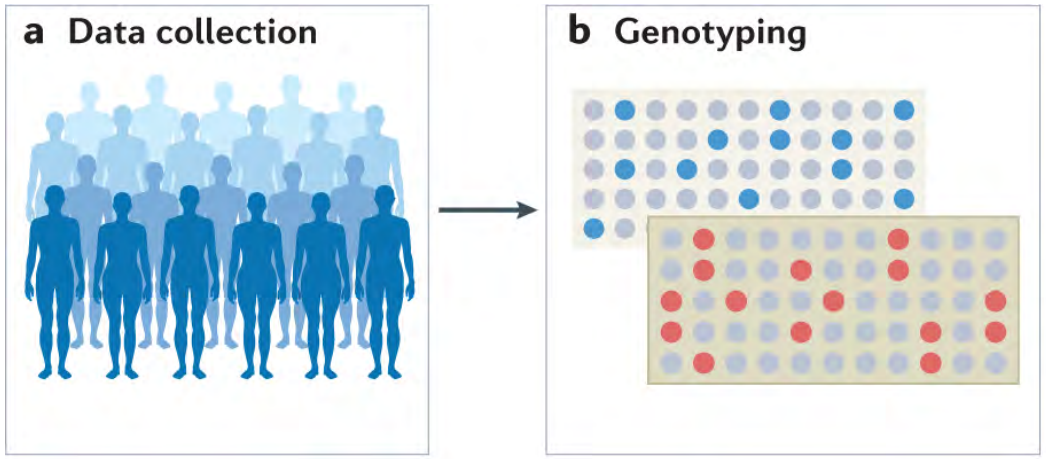
As of 2022, the GWAS Catalog contains
6096 publications and 434351 associations

MAPPING THE SCHIZOPHRENIA GENOME

Dr. Antonio Pardiñas

Nature. 2022 Apr;604(7906):502-508. doi: 10.1038/s41586-022-04434-5. Epub 2022 Apr 8.

Mapping genomic loci implicates genes and synaptic biology in schizophrenia

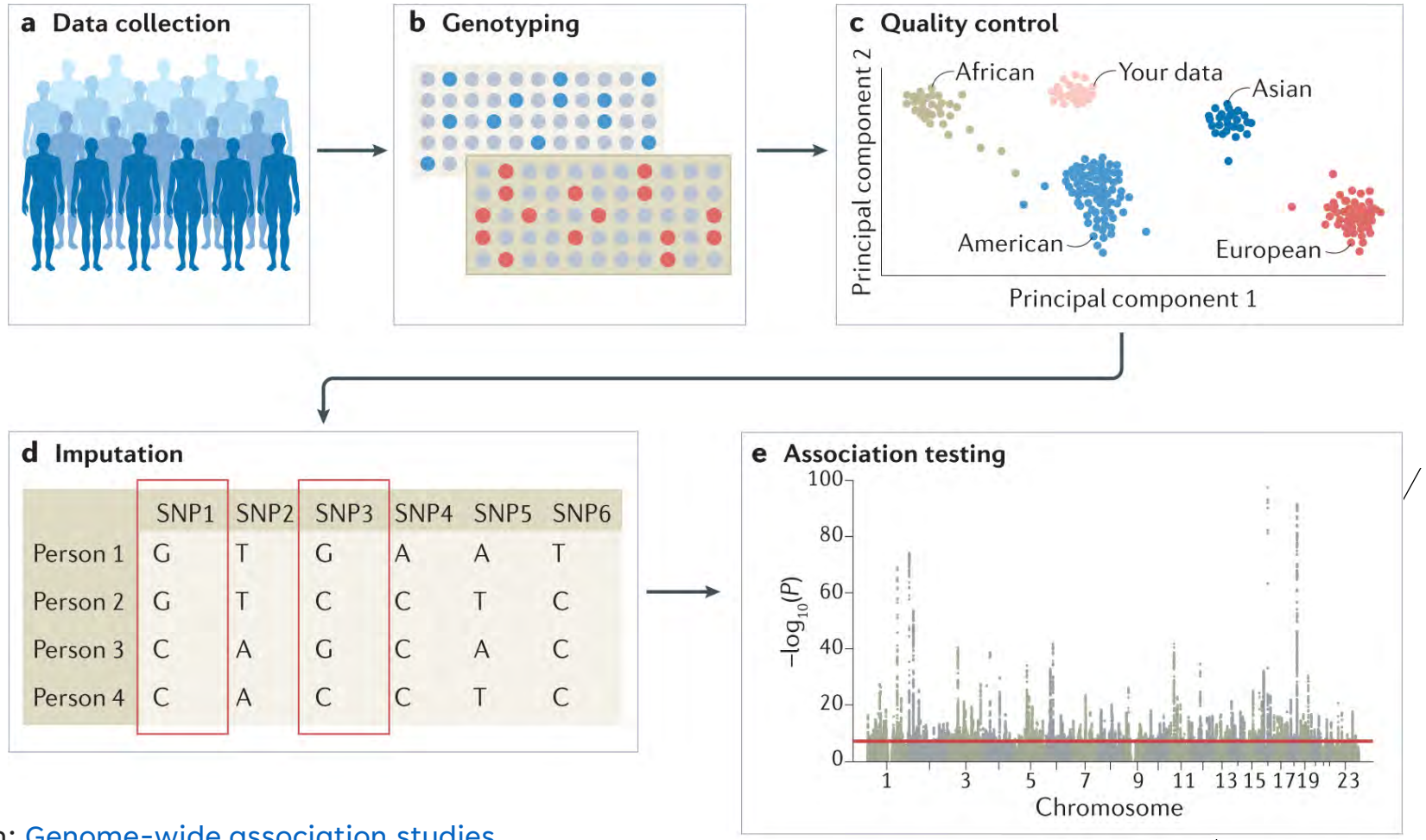


76,755 individuals with schizophrenia
and 243,649 control individuals

Figure from: Genome-wide association studies
doi.org/10.1038/s43586-021-00056-9

MAPPING THE SCHIZOPHRENIA GENOME

Dr. Antonio Pardiñas

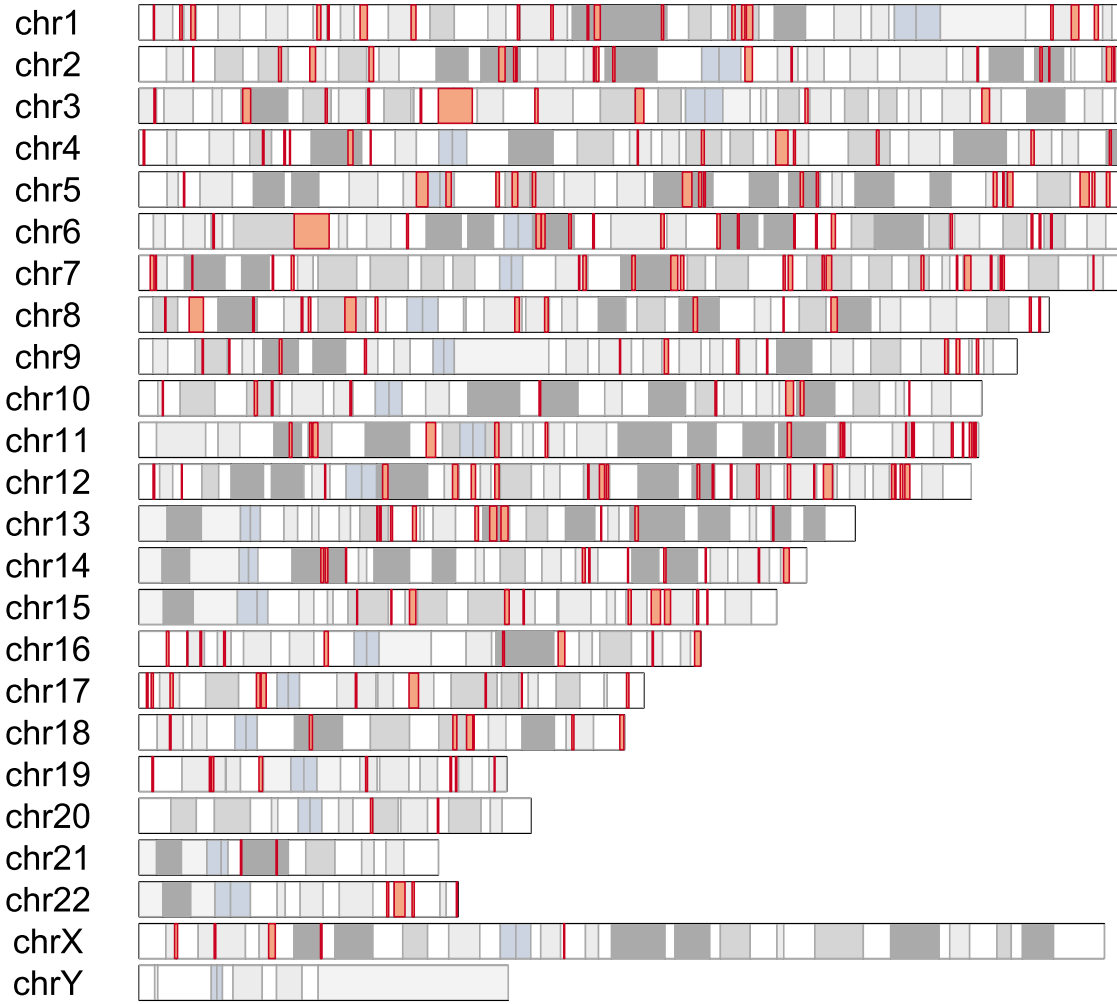


- Several terabytes of data
- Reach 1000 parallel jobs
- Creates 10000 files

From: Genome-wide association studies
doi.org/10.1038/s43586-021-00056-9

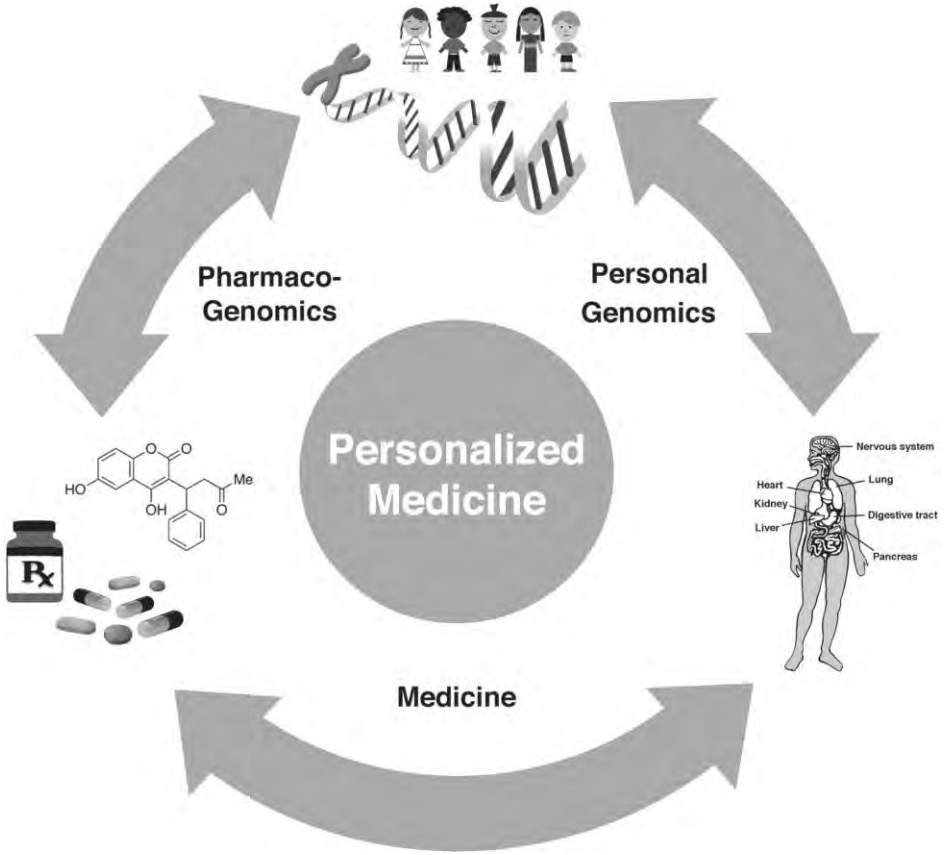
MAPPING THE SCHIZOPHRENIA GENOME

Dr. Antonio Pardiñas

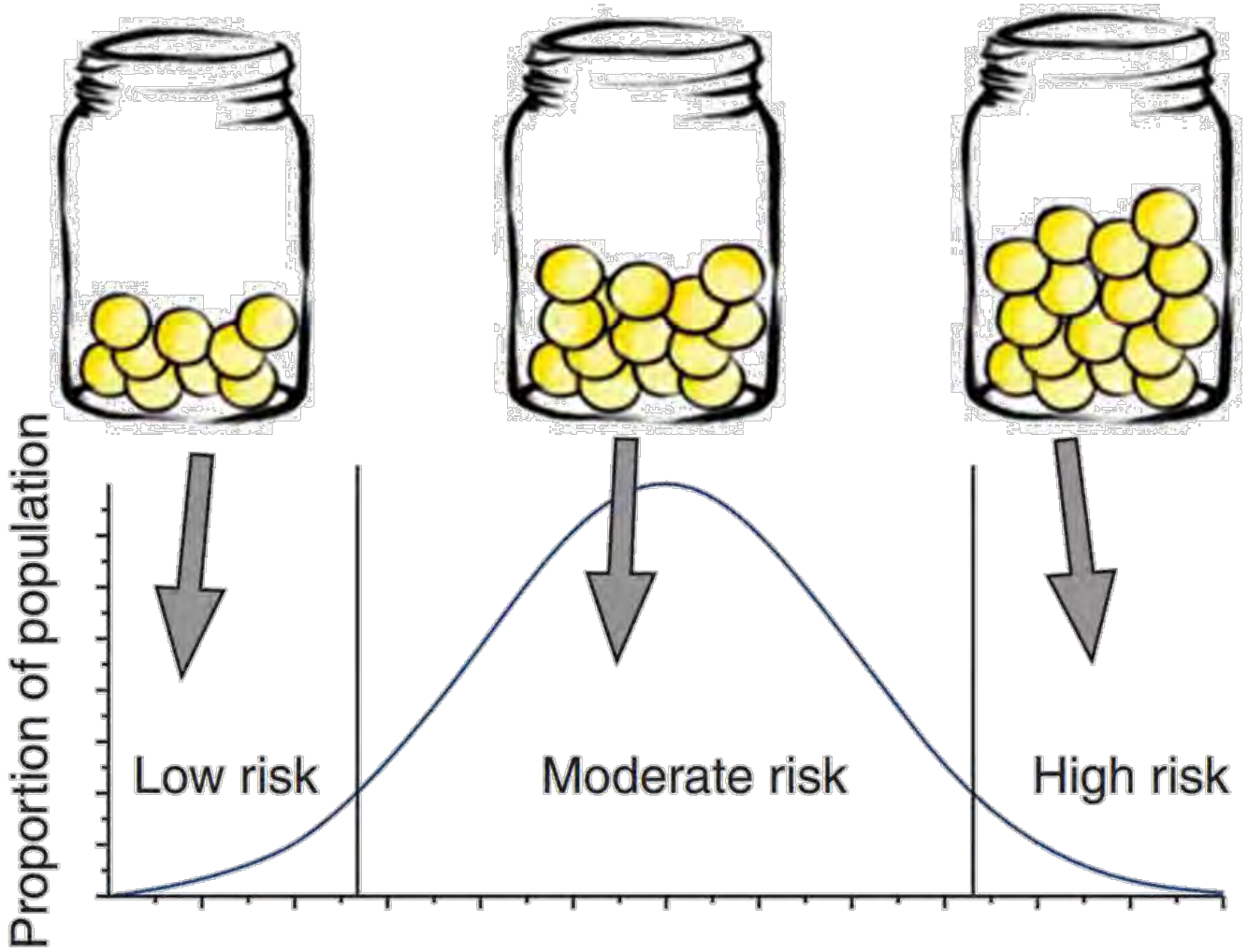


- 287 genomic regions (over 1000 genes)
  - 120 likely to be causal
- Associations concentrated in genes expressed in excitatory and inhibitory neurons of the central nervous system

WHY THIS MATTERS?



doi: 10.1093/bioinformatics/btr295

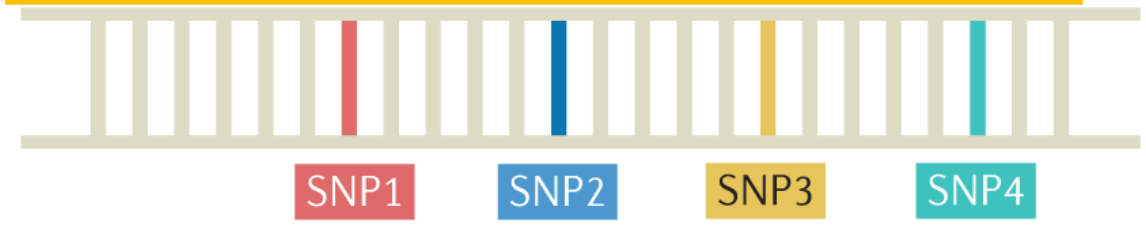


POLYGENIC RISK SCORES (PRS)

Polygenic scores capture (part of) someone's genetic "risk" by summing all risk alleles weighted by the effect sizes estimated in a GWAS

① GWAS summary statistics

Allele	A	C	T	A
Effect	+1.5	−0.5	+2.0	−1.5



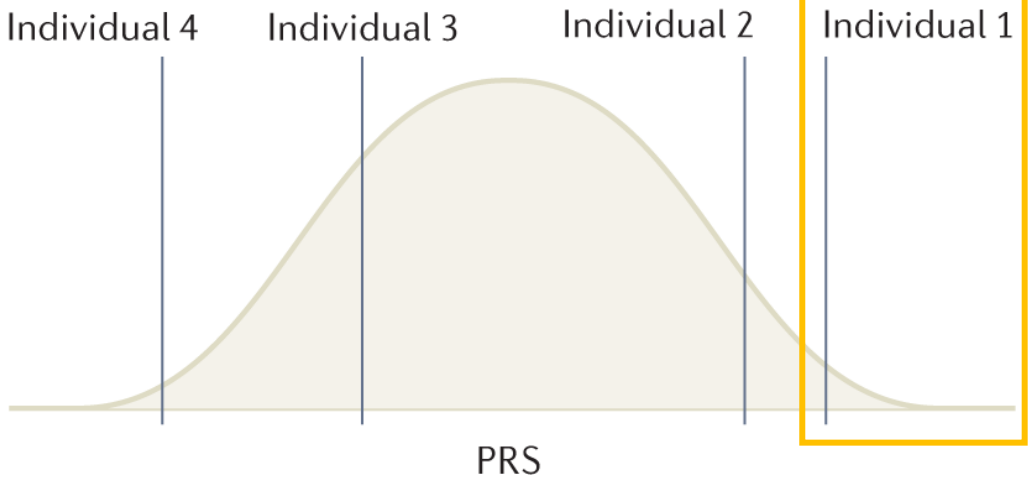
② Genotype data

	SNP1	SNP2	SNP3	SNP4
Individual 1	AT	CG	TT	CC
Individual 2	TA	GG	GT	CA
Individual 3	TT	CC	GT	CA
Individual 4	TT	CC	GG	AA

③ Polygenic risk score

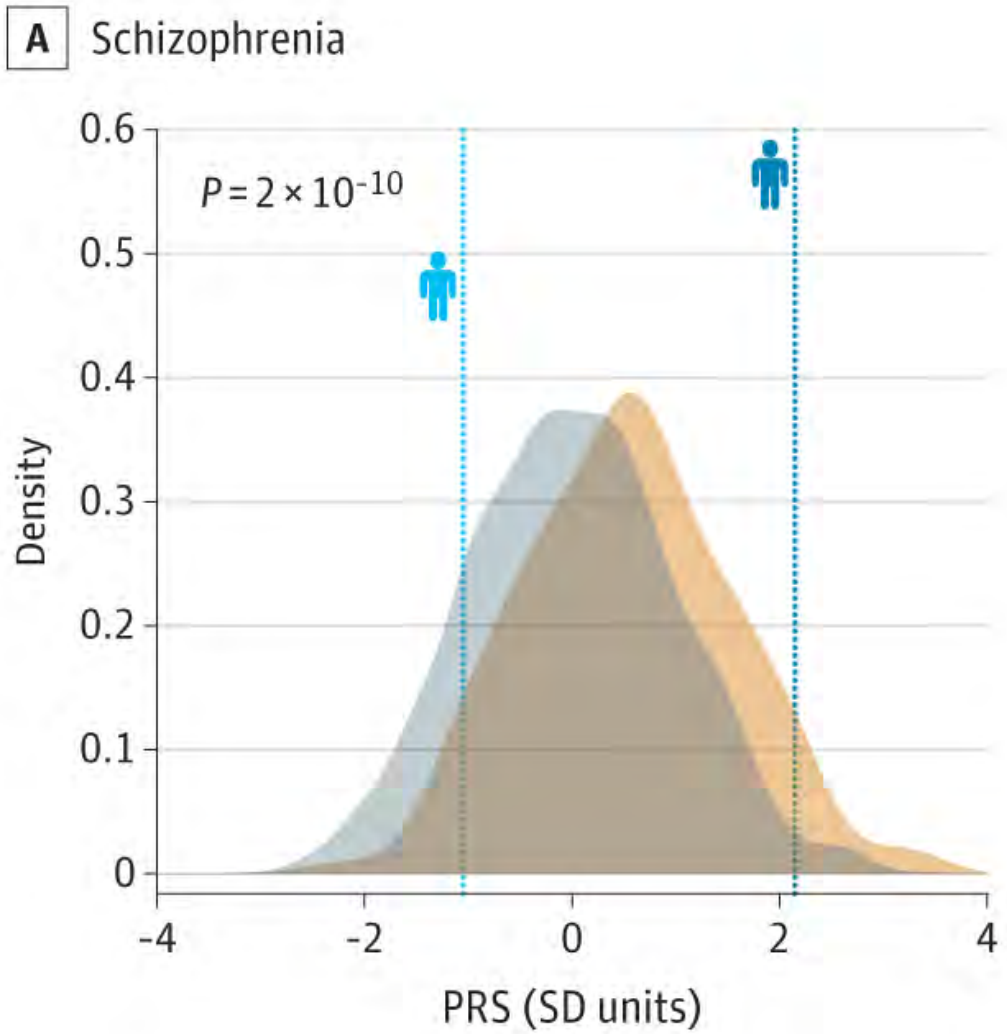
Individual 1	1.5	−	0.5	+	4.0	−	0.0	= **5.0**
Individual 2	1.5	−	0.0	+	2.0	−	1.5	= **2.0**
Individual 3	0.0	−	1.0	+	2.0	−	1.5	= **−0.5**
Individual 4	0.0	−	1.0	+	0.0	−	3.0	= **−4.0**

④ PRS distribution



From: Genome-wide association studies
doi.org/10.1038/s43586-021-00056-9

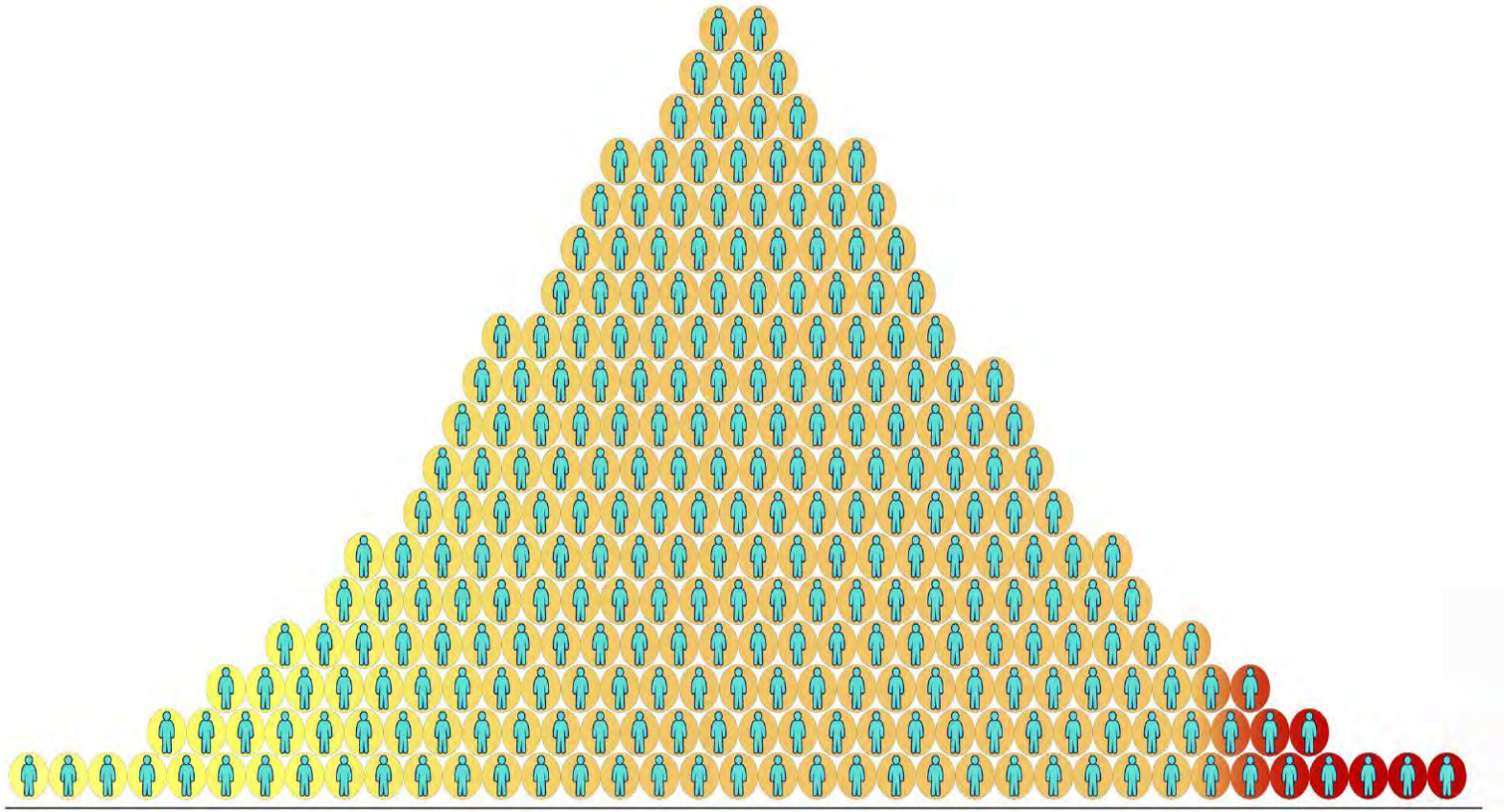
QUANTIFYING GENOMIC RISK



- Everyone carries mutations that confer some risk of schizophrenia, but not everyone has the same amount
- The same is true if we look at those who have already developed the disorder

doi:10.1001/jamapsychiatry.2020.3042

QUANTIFYING GENOMIC RISK



LOW RISK

BETTER PROGNOSIS?

RESPOND TO TREAMENT?

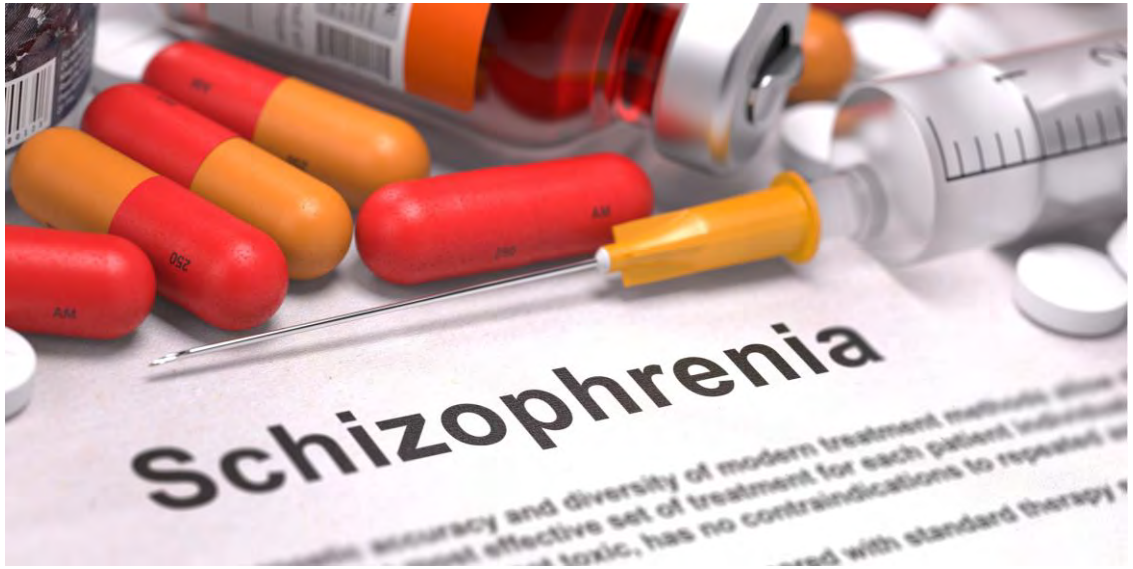
HIGH RISK

WORSE OUTCOMES?

TREATMENT- RESISTANT?

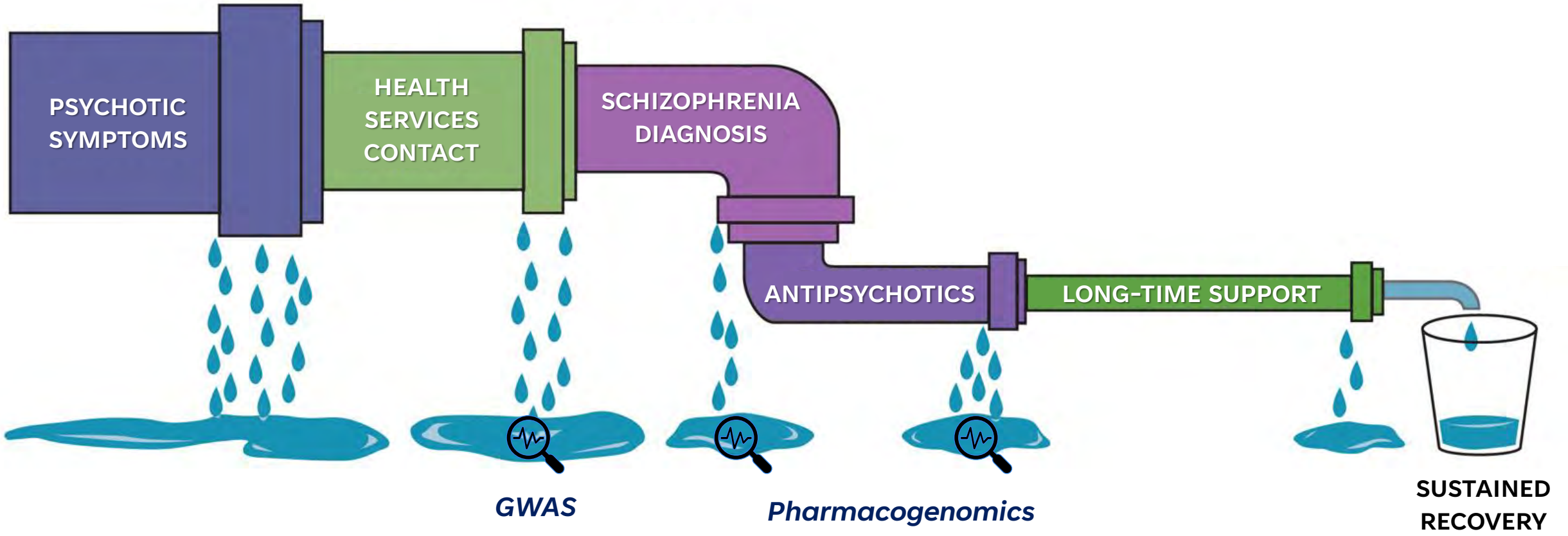
- Everyone carries mutations that confer some risk of schizophrenia, but not everyone has the same amount
- The same is true if we look at those who have already developed the disorder

CHALLENGES IN SCHIZOPHRENIA TREATMENT



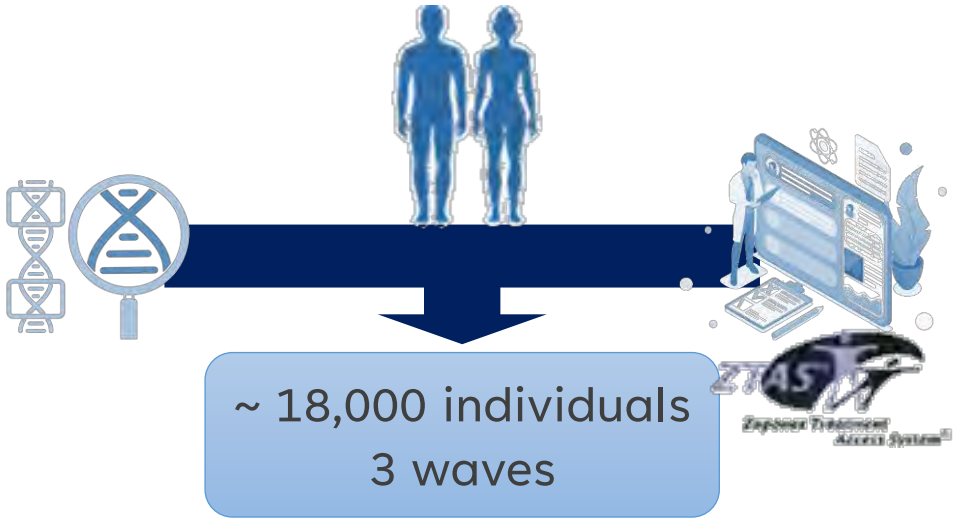
- Antipsychotic treatment hasn't changed
- 30-50% treatment failure
- Unpredictable response to the treatment of choice

CHALLENGES IN SCHIZOPHRENIA TREATMENT



THE CLOZUK STUDY

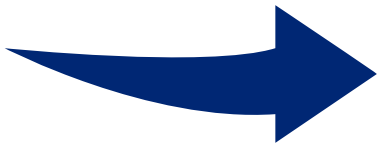
- Clozapine monitoring service
  - Blood tests and clozapine/norclozapine levels
  - Variable follow-up of these individuals



Sequencing (rare variants)	Clozapine blood levels
Array genotyping (SNPs – common variants)	Blood tests

GENOMIC PREDICTION OF CLOZAPINE DOSE

- Medication of choice for TRS
- Very effective but prone to severe ADRs, if not carefully managed
- Severe under-prescription
- Potentially long times on suboptimal doses



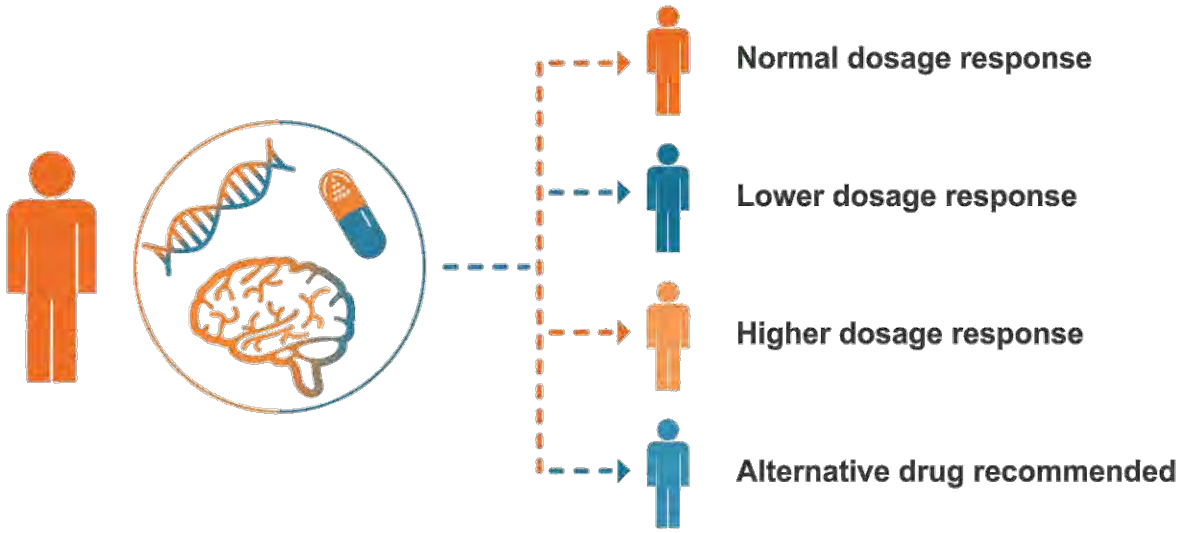
Safer, more efficient way to use clozapine

GENOMIC PREDICTION OF CLOZAPINE DOSE

"historically (…) prescribing practices have therefore been referred to as more of an art than a science"

www.thelancet.com/psychiatry **Published online April 25, 2016** **http://dx.doi.org/10.1016/S2215-0366(16)00017-1**

- Clozapine doses are gradually escalated (from 100-300 mg/day), prescribing the correct dose can take weeks or months.
- Dose-adjusting is probably one of the easiest ways to implement pharmacogenomic findings



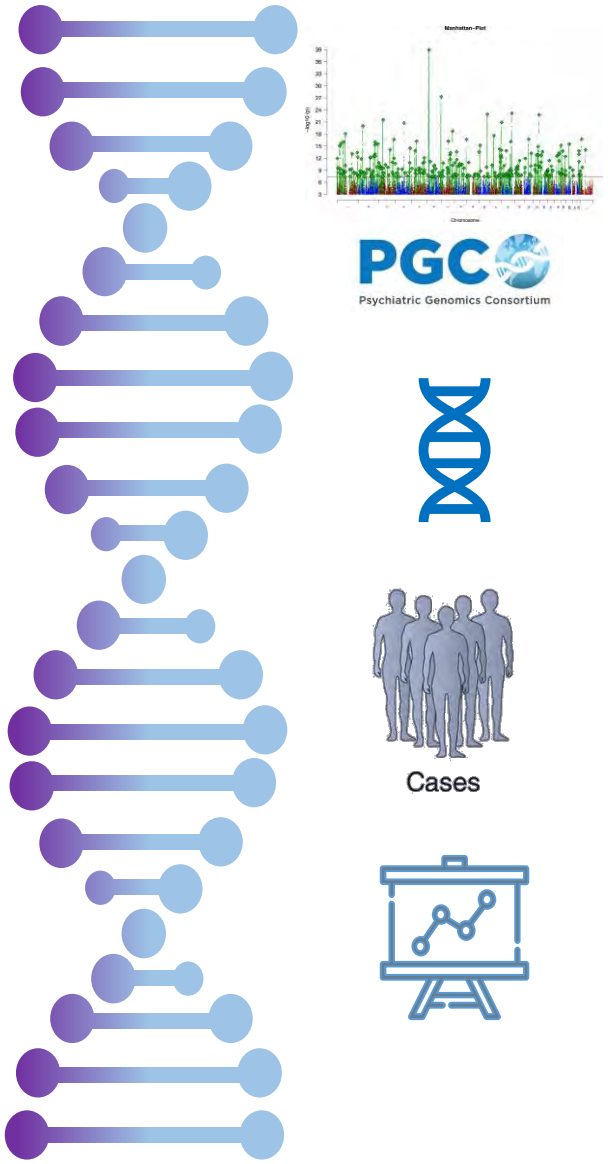
GENOMIC PREDICTION OF CLOZAPINE DOSE

Genomic Stratification of Clozapine Prescription Patterns Using Schizophrenia Polygenic Scores.

Kappel DB, Legge SE, Hubbard L, Willcocks IR, O'Connell KS, Smith RL, Molden E, Andreassen OA, King A, Jansen J, Helthuis M, Owen MJ, O'Donovan MC, Walters JTR, Pardiñas AF.

Biol Psychiatry. 2022 Aug 5:S0006-3223(22)01449-4. doi: 10.1016/j.biopsych.2022.07.014. Online ahead of print.

PMID: 36244804 Free article.



Schizophrenia GWAS

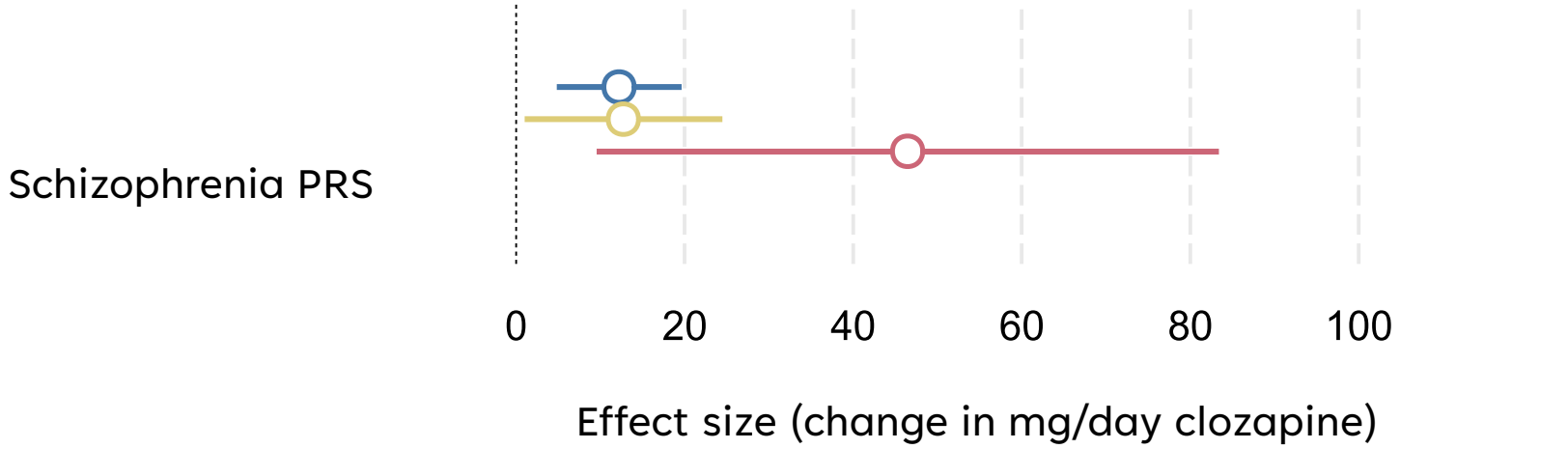
Polygenic risk

4,000 individuals with Schizophrenia

Clozapine prescription patterns

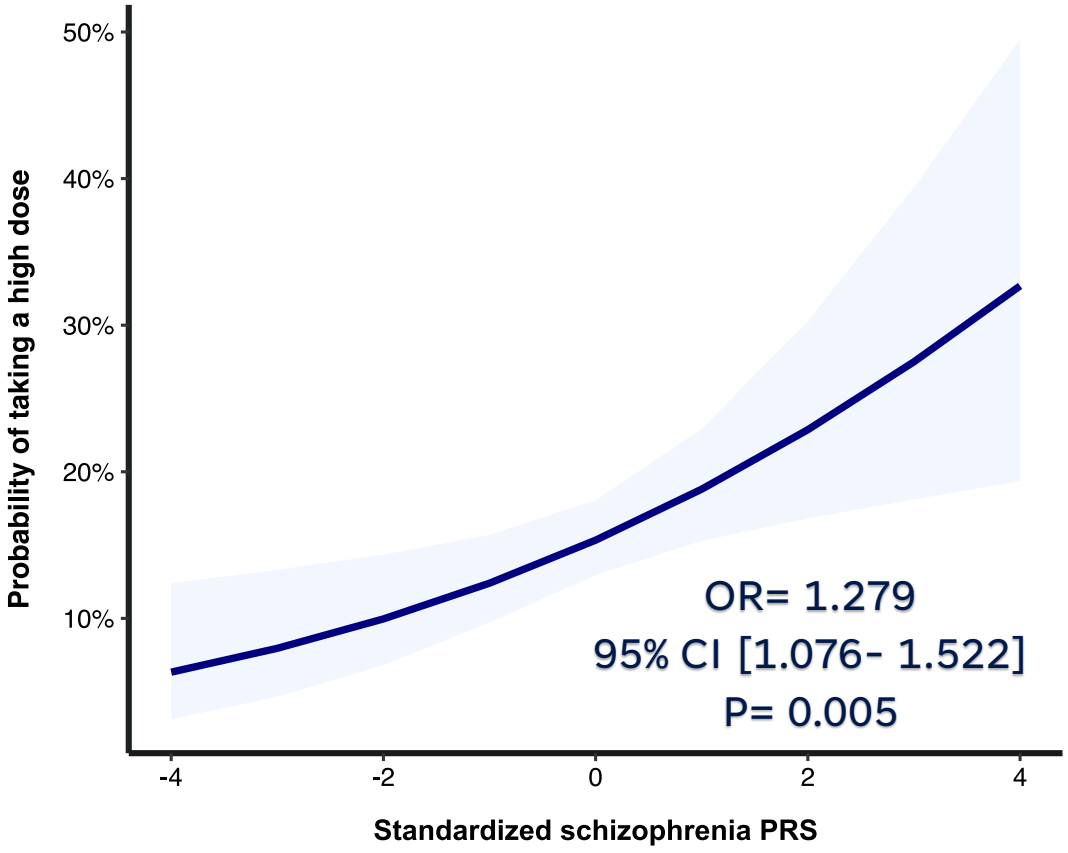
Kappel et al., 2022

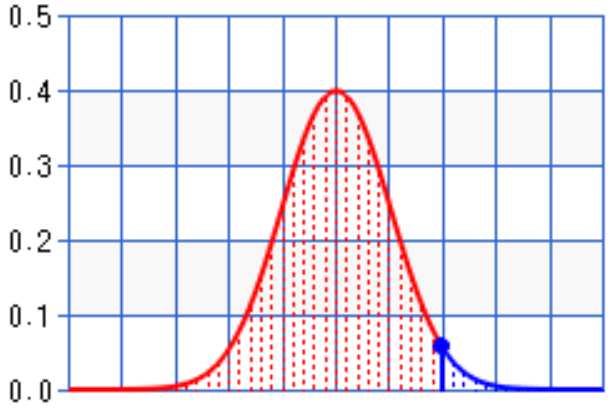
GENOMIC PREDICTION OF CLOZAPINE DOSE



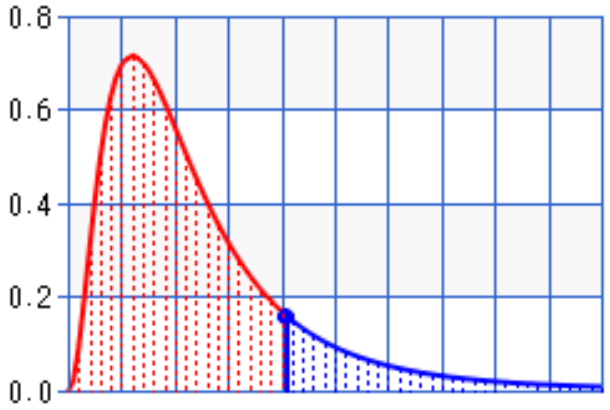
Kappel et al., 2022

GENOMIC PREDICTION OF CLOZAPINE DOSE





Schizophrenia risk: **Upper 2.5%**



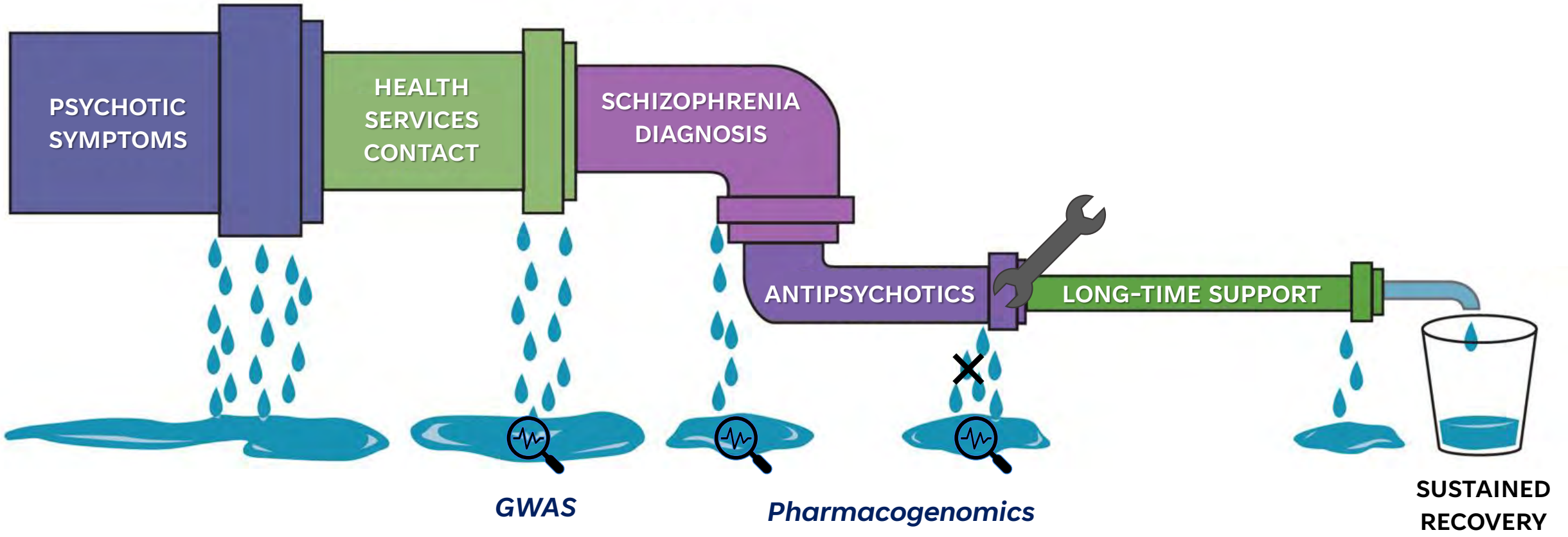
Antipsychotic dose: **>600 mg/day**

- Those at high genomic risk of schizophrenia were twice more likely to be prescribed a high antipsychotic dose.
- Genomics might be the first predictor of needing high doses.

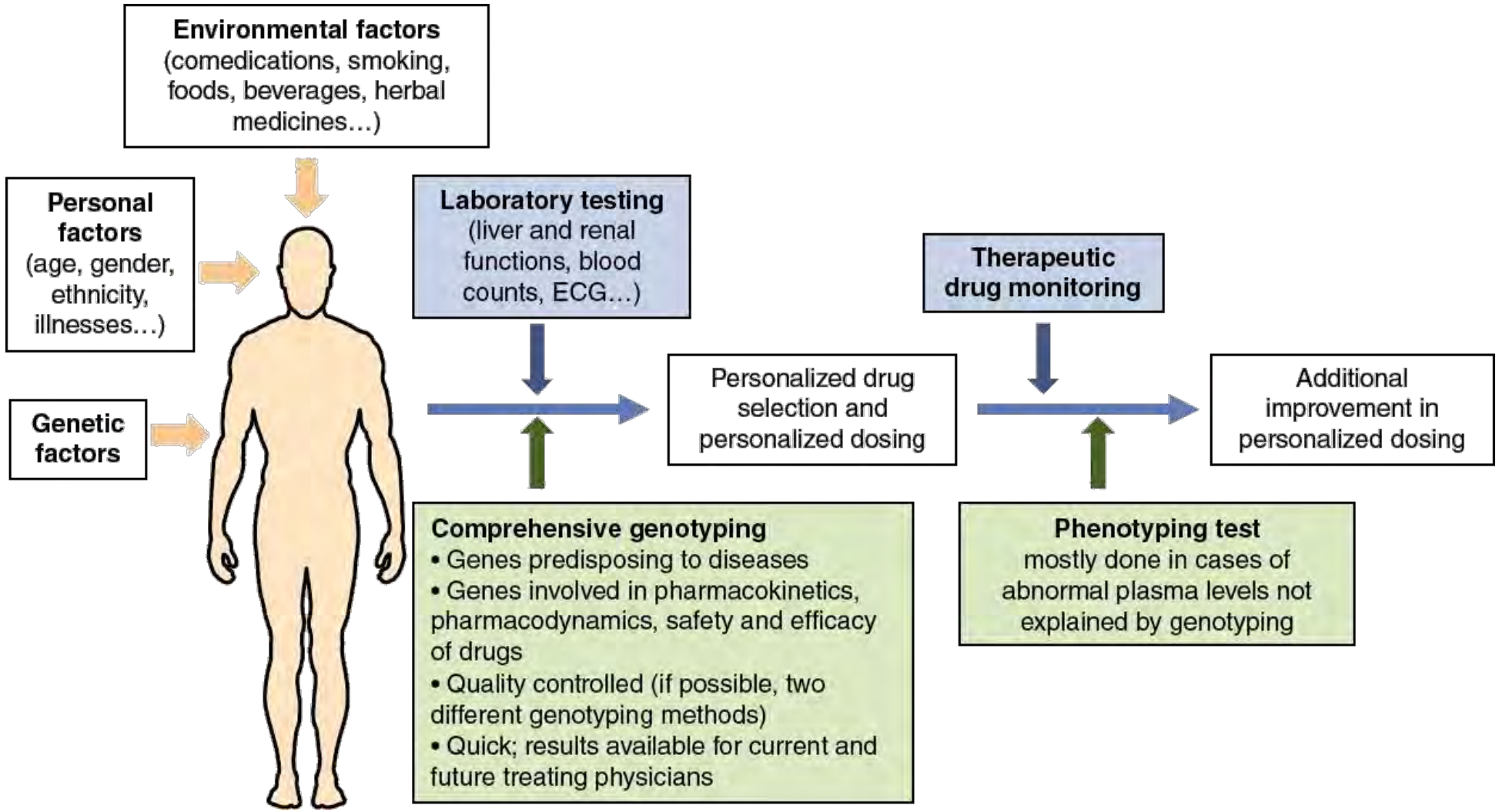
Kappel et al., 2022

Centre for Neuropsychiatric Genetics and Genomics
Canolfan Geneteg a Genomeg Niwroseiciatrig
The Academy of Medical Sciences

CHALLENGES IN SCHIZOPHRENIA TREATMENT

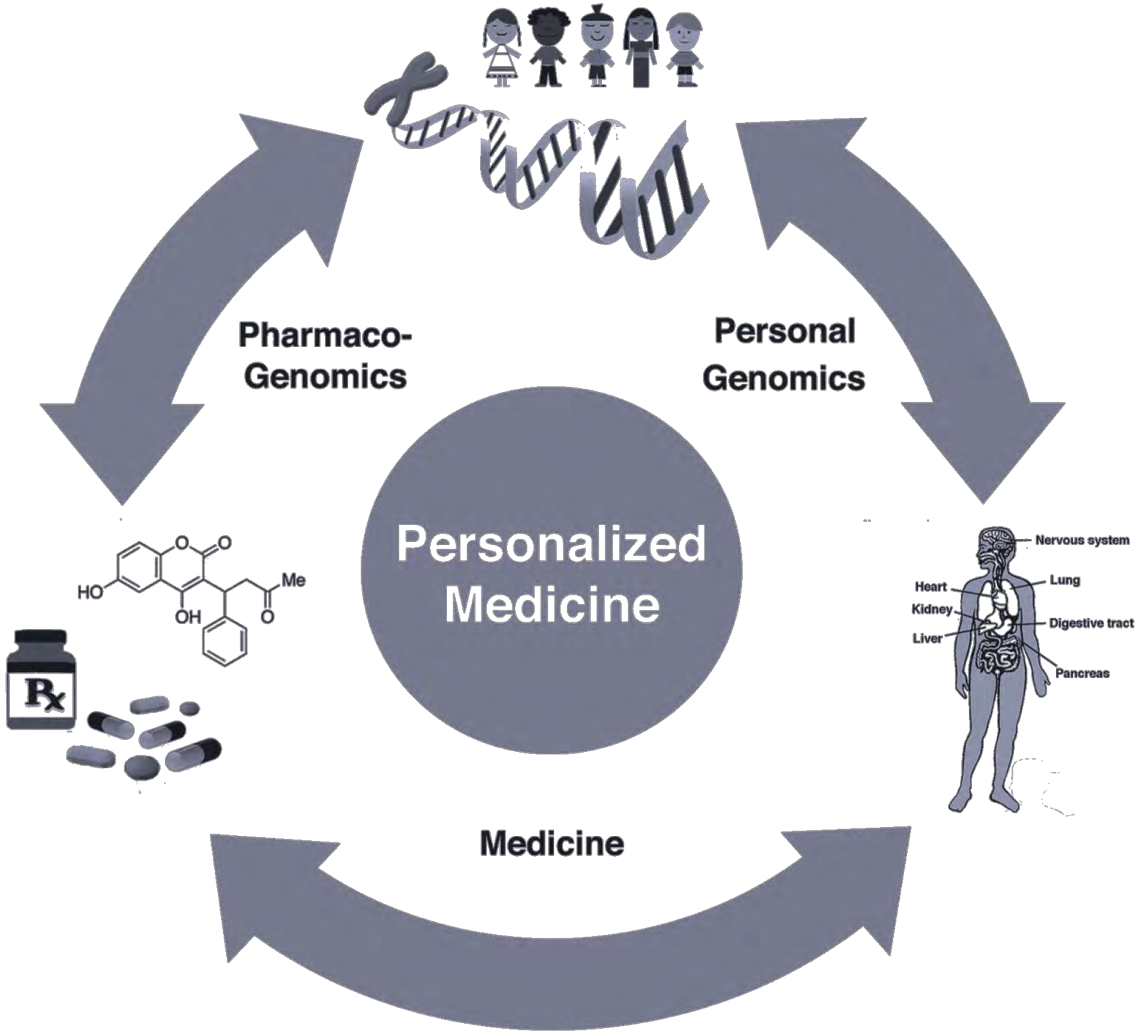


GENETICS OF CLOZAPINE METABOLISM



doi.org/10.1038/clpt.2013.221

GENETICS OF CLOZAPINE METABOLISM



Pharmacogenomics - use of genomic data to understand drug metabolism and response

doi: 10.1093/bioinformatics/btr295

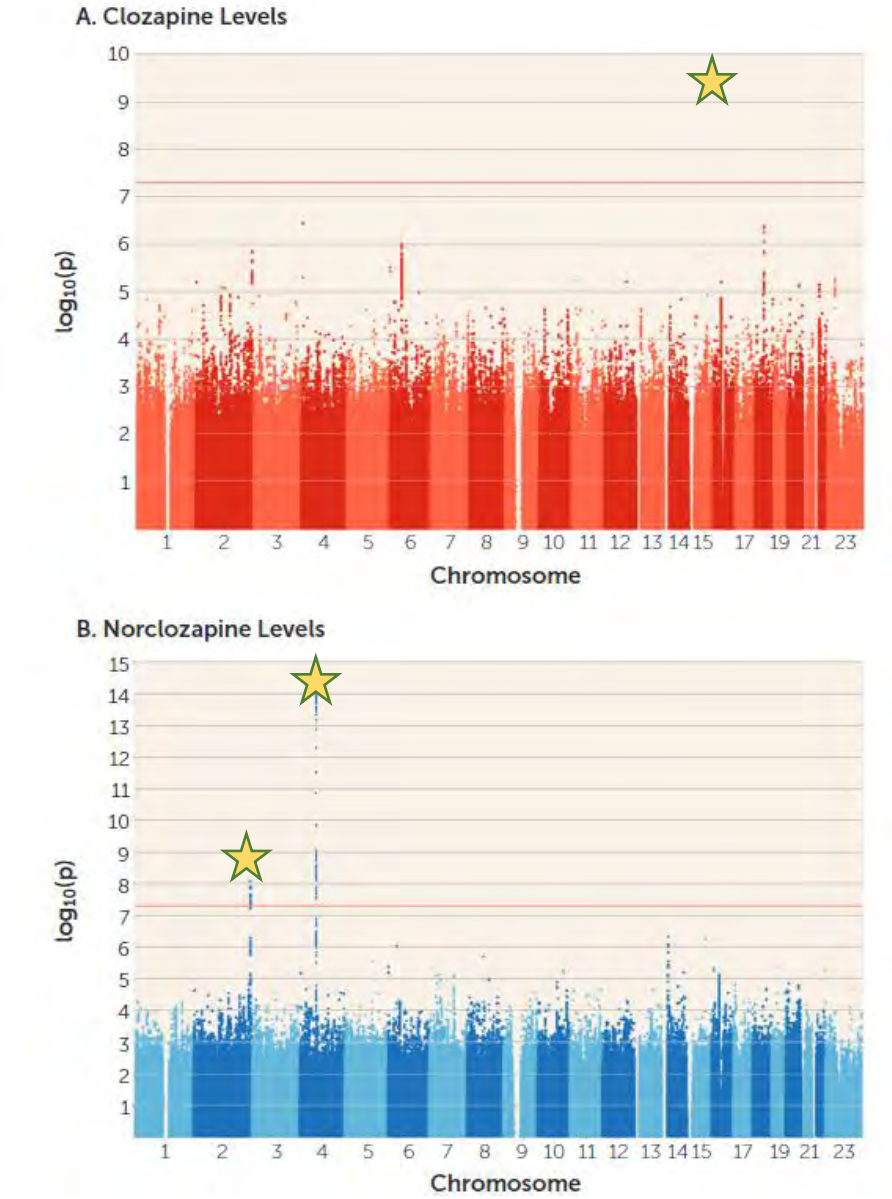
GENETICS OF CLOZAPINE METABOLISM

You can do a GWAS for anything...

› Am J Psychiatry. 2019 Jun 1;176(6):477-486. doi: 10.1176/appi.ajp.2019.18050589. Epub 2019 Mar 29.

Pharmacogenomic Variants and Drug Interactions Identified Through the Genetic Analysis of Clozapine Metabolism

Antonio F Pardiñas [1], Mariana Nalmpanti [1], Andrew J Pocklington [1], Sophie E Legge [1], Christopher Medway [1], Adrian King [1], John Jansen [1], Marinka Helthuis [1], Stanley Zammit [1], James MacCabe [1], Michael J Owen [1], Michael C O'Donovan [1], James T R Walters [1]



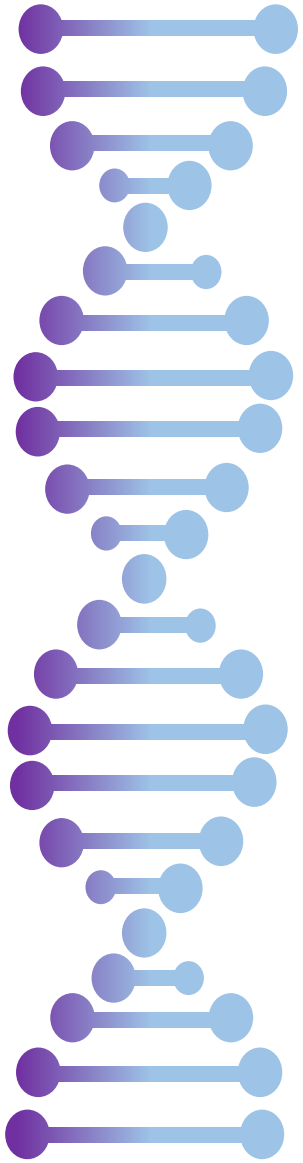
GENETICS OF CLOZAPINE METABOLISM

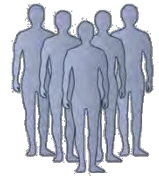
Pharmacokinetics and pharmacogenomics of clozapine in an ancestrally diverse sample:A longitudinal analysis and GWAS using clinical monitoring data from the UK

Antonio F. Pardiñas, Djenifer B. Kappel, Milly Roberts, Francesca Tipple, Lisa M. Shitomi-Jones, Adrian King, John Jansen, Marinka Helthuis, Michael J. Owen, Michael C. O'Donovan, James T.R. Walters

doi: https://doi.org/10.1101/2022.09.23.22280299

medRxiv THE PREPRINT SERVER FOR HEALTH SCIENCES





Cases

4,495 individuals with schizophrenia taking clozapine

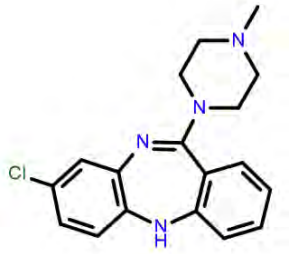
Longitudinal trajectory GWAS

Clozapine metabolite levels

R: 10s per SNP (~1 Year)
Julia: 6h

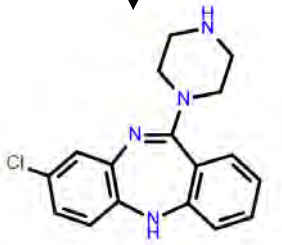
Pardiñas et al., 2022 *in review*

GENETICS OF CLOZAPINE METABOLISM

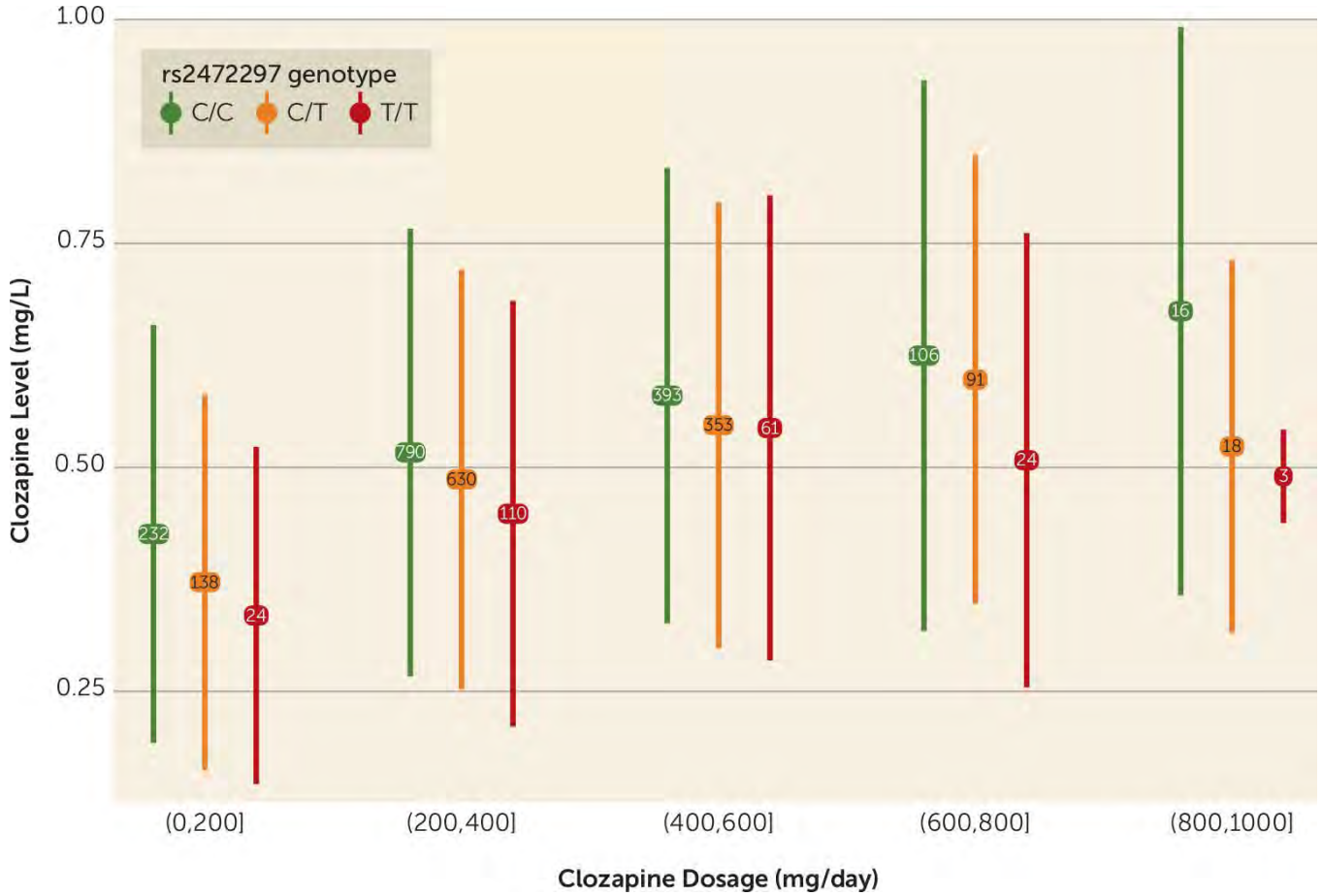


CLOZAPINE

CYP1A2



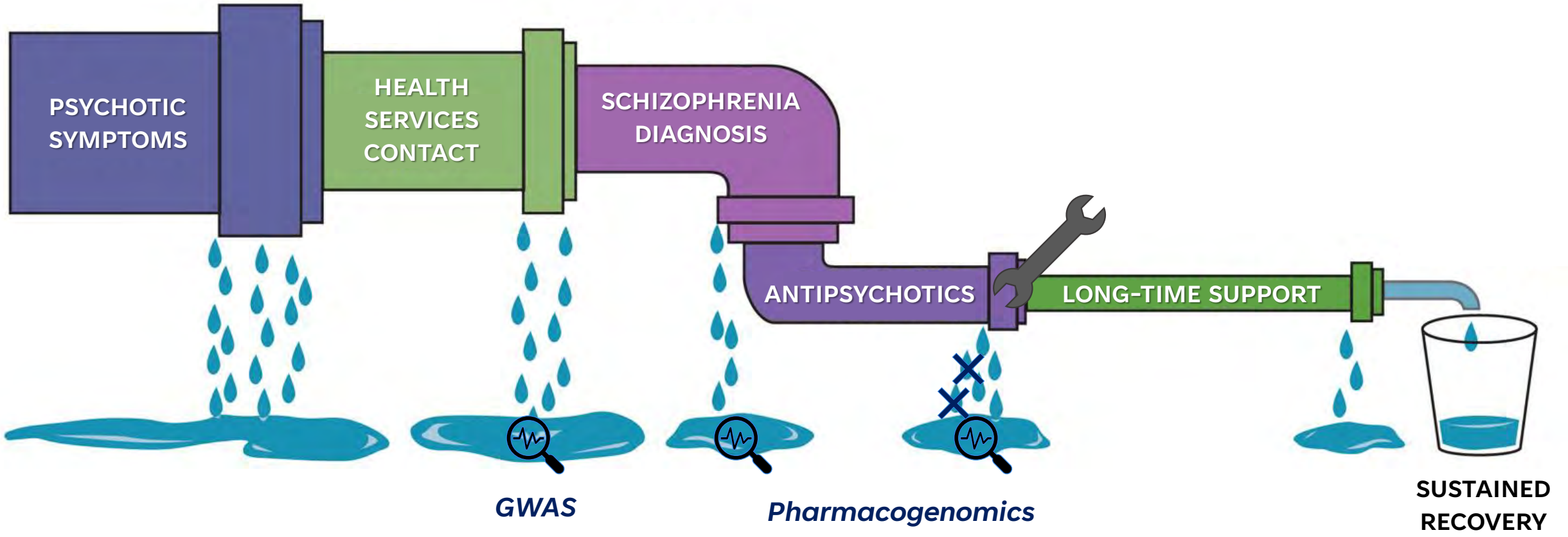
NORCLOZAPINE



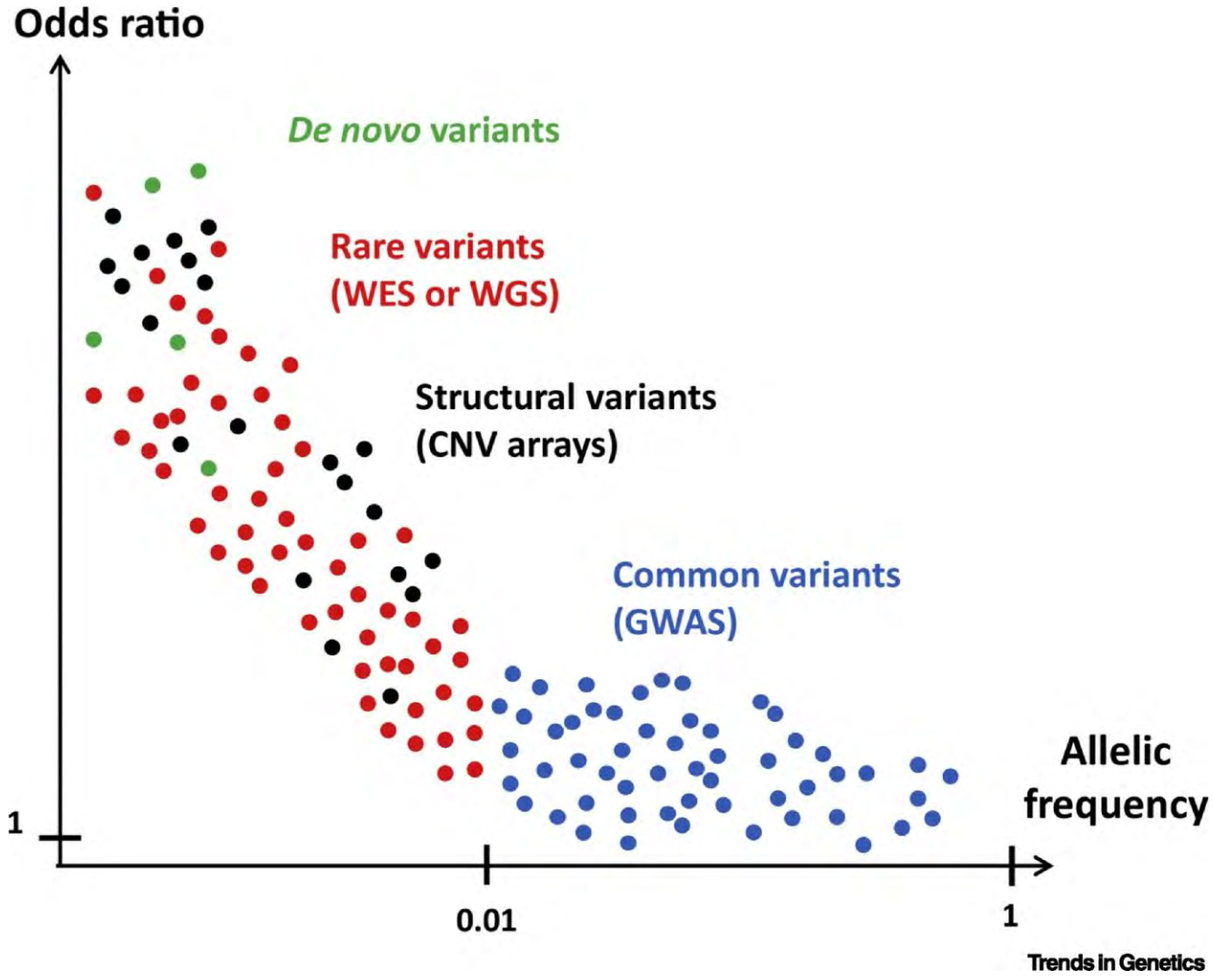
- Carrying a CYP1A2 mutation is equivalent to reducing antipsychotic doses in 50 mg/day.
- Might be an important consideration when choosing an individual's dose.

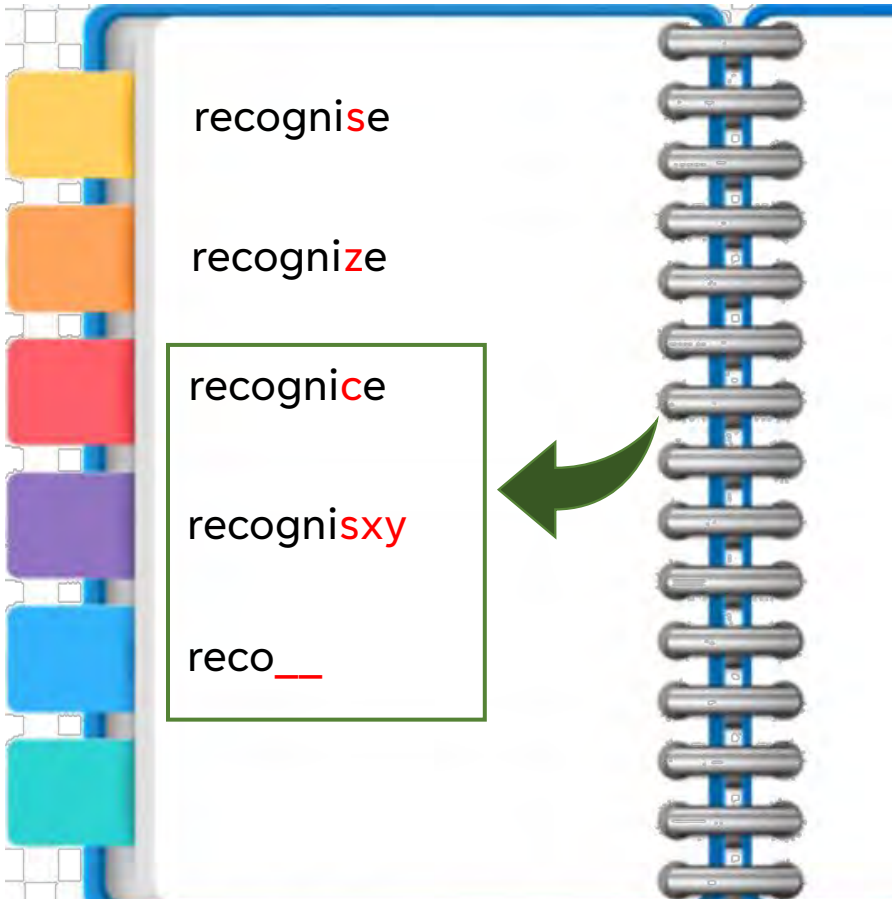
Pardiñas et al., 2022 *in review*

CHALLENGES IN SCHIZOPHRENIA TREATMENT



Complex trait genetics are influenced by the additive risk of both common and rare variants

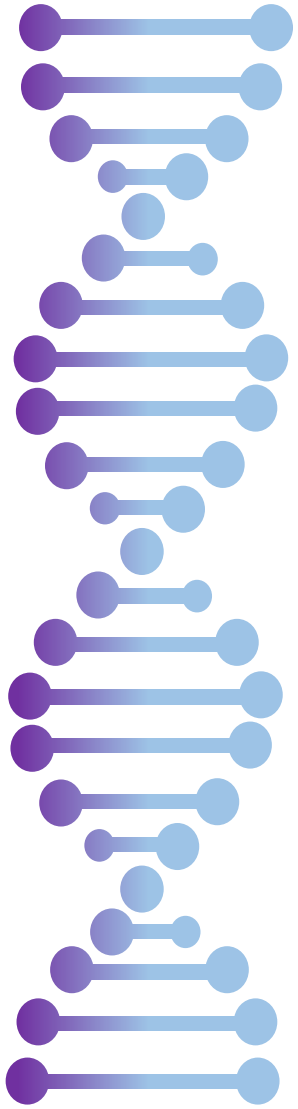




Centre for Neuropsychiatric Genetics and Genomics
Canolfan Geneteg a Genomeg Niwroseiciatrig
The Academy of Medical Sciences

RARE VARIANTS AND CLOZAPINE METABOLISM

Investigate whether rare genomic variation in genes linked to drug metabolism contributes to inter-individual differences in clozapine plasma concentrations



2,000 individuals with schizophrenia taking clozapine

Rare genetic variants

Clozapine metabolite levels

Generalized mixed-effect models

Kappel et al., *in prep*

Centre for Neuropsychiatric Genetics and Genomics
Canolfan Geneteg a Genomeg Niwroseiciatrig
The Academy of Medical Sciences

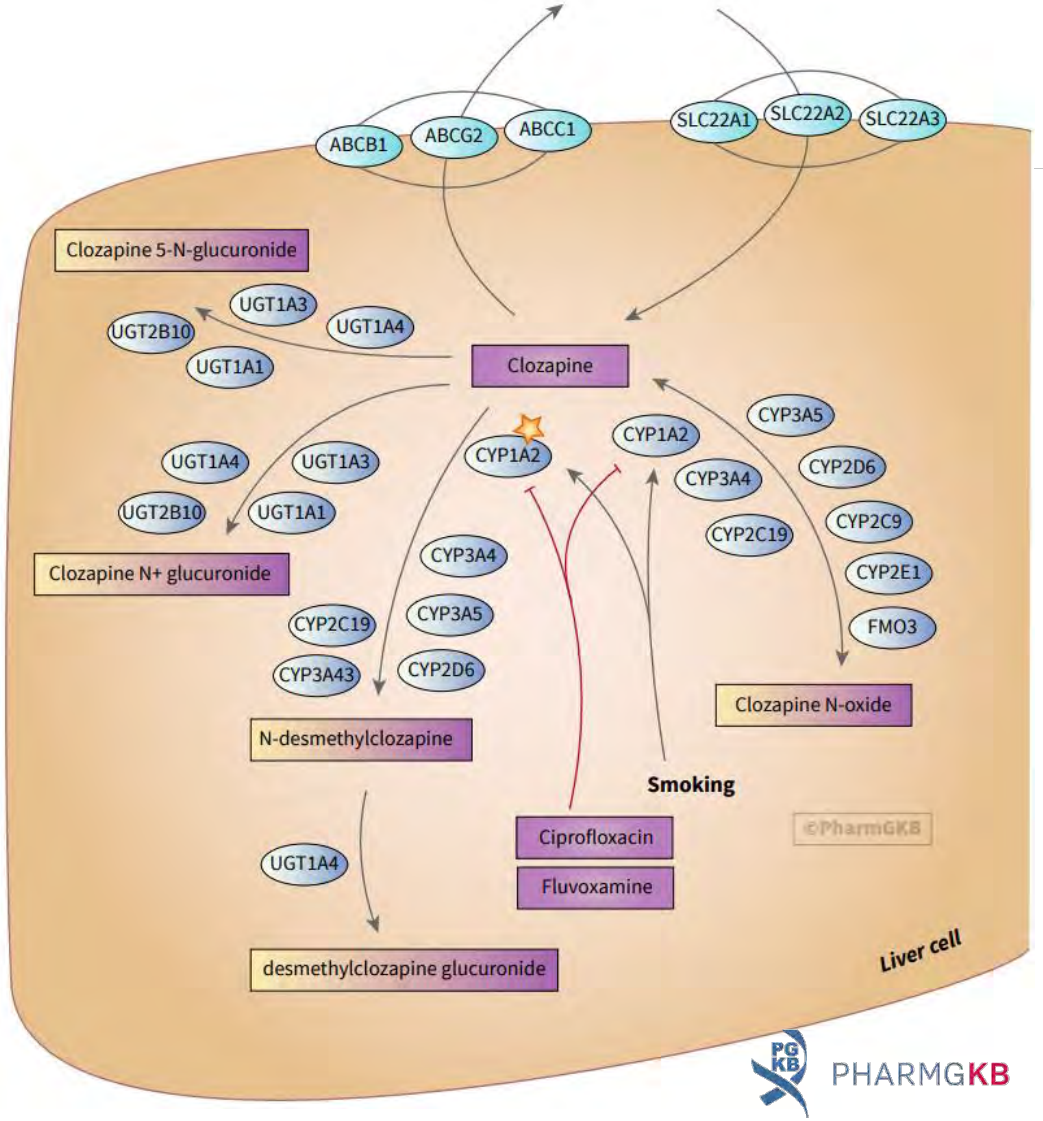
RARE VARIANTS AND CLOZAPINE METABOLISM

SPARK distribution compatible

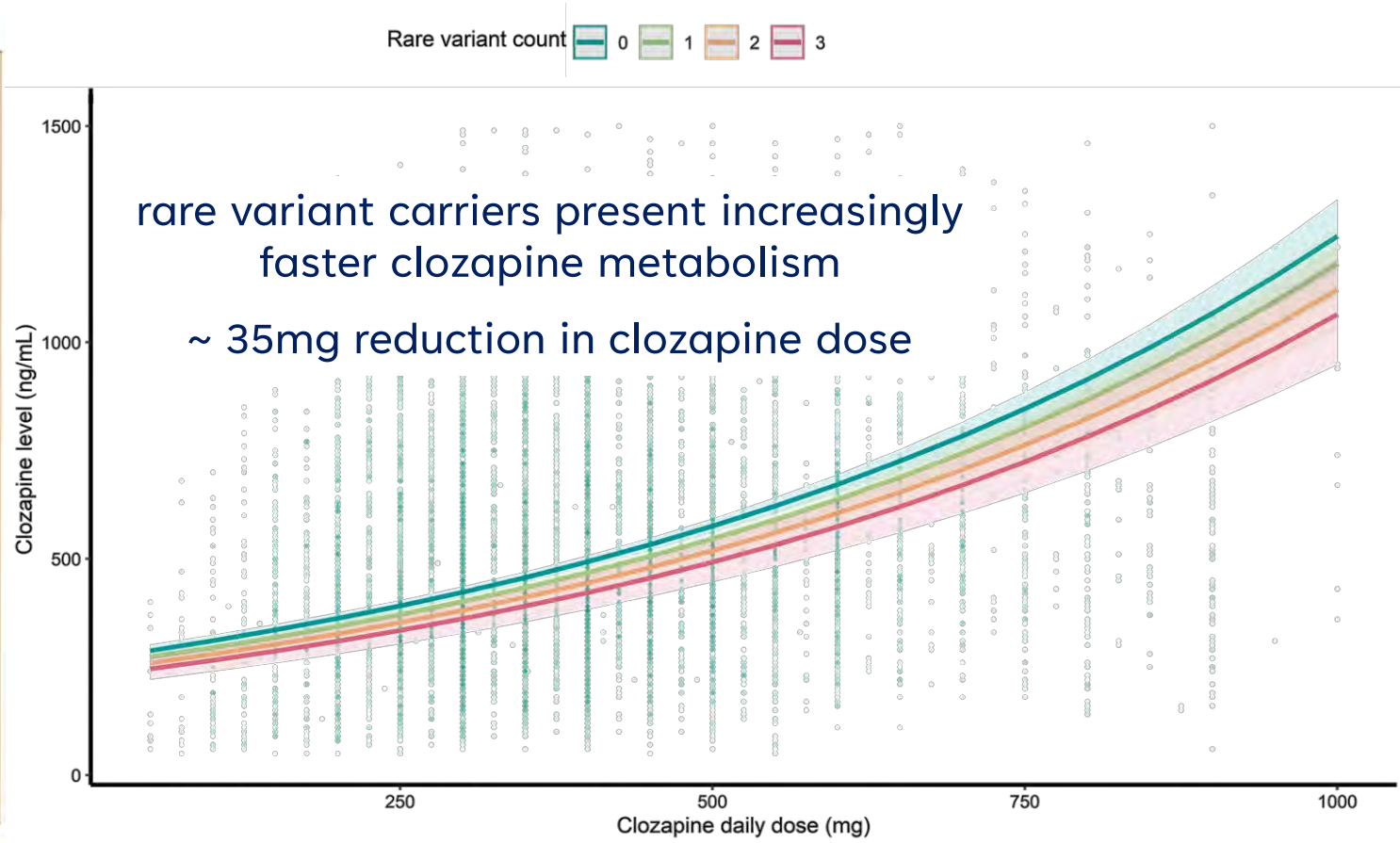
Kappel et al., *in prep*

Centre for Neuropsychiatric Genetics and Genomics
Canolfan Geneteg a Genomeg Niwroseiciatrig
The Academy of Medical Sciences

RARE VARIANTS AND CLOZAPINE METABOLISM

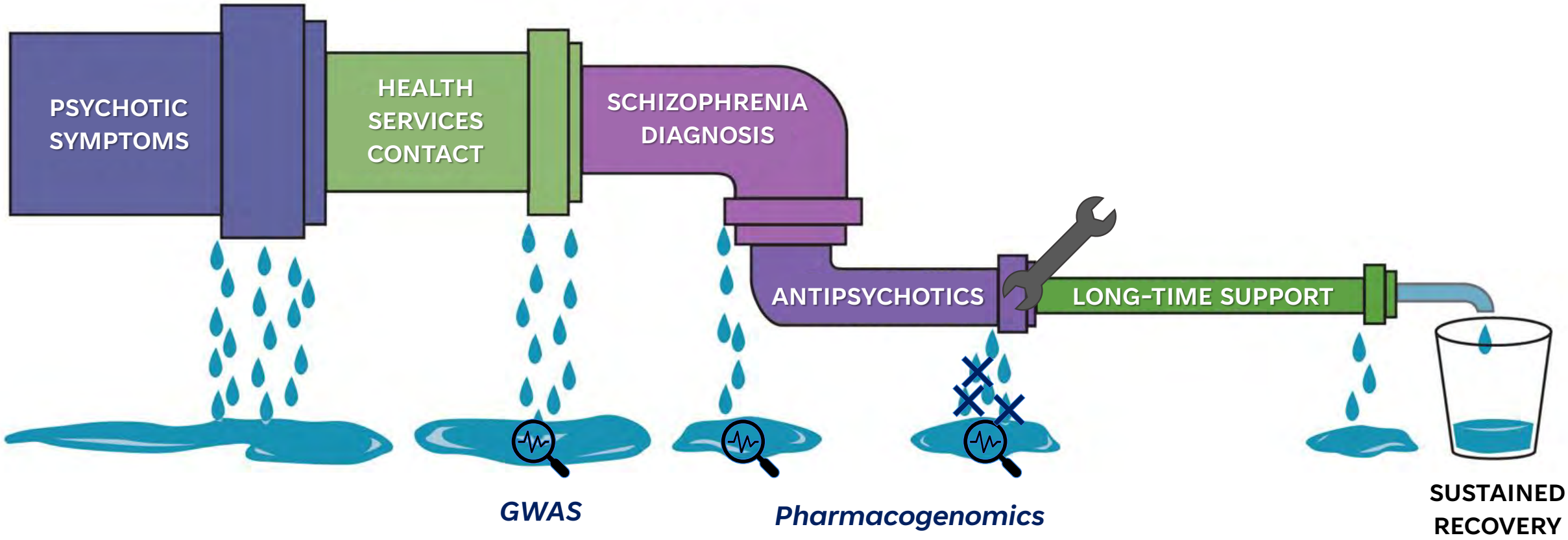


PHARMGKB



Kappel et al., *in prep*

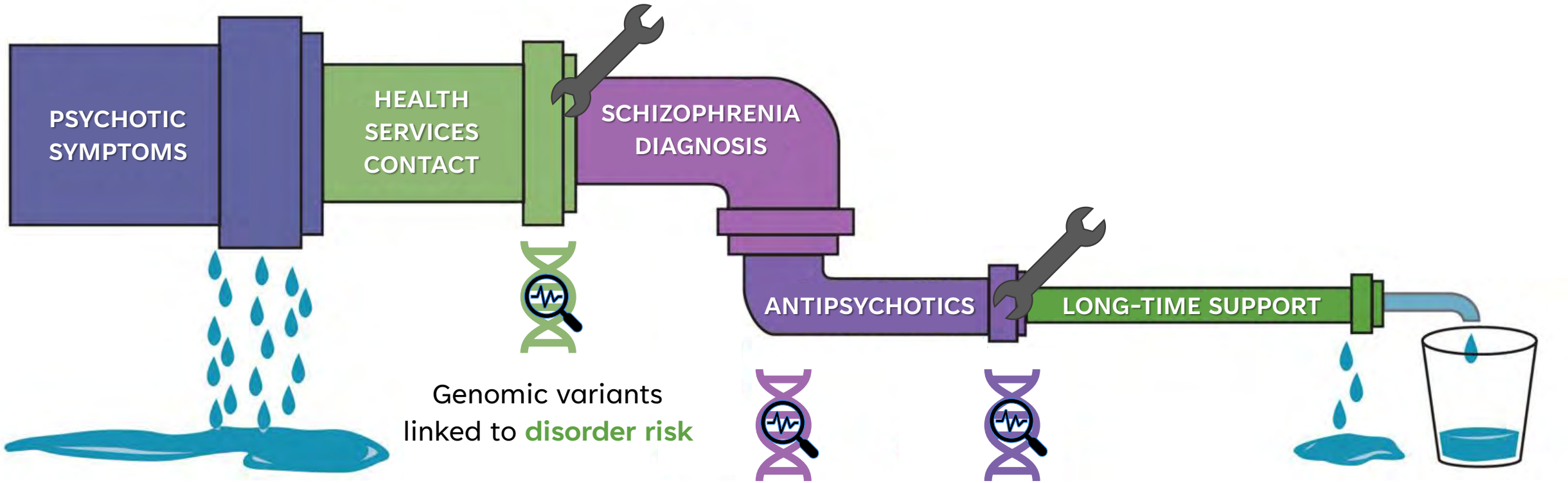
CHALLENGES IN SCHIZOPHRENIA TREATMENT



Centre for Neuropsychiatric Genetics and Genomics
Canolfan Geneteg a Genomeg Niwroseiciatrig
The Academy of Medical Sciences

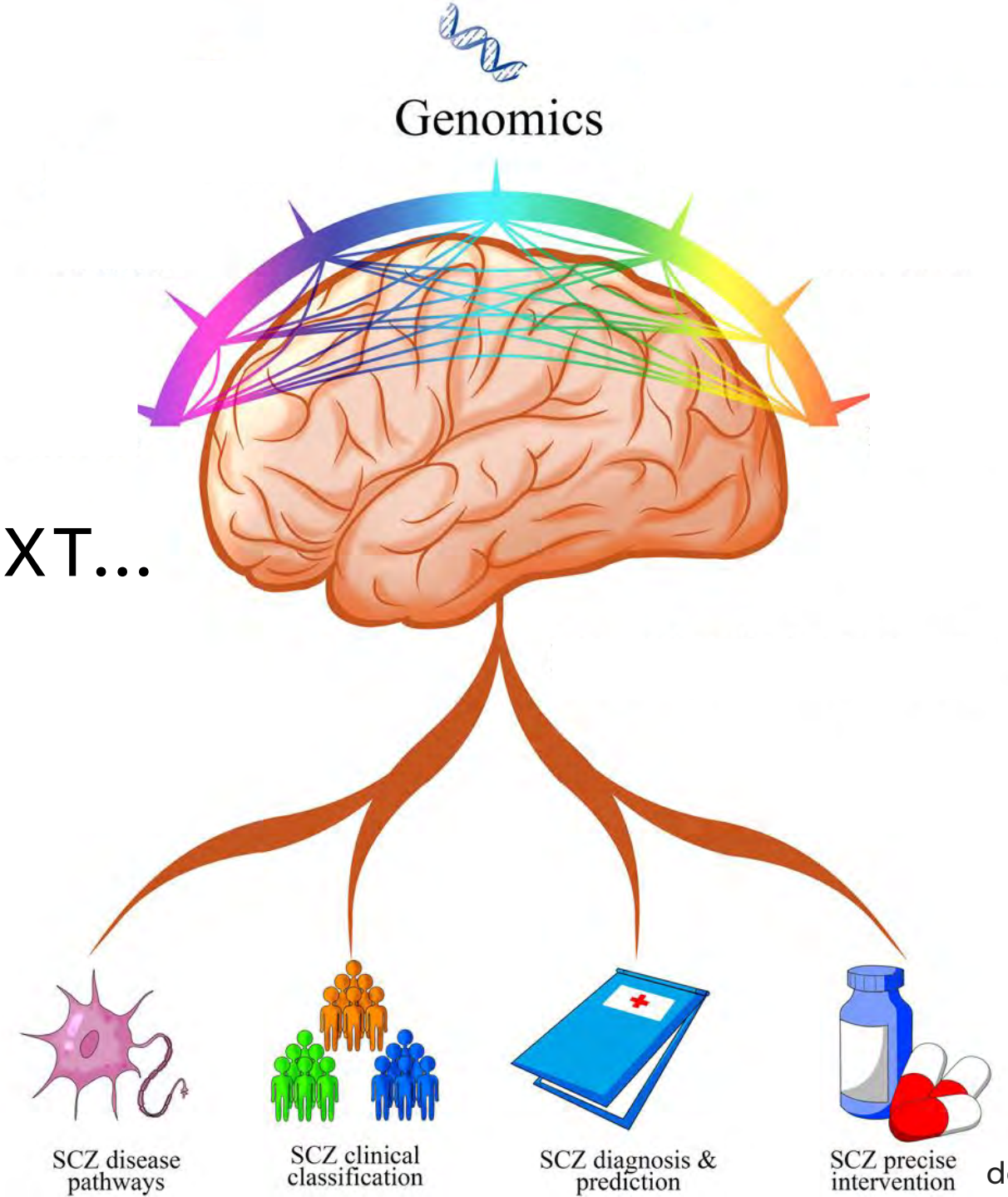
SUMMARY - FIXING THE LEAKS

Developing a “genomics-aware” standard of psychiatric care



CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD
Centre for Neuropsychiatric Genetics and Genomics
Canolfan Geneteg a Genomeg Niwroseiciatrig
The Academy of Medical Sciences

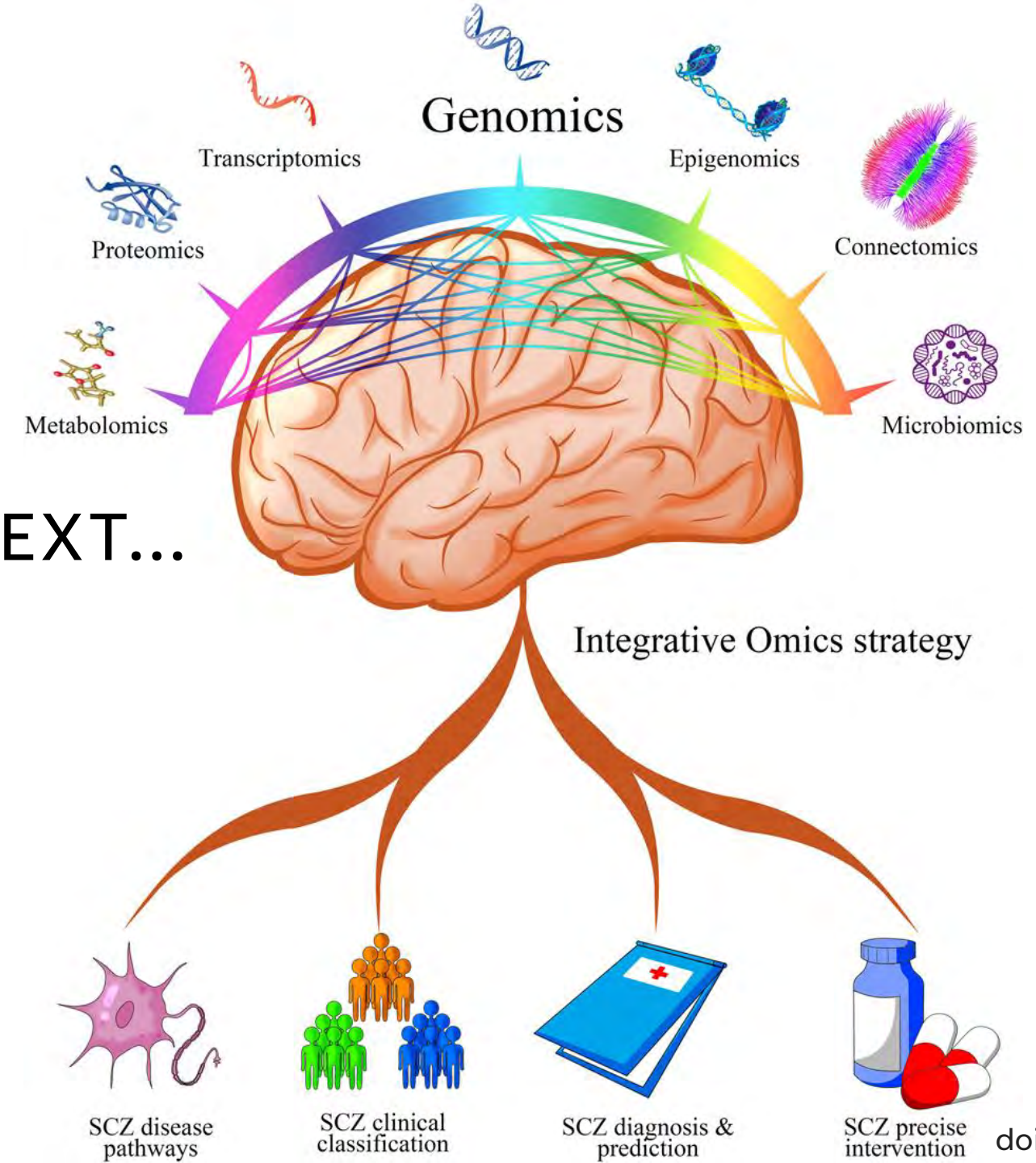
WHERE TO NEXT...



doi.org/10.1038/s41380-021-01201-2

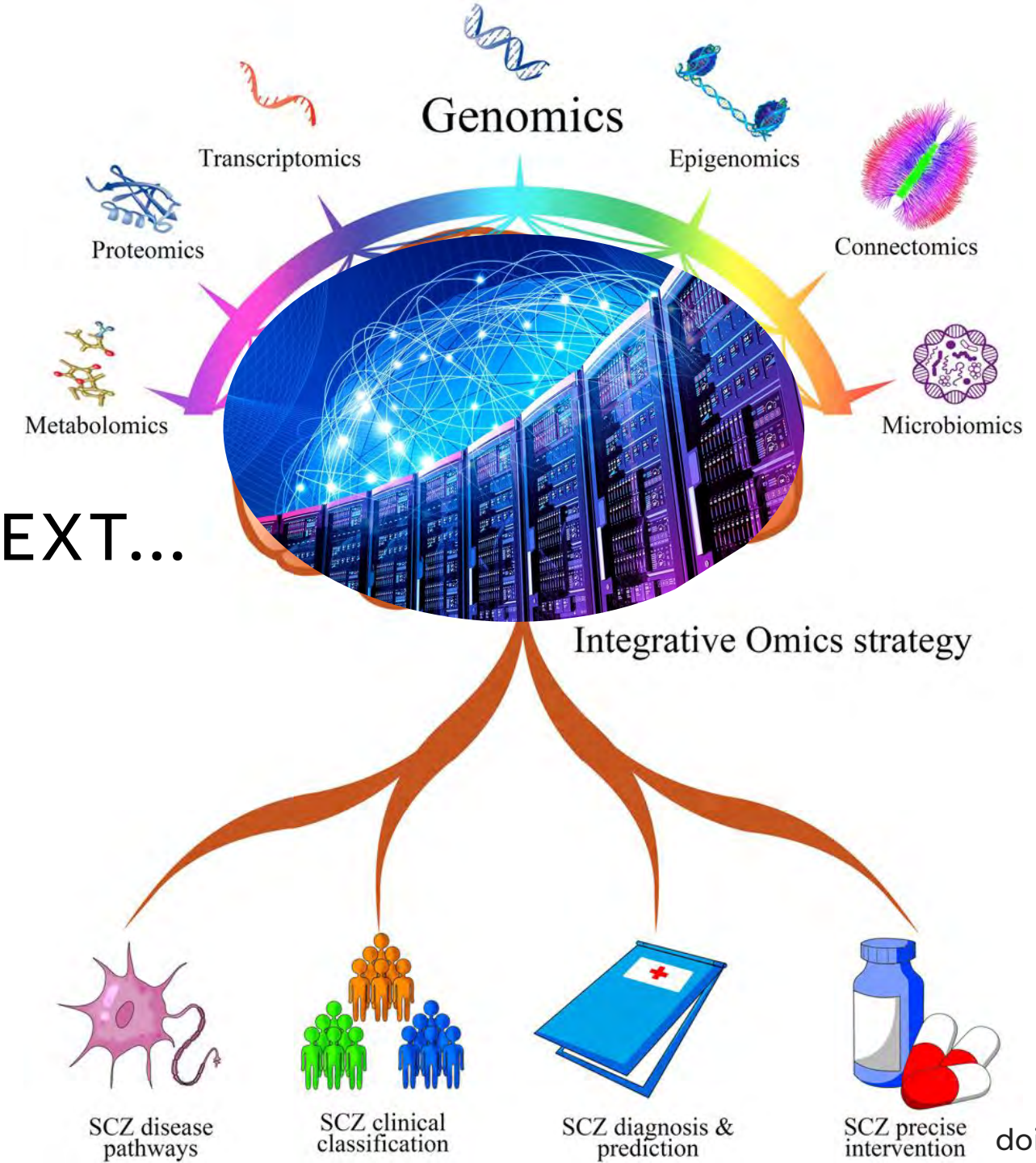
Centre for Neuropsychiatric Genetics and Genomics
Canolfan Geneteg a Genomeg Niwroseiciatrig
The Academy of Medical Sciences

WHERE TO NEXT...

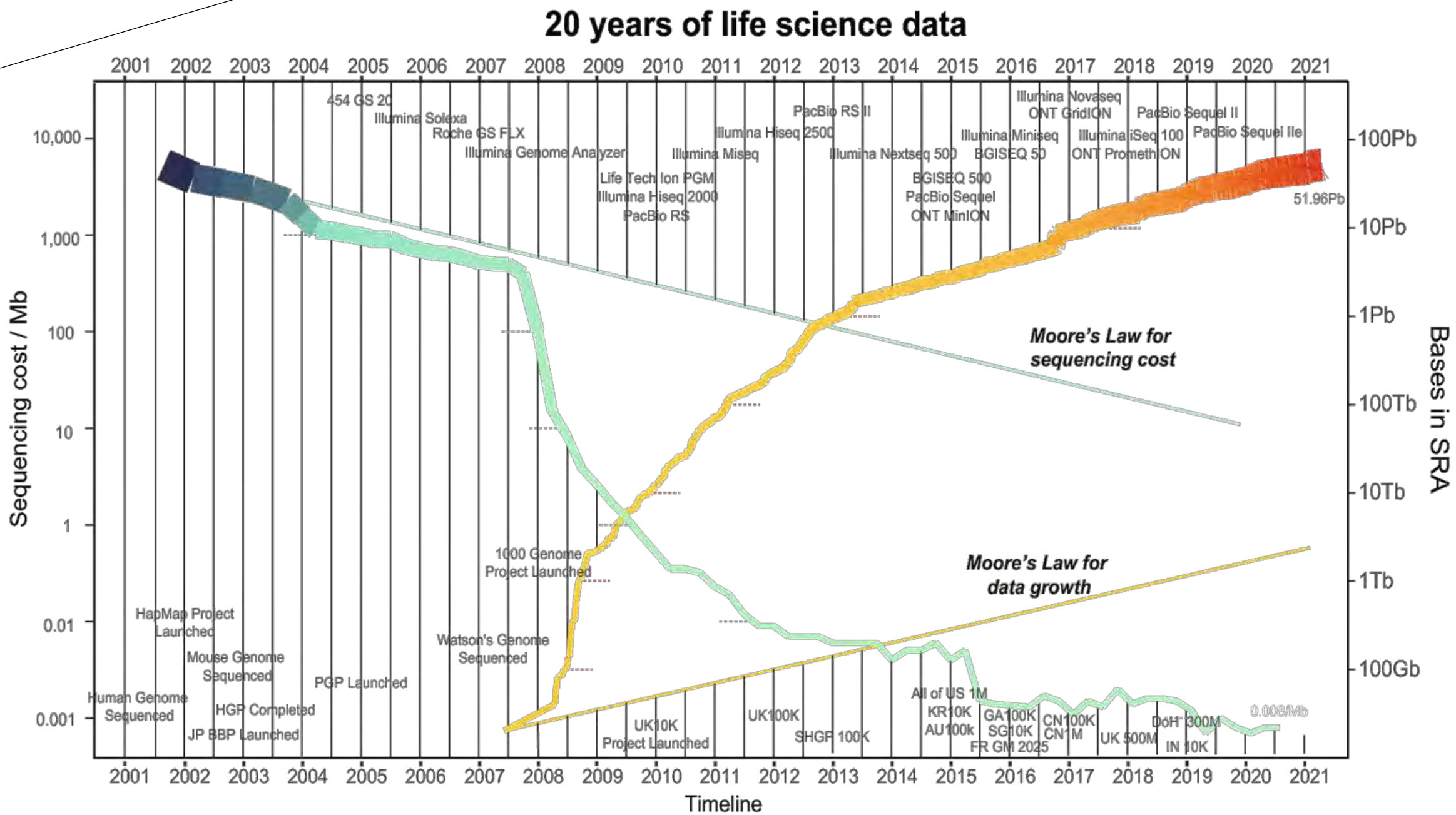


doi.org/10.1038/s41380-021-01201-2

WHERE TO NEXT...



doi.org/10.1038/s41380-021-01201-2



From: [Applications and challenges of high performance computing in genomics](https://doi.org/10.1007/s42514-021-00081-w)
doi.org/10.1007/s42514-021-00081-w

Thank you!

@dbkappel

KappelD@cardiff.ac.uk

Louis Wain (1860 – 1939)

Centre for Neuropsychiatric Genetics and Genomics

Canolfan Geneteg a Genomeg Niwroseiciatrig

The Academy of Medical Sciences

Leyden Delta

Magna Laboratories

UiO : University of Oslo

PGC Psychiatric Genomics Consortium

REALMENT

SUPERCOMPUTING WALES
UWCHGYFRIFIADURA CYMRU

arcca

Antonio Pardiñas
Sophie Legge
Issie Willcocks
Leon Hubbard
Eilidh Fenner
Elliott Rees
Adrian King
Marinka Helthuis
John Jansen
Kevin O'Connell
Robert Smith
Espen Molden
Ole Andreassen
Peter Holmans
Michael O'Donovan
Michael Owen
James Walters

Dr Rosemary Francis (Chief Scientist HPC, Altair)

Ten Ways in Which Altair is Saving the Planet with HPC

Abstract: Most of the HPC community know Altair for HPC infrastructure products such as PBS Professional or Altair Grid Engine, but most of our business is in the manufacturing and simulation space as a user of HPC. In this presentation I'll be looking at ten ways in which Altair and our customers are saving the planet through reducing the environmental cost of manufacturing, increasing product life spans, and using HPC in place of expensive real-life experiments. The quest for net zero needs to be done at every level so I'll also be talking about what you can do in the datacentre to increase efficiency, reduce power consumption and increase the life span of your hardware. HPC is usually cost efficient and power efficient compared with the alternatives, but we owe it to the planet to ensure that HPC is as green as possible and that every gram of carbon emissions count.

Bio: Dr Rosemary Francis founded Ellexus, the I/O profiling company, in 2010, and Ellexus was acquired by Altair in 2020. Rosemary obtained her PhD in computer architecture from the University of Cambridge and worked in the semiconductor industry before founding Ellexus. She is now chief scientist for HPC at Altair, responsible for the future roadmap of workload managers Altair® PBS Professional® and Altair® Grid Engine®. She also continues to manage I/O profiling tools, Altair® Breeze™ and Altair® Mistral™ and is shaping analytics and reporting solutions across Altair's HPC portfolio. Rosemary is a member of the Raspberry Pi Foundation, an educational charity that promotes access to technology education and digital making. She has two small children and is a keen gardener and windsurfer.

ALTAIR

TEN WAYS IN WHICH ALTAIR IS SAVING THE PLANET WITH HPC

Dr Rosemary Francis, Chief Scientist HPC

About Altair

$459M
FY19
Revenue

3,000+
Engineers, Scientists,
and Creative Thinkers

1985
Founded and
Headquartered in
Troy, MI U.S.

11,000+
Customers
Globally

86
Offices in
25 Countries

150+
Altair and Partner
Software Products

ALTAIR

Altair

Transforming the Future Through Computational Science and Artificial Intelligence (AI)

Transform decision making with simulation, data analytics, and high-performance computing.”

Jim Scapa, Founder and CEO

Sustainable product design

Altair® simulation has reduced many **millions of tons** of CO2 emissions

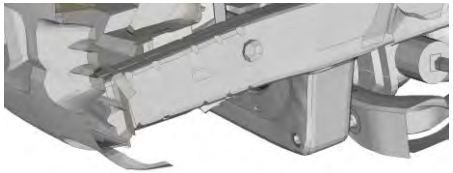
- Manufacturing materials and processes
- Transportation and logics
- Product life span and fault tracking

Crash & Safety

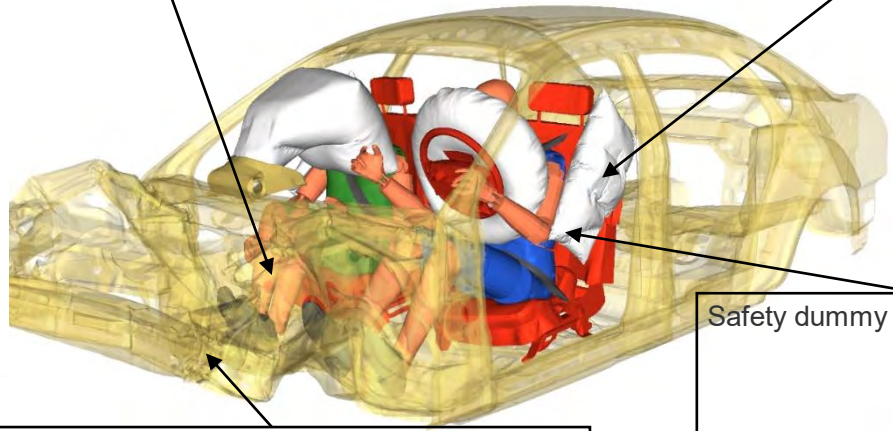
Health and safety is critical to sustainability

Simulation is more sustainable than physical testing

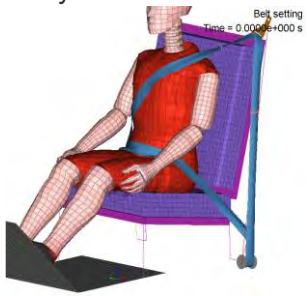
Failure risk assessment:



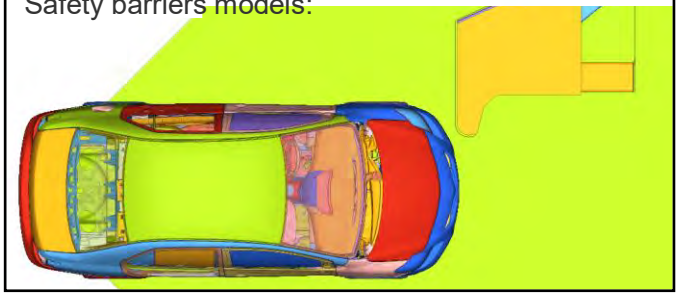
Airbag Folding & Deployment:



Safety dummy models:



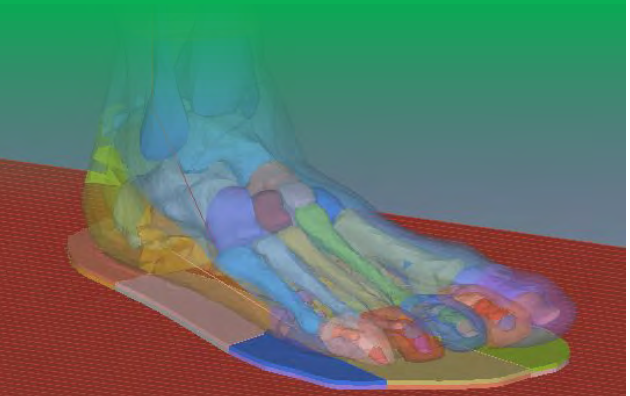
Safety barriers models:

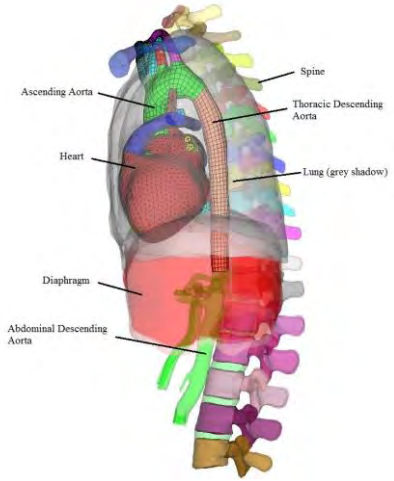


ALTAIR

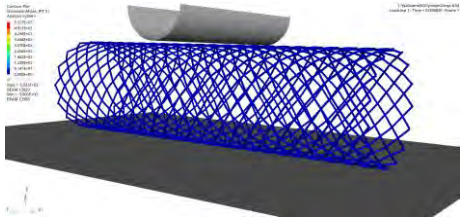
Biomechanics

Improving safety with accurate models

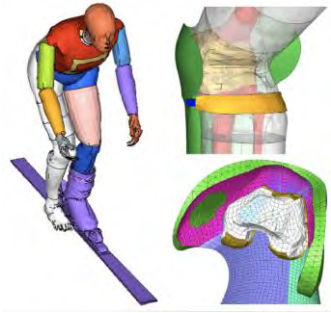




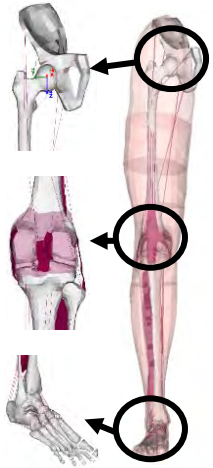
LBA Heart Model

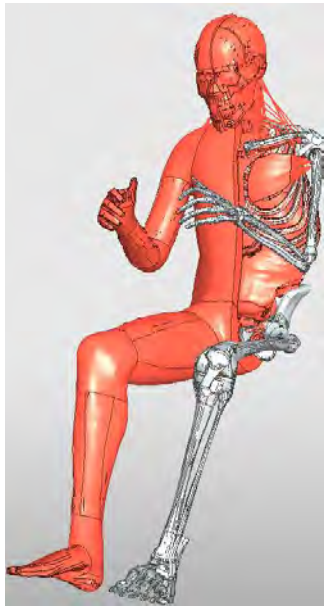


Stent with Shape Memory Alloy material (Nitinol)



LBA Lower Limb Model for Safety (LLMS)



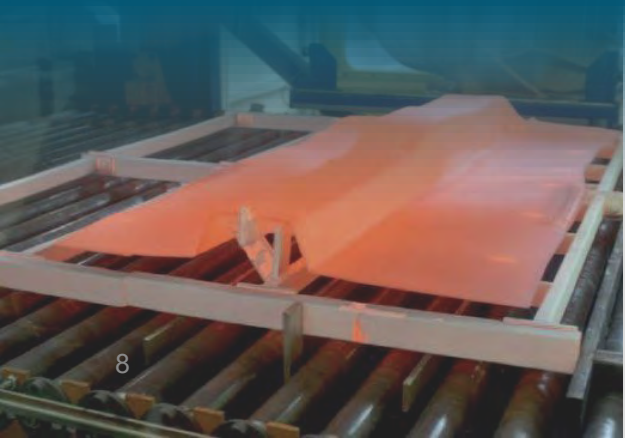


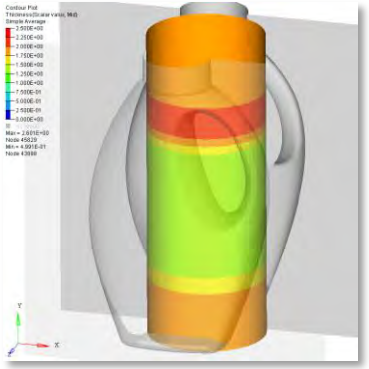
HBM - THUMS AM50 v6.1

ALTAIR

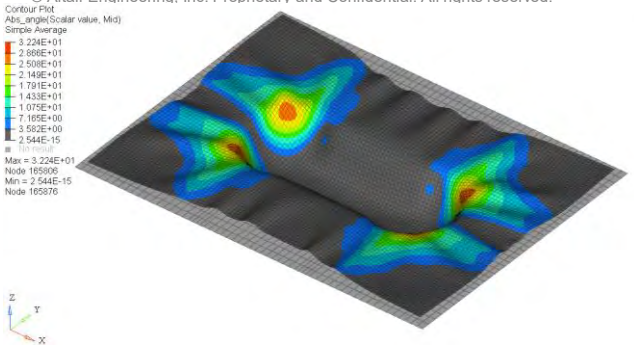
Manufacturing

Choosing more sustainable processes and materials



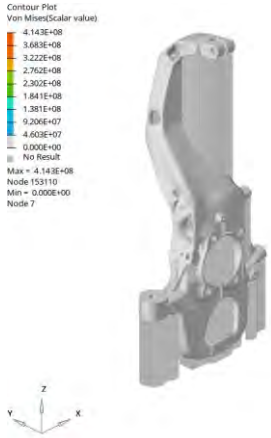


Blow molding

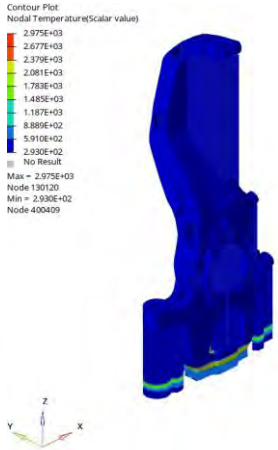


Composite forming

Residual deformation



Stresses



Temperature

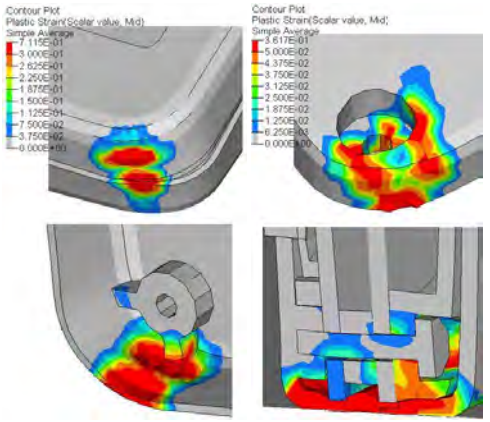
MAESTRO
Horizon 2020 project

Additive Manufacturing for knuckle part

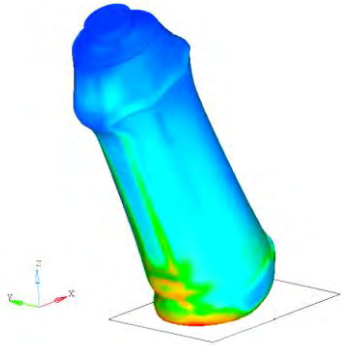
ALTAIR

Drop and Impact

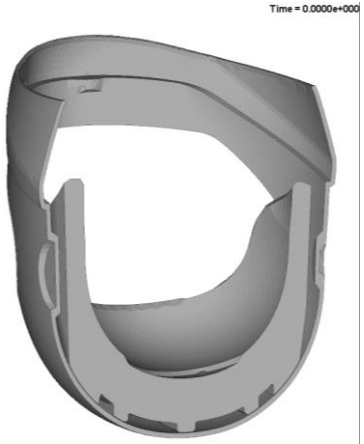
Increasing the product life span



Cellphone



Bottle



Moto helmet

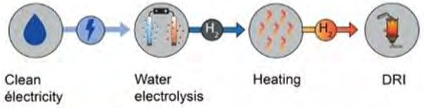
ALTAIR

The Altair Enlighten Award

Great Minds Think Light

Nemak, Recycled Materials for Sustainable Manufacturing & Products, Sustainable Materials Winner

Hydrogen DRI
H_2 use for DRI production



ArcelorMittal, First Industrial Large Scale H2 DRI Test, Sustainable Process Runner-up

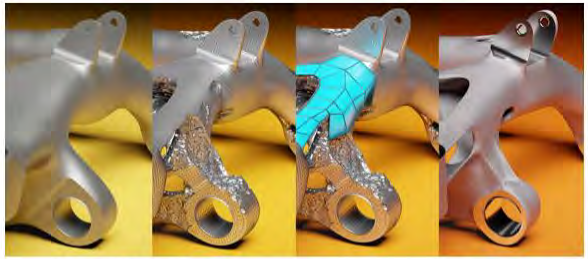
Lacks Enterprises, Composite Wheel Technology, Sustainable Product Runner-up

Adient Design & Technology, UltraThin Seating, Future of Lightweighting Runner-up

Nucor Corporation, Econiq – The World's First Net Zero Carbon Steel at Scale Sustainable Material Runner-up

Bocar Group, Tundra Rear End Post – Lightweight Design, Module Runner-up

Human Horizons, Integrated HPDC rear cabin, Enabling Technology Runner-up



Bionic Mesh Design, Lightweight Design without Compromises, Enabling Technology Runner-up

csi entwicklungstechnik GmbH, AMC, BMW M, DITF, NaMiKo – Biobased Automotive Center Console, Future of Lightweighting Honorable Mention

Yanfeng International, IP PAB Integration, Future of Lightweighting Runner-up

The Altair Enlighten Award

Great Minds Think Light

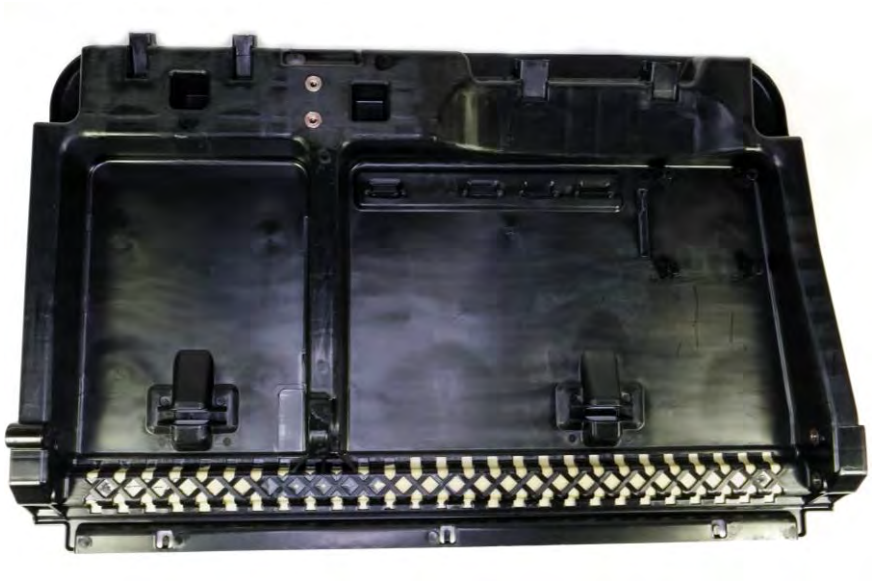
- **Sustainable Product**
  - **Ford Motor Company**
    - 100% Post-Consumer Recycled (PCR) Ocean Plastic Wiring Harness Clips (PA6)
    - 2022 F-150 Lightning

ALTAIR

The Altair Enlighten Award

Great Minds Think Light

- **Module Lightweighting**
  - **BASF Corp., Toyota, and L&L Products** Toyota Tundra Second Row Seat Structure



ALTAIR

In 2020, electric and hybrid models accounted for just 2% of new car sales.

We need to stop selling petrol cars this year to meet the <2C goals

ALTAIR

Electric vehicles

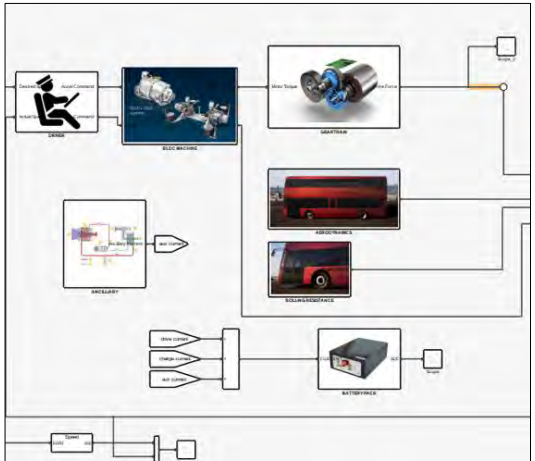
At Altair we work on

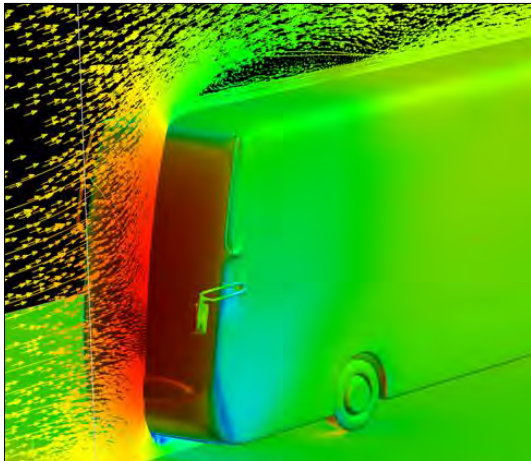
- Battery technology
- Lightweighting and manufacturing
- Vehicle range
- Charger technology
- Charging infrastructure
- Adoption and incentive efficacy

SWITCH

Worked with Switch Mobility to create a digital twin of electric buses to **predict range in real world use conditions**.

Accurate range prediction gives **confidence to transport authority clients**, helping them to meet their green mobility objectives.

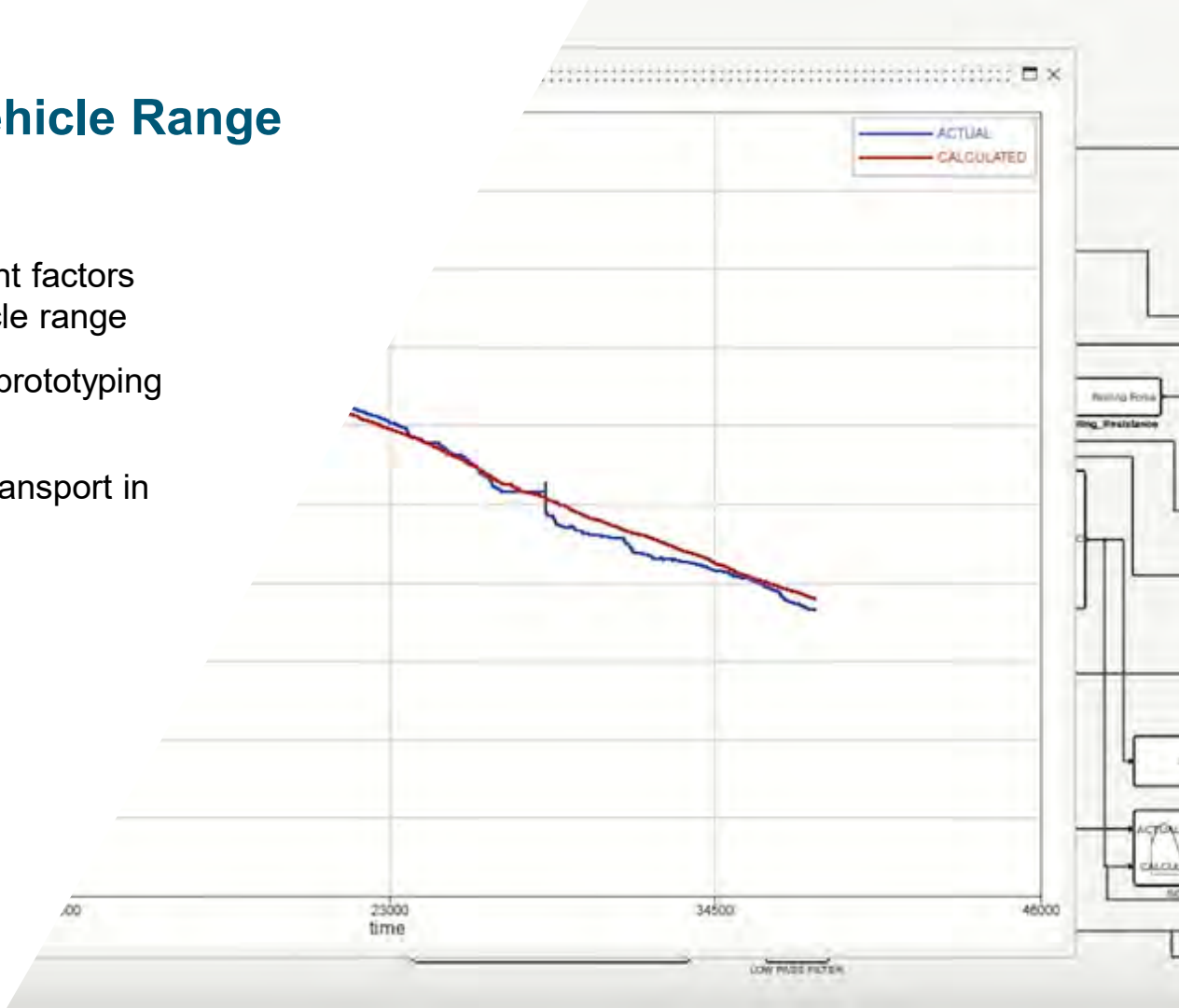




Predicting Electric Vehicle Range

Through Digital Twins

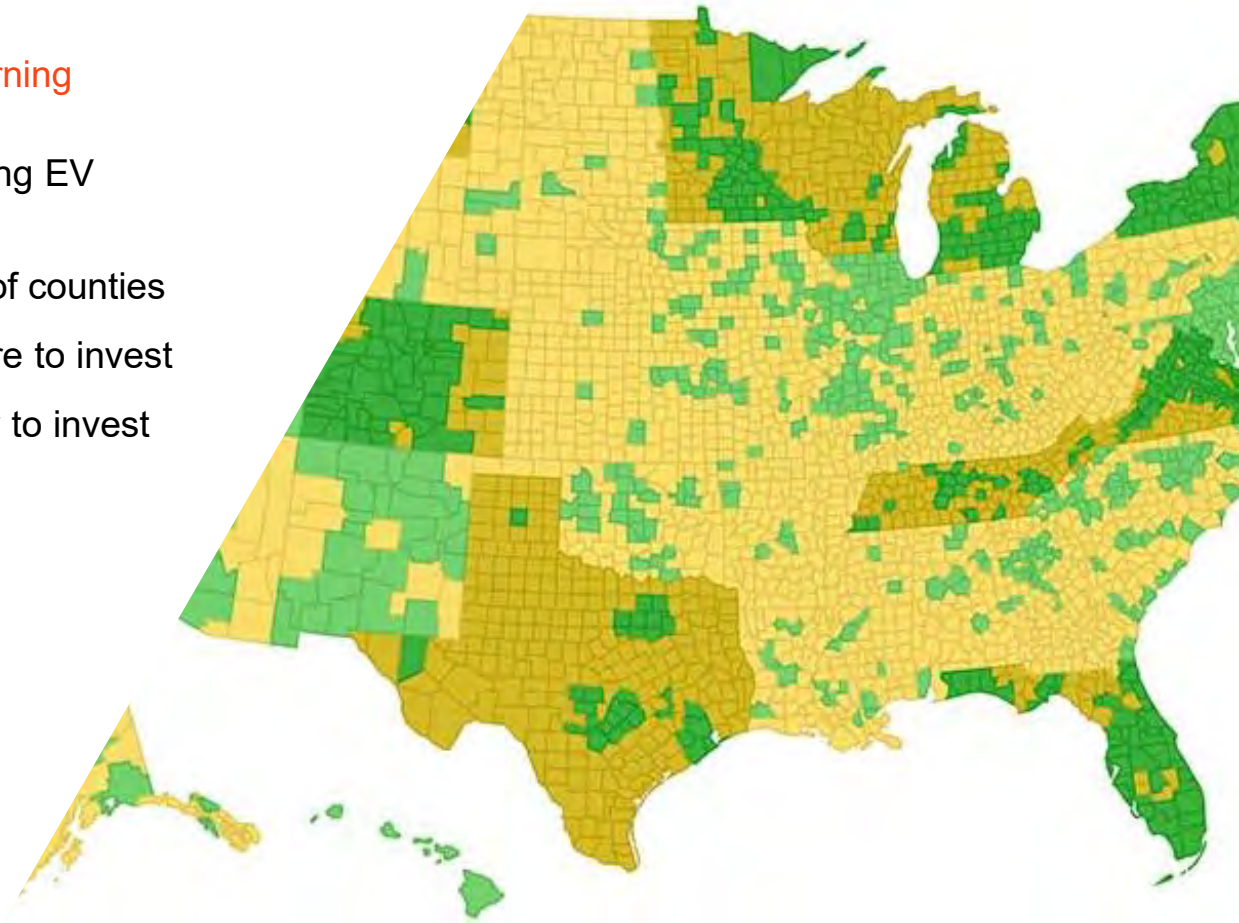
- **Switch** looked at how different factors affected battery life and vehicle range
- Provided confidence prior to prototyping and physical testing
- Proved the feasibility of EV transport in London



Electric vehicles

Driving adoption with machine learning

- Altair looked at factors predicting EV adoption in US counties
- Data was collected from 15% of counties
- Results show businesses where to invest
- Results show government how to invest in infrastructure and incentives



GREENER HPC

ALTAIR

Altair is the leader in HPC

- 19 of the top 20 **Automotive** companies
- 24 of the top **Aerospace** companies
- 27 of the top **Weather / Earth Sciences** organizations
- 15 of the top **Life Science** companies
- 11 of the top **Oil & Gas** companies
- 21 of the top **Government and Research** Organizations
- 3 of the top 4 **IaaS Cloud** providers
- 2 of the top 4 **IaaS Cloud** providers for their own Semiconductor Design

ALTAIR

Efficient scheduling for HPC and Cloud

Altair HPC and Cloud software is used to schedule **billions of core-hours** per year.

Today's high-performance processors can consume **more than 200 watts of power** per processor

ALTAIR

Efficient scheduling for HPC and Cloud

Advanced job scheduling can **improve system utilization by 15%**

Applying power profiles has been shown to **decrease power consumption by up to 18%**

ALTAIR

Green scheduling with NCAR and PBS Professional

- NCAR-Wyoming Supercomputing Center, HPE Cray EX system
- 19.87 petaflops with a combination of CPU and GPU nodes
- Funded by the National Science Foundation (NSF)
- **Green scheduling and energy aware scheduling** with Altair PBS Professional and Altair Accelerator Plus

Cylc & Altair's PBS Professional™

Weather Modeling at Australia's Bureau of Meteorology

- Bureau of Meteorology models the often-harsh natural environment: drought, floods, storms, and tropical cyclones throughout Australia.
- **Cylc** is an open-source Python workflow engine for cycling systems
  - executes tasks with detailed schedules and dependencies
  - used in climate modeling, weather forecast, data processing.
- The Cylc + PBS Professional integration unifies a large production system with many workflows

Maximising System Utilisation with Altair Grid Engine

Information Sciences Institute, Viterbi School of Engineering

- ISI have extensive use of machine learning, but suffered from low system utilization
- ISI chose Altair Grid Engine due to the built-in advanced GPU support, detailed documentation, ongoing product upgrades and customer support.

"With Altair Grid Engine, we have an infrastructure that schedules workloads to GPUs. We operate our infrastructure at 95% capacity with lower overall costs."

Stephen Rawls, Research Analyst

Maximising System Utilisation

Higher utilisation and lower total cost of ownership

- Support for GPUs, Containers and MPI technology
- Cloud bursting automation
- High system reliability and world-class support
- Support for heterogeneous systems

Ten ways in which Altair is saving the planet

1. **Reducing the environmental cost of manufacture** through simulation and optimisation
2. **Increasing product lifetime** with digital twin and AI
3. **Improving safety though simulation** with real human models
4. **Making electric vehicles a reality** though lightweighting and battery simulation
5. **Enabling innovative use of green materials and processes** such as recycled plastics

Ten ways in which Altair is saving the planet

6. **Power-aware and green scheduling**
7. **Improving system utilisation**
8. **Reducing cloud waste** though benchmarking, automation and budget tracking
9. **Solutions for climate modelling**
10. **Increase lifespan of hardware investment**

Dr Crispin Keable (Senior HPC Architect, Global HPC Strategic Sales, Atos)

Sustainability issues as we move towards exascale class HPC architectures

Abstract: High performance Computing is evolving from limited options with high costs towards flexible and simple computing resources available to end-users. At the same time with increased parallelism the journey towards Exascale entails numerous challenges. As energy costs spike, ignoring these costs is not an option. While an exascale system can consume the electricity of a small city, how to reconcile "performance" with "sustainability"?

A trade-off has to be made between achieving results, computational performance and energy consumption. "Greener" is as important as "faster", in defining the scientific value delivered by these systems. We need to consider user goals to reach net zero, and also looming carbon pricing as we evaluate what "Greener" really means. At the same time, more options are opening up to deliver scientific value using HPC in the cloud.

Bio: With over 30 year's experience in the HPC industry, I have worked through the evolution from proprietary supercomputers to Unix and then Open systems. While the development of technology is critically important, it is only there to do a job – forecast the weather, design new materials or drugs, build better cars or improve energy systems. I have always tried to keep this maxim at the forefront in my design choices.

Sustainability in Exascale computing

Crispin Keable

December 2022

1

Atos HPC, AI & Quantum
Our Vision

Help our client achieve **breakthrough science or business benefits** through HPC, AI and Quantum simulation

Design innovative, energy efficient on-premise and cloud based **High Performance Computing** (**HPC**) solutions

Give access to **Artificial Intelligence** (**AI**) through innovative solutions such as Atos Trusted SuperPODs

Atos decarbonization ambition
To reach "net-zero"

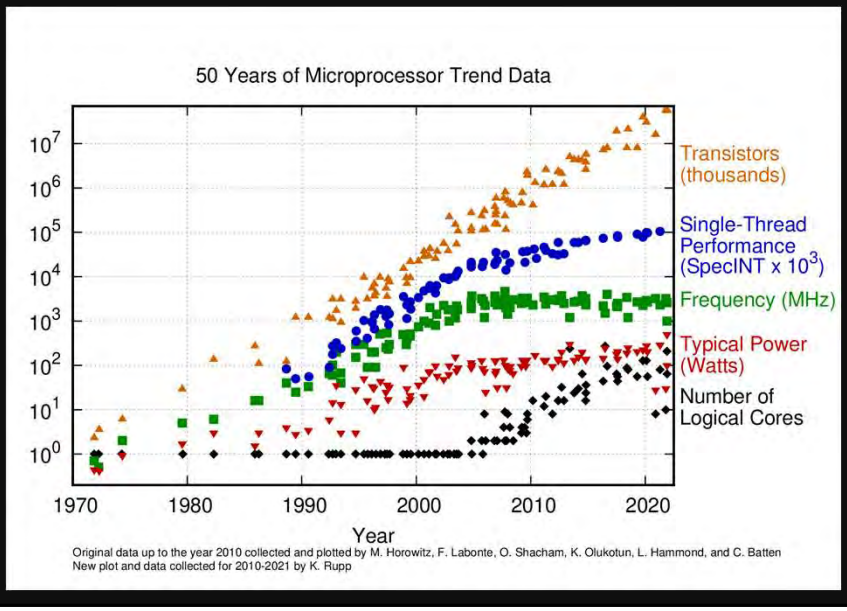
Step into the future with the **Quantum Computing** (**QC**) universal gateway

2

2

CPUs and GPUs are getting more powerful

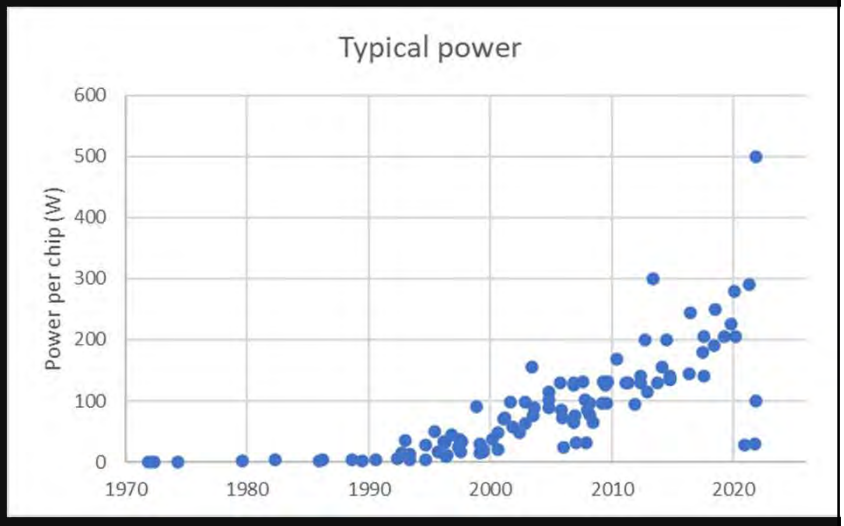
- Log graph shows the continuing Moores law climb
- Log graph
  - Great for showing exponential growth
  - Not so great for upward drift of CPU power used



3

Processor power is growing

- Switching to linear plot for power only, the growth in CPU and GPU power use is clear
- Narrowing the time window makes it even more clear
- Adding data for upcoming processors starts to look worrying!



4

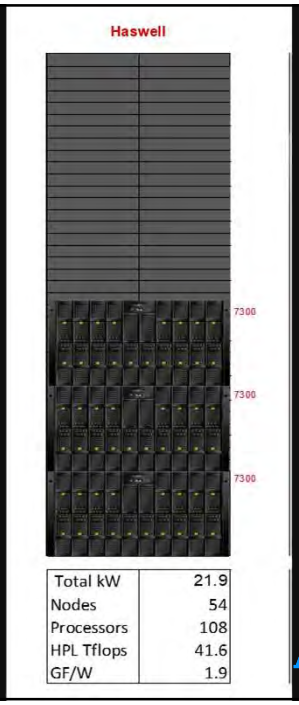
What can you put in an air cooled rack?

- Thinking back, HPC has been air cooled for decades
- Looking at typical HPC technologies in the last years, we consider
  - Haswell, with a typical rack configuration from 2014
  - Haswell, with water cooled doors
  - SkyLake, typical HPC rack (2018)
  - NVIDIA DGX1 – Ampere (2021)
  - NVIDIA DGX1 – Hopper (2023)
- Assume a DC rack limit of 20kW per rack to air
- GF per rack goes up, as does the GF/W
- However, rack density goes way down – ballooning DC costs

5

2014, Haswell ac

- Standard 19" rack, with high density 7U chassis
- Air cooled
- 18 dual socket nodes per chassis (Bull B500)
- Rack is 50% full



Total kW	21.9
Nodes	54
Processors	108
HPL Tflops	41.6
GF/W	1.9

6

2014 – Haswell, WCD

- 19" rack, with water cooled door
- Same Haswell compute chassis/nodes
- Bull B500 again

- Rack is 83% full

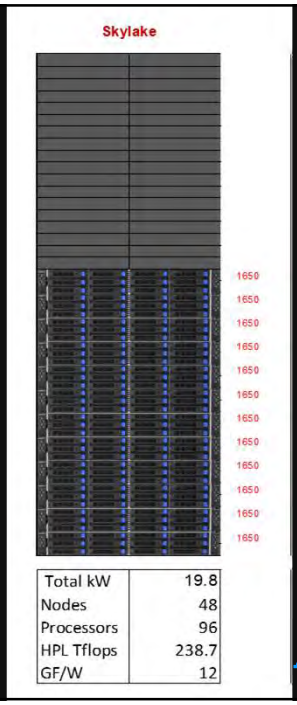
Total kW	36.5
Nodes	90
Processors	180
HPL Tflops	69.3
GF/W	1.9

7

2018, Skylake

- 19" rack, air cooled
- 2U, 4 node dual socket configuration
- Bull X400

- Rack is 57% full



Total kW	19.8
Nodes	48
Processors	96
HPL Tflops	238.7
GF/W	12

8

2021, NVIDIA A100

- 19" rack air cooled
- DGX1 – 8x A100 in 6U per node

- Rack is 43% full

9 | Dec 2022 – CIUK | © Atos

9

2023, NVIDIA next gen

- NVIDIA Hopper based DGX
- 6U, 8 GPUs per node

- Rack is 28% full

10 Dec 2022 – CIUK | © Atos

10

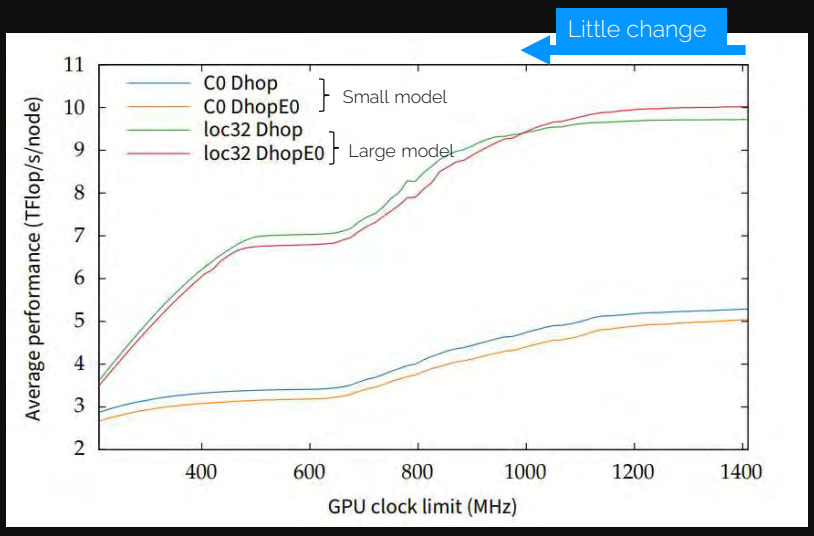
Edinburgh HPC energy use study

- Based on the 'Tursa' system, live in early 2022
- System is a BullSequana XH2000 configured with 112x A100-40 blades.
- Each blade is one node, with
  - 4x NVIDIA A100-40 GPUs
  - 2x AMD Rome
  - 4x NVIDIA InfiniBand 200Gbps networks
- High performance network is unblocked HDR, providing 800Gbps to the node, or 200Gbps to each GPU
- System is designed for maximum applications scalability for GPU enabled codes
- Apart from applications performance, optimising energy use is **critical**

11

Edinburgh energy study

- Using real application code to measure energy use
- This study runs the code, while measuring energy use per job
- Study maps performance against clock rate of A100
- Very little performance gain above 1GHz
- Translating to cost, clocking back A100 saves ~20% energy cost for virtually no loss of science
- Thanks to Antonin Portelli, Edinburgh University



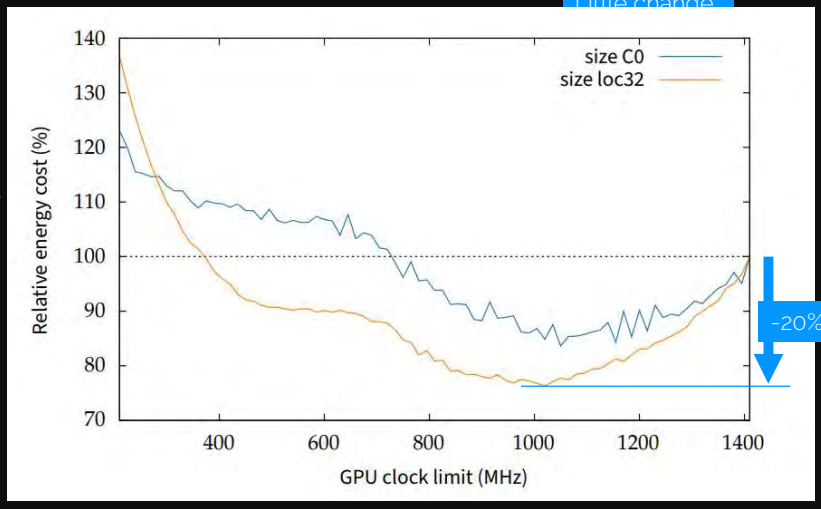
https://zenodo.org/record/7057319#.Y2okWXbP2Um

12

Edinburgh energy study

- Using real application code to measure energy use
- This study runs the code, while measuring energy use per job
- Study maps performance against clock rate of A100
- Very little performance gain above 1GHz
- Translating to cost, clocking back A100 saves ~20% energy cost for virtually no loss of science

- Thanks to Antonin Portelli, Edinburgh University

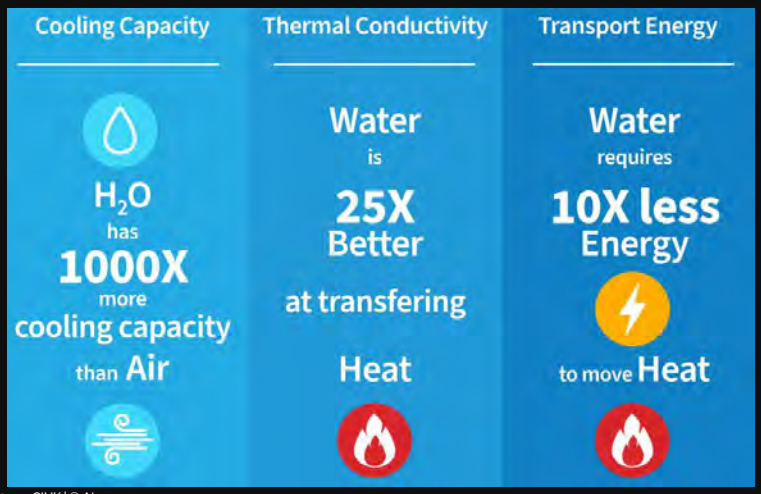


https://zenodo.org/record/7057319#.Y2okWXbP2Um

13

Overview of cooling technologies
Using water instead of air for components cooling

- Atos has offered Direct Liquid Cooling servers since 2011



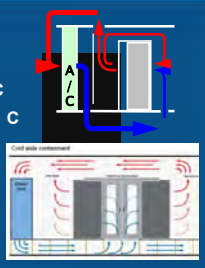
Atos continues to investigate other technologies such as liquid immersion cooling, which may be needed in the future as processor power continues to climb

14

Impact on PUE and DC operating costs (Tier2 assumption)

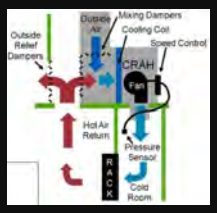
Air-cooled
5-20 kW/rack
Room 20° C
A/C water 7-12° C

PUE 1,6-2



Air-cooled with inrow
10-30 kW/rack
Room 20° C
A/C water 7-12° C

PUE 1,4-1,6

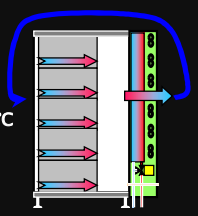


Free cooling air/air
In container or next generation datacenter
Room up to 35°C

PUE < 1.1

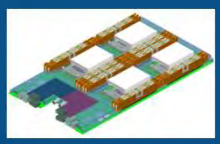
Water-cooled doors
40 kW/rack
Room 23°C
Water 12-18°C

PUE 1.3-1.5



DLC Water Block
80 kW/rack
Room up to 27°C
Water 38-47°C

PUE 1,2-1,3



Full DLC
150 kW/rack
Room up to 27°C
Water 38-47°C

PUE < 1.1

15

5 MW Cost Study comparing Atos HW technology

PUE, energy use, and CO2 are key factors in the cost of a large-scale HPC/AI solution

Based on 4000 X400 or XH3000 nodes with the following assumption :

- Power consumption 1250 W/nodes
- Price 7500 € per node
- X400 CooliT nodes are limited to 60kW 12 double-twin per rack,
- air solution is based on 20kW/rack datacenter
- 4-years TCO

5 MW - 4000 compute nodes	Full DLC system XH3000	DLC water block system X400 CooliT	Air system X400 air
Rack number	42	83	250
PUE	1.1	1.3	2
IT estimated cost	€ 42,416,667	€ 32,500,000	€ 30,750,000
Infrastructure	€ 4,166,667	€ 8,333,333	€ 25,000,000
Cooling	€ 1,000,000	€ 3,000,000	€ 4,000,000
Total CAPEX	€ 47,583,333	€ 43,833,333	€ 59,750,000
CAPEX gbp	£ 41,557,496	£ 38,282,387	£ 52,183,406

	€/KWH	XH3000	X400 CooliT	X400 air
Finland	0.04	€ 55,292,133	€ 52,943,733	€ 73,766,000
France	0.25	€ 95,763,333	€ 100,773,333	€ 147,350,000
Germany	0.4	€ 124,671,333	€ 134,937,333	€ 199,910,000
United Kingdom	0.229	£ 62,335,667	£ 67,468,667	£ 99,955,000

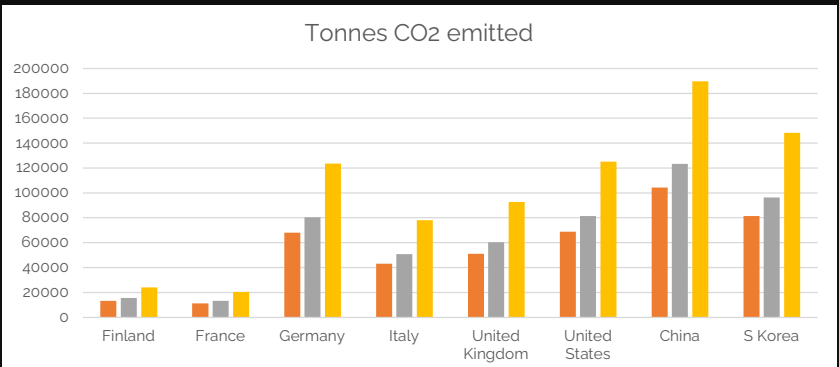
16

5 MW Cost Study comparing Atos HW technology

PUE, energy use, and CO2 are key factors in the cost of a large-scale HPC/AI solution

Based on 4000 X400 or XH3000 nodes with the following assumption :

- Power consumption 1250 W/nodes
- Price 7500 € per node
- X400 CooliT nodes are limited to 60kW 12 double-twin per rack,
- air solution is based on 20kW/rack datacenter
- 4-years TCO



	XH3000		X400 CooliT		X400 air	
	Infra + Power	Carbon	Infra + Power	Carbon	Infra + Power	Carbon
Finland	€ 55,292,133	€ 662,957	€ 52,943,733	€ 783,494	€ 73,766,000	€ 1,205,376
France	€ 95,763,333	€ 562,742	€ 100,773,333	€ 665,059	€ 147,350,000	€ 1,023,168
Germany	€ 124,671,333	€ 3,395,726	€ 134,937,333	€ 4,013,131	€ 199,910,000	€ 6,174,048
United Kingdom	€ 71,374,338	€ 2,548,722	€ 77,251,623	€ 3,012,126	€ 114,448,475	€ 4,634,040

17

BullSequana XH3000

6 Key Value Proposition Points

6 Secured by design

- Embedded security with hardware-anchored secure boot
- Vulnerability monitoring and threat anticipation

1 Ready for Accelerated Hybrid computing

- Most powerful computing technologies embedded
- Ready for GPU and QPU acceleration next

intel AMD NVIDIA SiPEARL

5 Open & Future-proof

- Standardized design for easy future evolutions
- OpenSequana program to seamlessly integrate 3rd party innovations

2 Any Scale

- "All-In-One" rack design
- Seamless scalability to ExaScale

4 Unrivaled efficiency

- 4th generation DLC technology with 40°C inlet water temperature
- Smart Power Management Suites
- Lowest TCO with free cooling and PUE ~1.0

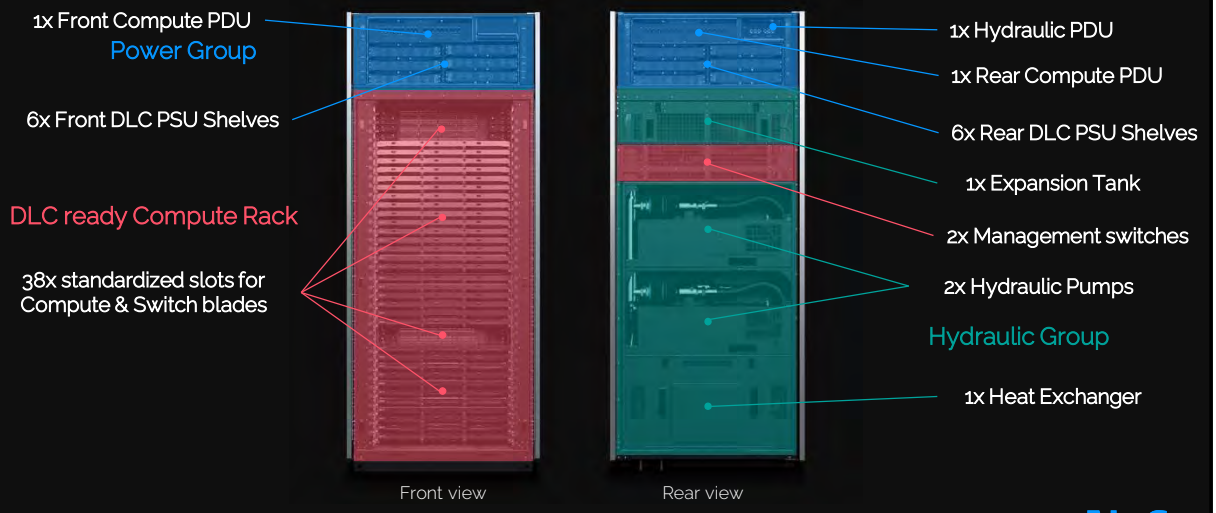
3 Flexible and dense by design

- Standardized slots to mix and interchange compute and switch blades
- Large choice of computing and interconnect technologies
- Industry-leading density

18

Infrastructure overview

Architecture Overview



19

NetZero[1] | Context

we share the same challenges

1

Economic

40[2] billion tCO_2

Get to Net Zero

2

Social

3

Legal

European Commission

EU TAXONOMY

ERP COMPLIANT

ECODESIGN

GREEN 500 level 2/3

EEHPC WG: Power Measurement Meth.

[1] On 28 October,, the Science Based Targets Initiative (SBTi) released the Net Zero Standard, which represents the first global standard for corporate Net-Zero target-setting in line with science

[2] Energy Transition Commission: Mind the Gap: How Carbon Dioxide Removals Must Complement Deep Decarbonization to Keep 1.5°C Alive (March 2022)

20

Sustainability | A holistic approach

everyone has a role to play

CONSIDERATIONS

Atos impact is possible across **the entire lifecycle:**

1. Eco **DESIGN**.. designing Green(er) IT. In some cases, the Green IT can be stand alone technology that can be leveraged as part of an IT for Green solution for our customers
2. Eco **MANUFACTURING & SUPPLY CHAIN**.. reduce energy consumption and waste in the Angers factory and influence supply chain behavior
3. Eco **OPERATIONS**.. influence and incentivize client behavior with new "business" elements that can in included in our Decarbonization Level Agreements (i.e. reduction on non-green energy use, DC PuE improvements, ecoact offsets...)

21

Frugal HPC

Our DNA, Green HPC Platforms

Sustainability with a holistic approach:

- REQUIREMENTS & DESIGN
- TENDER
- FACTORY
- DELIVERY
- PRODUCTION & EOL

TODAY

1/ Generation of compute... **less energy lost**
Direct Liquid Cooling with 40°C water at the inlet

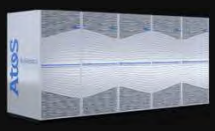
2/ **Understand** compute consumption
Simulate peak and operational consumption phase
Accurate modeling of production performance

3/ Effective use of compute... **less energy used**
Profile and optimize application behavior to minimize its energy footprint by optimizing runtime (MPI. etc.) and I/O flows (FastIO Library, SBB/SBF)

Measure energy and temperature metrics from all components of the supercomputer and link them with job execution information (provide per-job information)

Policy driven control and capping of power consumption. Static rules today, tomorrow "tight coordination" with resource planning

4/ **Ensure Availability** of compute
Predictive maintenance to reduce energy waste by optimizing execution and/or avoiding unproductive jobs



eco friendly design

software solutions to measure & control energy to solution

high availability supported by predictive maintenance

sovereignty via European Centric Supply Chain

22

From frugal to carbon neutral HPC

covering CO_2 footprint measurement, reduction, reporting and offsetting

Sustainability with a holistic approach:

Requirements & Design

Tender

Factory

Delivery

Production & EOL

Today

1/ 3rd generation of DLC compute blades…
Energy efficient design
Direct Liquid Cooling with 40°C water at the inlet

2/ **Understand** supercomputer consumption
Simulate peak and operational consumption phase
Accurate modeling of production performance
Automatic out-of-band data gathering in production

3/ Effective use of compute…
Optimize energy usage
Profiling tools (runtime, I/O, energy, etc.)
Live energy optimization (automated DVFS)
Power capping
Static rules today, tomorrow "tight coordination" with resource manager

4/ **Maximize solution availability**
Out-of-box monitoring (logging, events, etc.)
Predictive maintenance to reduce energy waste by optimizing execution and/or avoiding unproductive jobs

Tomorrow

5/ **Carbon Accounting**
Visibility on the Baseline CO_2 footprint impact on TCO (complete lifecycle supply chain, run, end of life)

European elements in our HPC TCO > 50% means reduction in supply chain CO_2 footprint related to transportation (70% by 2023/2024, 90/100% by 2030)

Provide CO_2 removal credits to offset the incompressible part of the platform carbon footprint

6/ Angers Factory, **carbon neutral** since 2017
Site Energy consumption, Transportation
Plastic free program, Waste generation…

7/ **Decarb Levers**, to reduce the CO_2 baseline
Build Green Co-innovation programs with Atos R&D teams to measure, reduce and report on CO_2 footprint

Leverage a set of existing – or develop new – "decarbonization levers" that unlock potential carbon footprint and energy savings

Provide user-land information (job-related) related to energy consumption and CO_2 footprint (end-of-job report, dynamic power optimizer, etc.)

CO_2 footprint per BOM component

Includes extraction, fabrication, transportation and end of life

Bureau Veritas EIME LCA database

40% to 60% of CO_2 total lifecycle footprint
(depending on energy source + system size)

23

From frugal to carbon neutral HPC

Innovation driving the next steps

Sustainability with a holistic approach:

Requirements & Design

Tender

Factory

Delivery

Production & EOL

Today

1/ 3rd generation of DLC compute blades…
Energy efficient design
Direct Liquid Cooling with 40°C water at the inlet

2/ **Understand** supercomputer consumption
Simulate peak and operational consumption phase
Accurate modeling of production performance
Automatic out-of-band data gathering in production

3/ Effective use of compute…
Optimize energy usage
Profiling tools (runtime, I/O, energy, etc.)
Live energy optimization (automated DVFS)
Power capping
Static rules today, tomorrow "tight coordination" with resource manager

4/ **Maximize solution availability**
Out-of-box monitoring (logging, events, etc.)
Predictive maintenance to reduce energy waste by optimizing execution and/or avoiding unproductive jobs

Tomorrow

5/ **Carbon Accounting**
Visibility on the Baseline CO_2 footprint impact on TCO (complete lifecycle supply chain, run, end of life)

European elements in our HPC TCO > 50% means reduction in supply chain CO_2 footprint related to transportation (70% by 2023/2024, 90/100% by 2030)

Provide CO_2 removal credits to offset the incompressible part of the platform carbon footprint

6/ Angers Factory, **carbon neutral** since 2017
Site Energy consumption, Transportation
Plastic free program, Waste generation…

7/ **Decarb Levers**, to reduce the CO_2 baseline
Build Green Co-innovation programs with Atos R&D teams to measure, reduce and report on CO_2 footprint

Leverage a set of existing – or develop new – "decarbonization levers" that unlock potential carbon footprint and energy savings

Provide user-land information (job-related) related to energy consumption and CO_2 footprint (end-of-job report, dynamic power optimizer, etc.)

Future

8/ **Global topic** supported by multiple groups
Green4BDS, dedicated R&D teams in BDS and at Atos Group level

9/ Increase **"green" energy** consumption
Local Green Energy with H2 cell designed for HPC

Model energy consumption profiles and optimize usage of smartgrid capabilities to manage peak loads

10/ Modular, extensible, adaptable **architectures**
Optimized heterogeneous scheduling across "sub clusters" with high variety of hardware (including shared memory models)

11/ **Data aware**, data usage efficiency (DuE)
Consider data path and data volume in scheduling policies linked to application profiles

24

Decarbonization Level Agreements | Benefits

25

Nimbix Supercomputing Suite
Global Elastic Supercomputing as-a-Service

3 Supercomputing "as-a-Service" Consumption Models

Elastic

Pay-as-you-go, on-demand, secure and scalable supercomputing through a single user interface.

Dedicated

Leverage powerful dedicated Bullsequana HPC servers as "Bare Metal as-a-Service" for the best of infrastructure and on-demand scalability, convenience, and agility.

Federated

Federated Supercomputing-as-a-Service offers a unified service console to manage all compute zones and regions in a public or private HPC, AI, and supercomputing federation

26

JARVICE XE software for Public, Private, or Hybrid HPC

Simplified User Experience

Simple point-click-run workflows on any infrastructure

Simplified Administration

Unified SaaS for HPC and Deep Learning

Unified platform for multi-cloud, multi-datacenter deployments

Containerized Application Distribution & Deployment

Platform-as-a-Service (PaaS) continuous integration and deployment for in-house algorithms or customization of commercial applications

Automatic synchronization with HyperHub Application Marketplace

Reduced Infrastructure Complexity

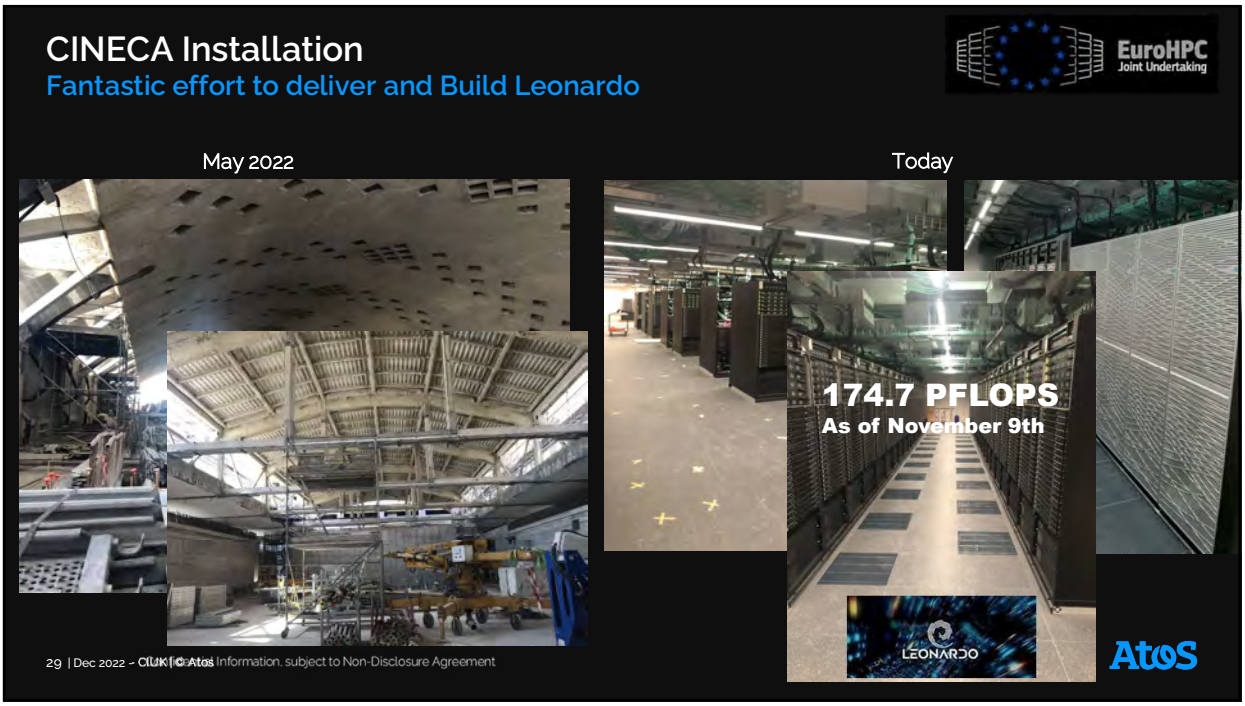
Unified infrastructure layer with Kubernetes

27

Summary

- Smart software tools are invaluable
- Systems and DCs need to be considered together
- Smart hardware design
- We need a holistic approach
- Sustainability considerations & Technology trends point in the same direction

28



29

30

Laura Foster (techUK)

Why is HPC integral to becoming a "science and technology superpower?"

Abstract: Through techUK's Future of Compute workstream, techUK members – a group of over 950 tech sector companies in the UK - have emphasised that HPC should be viewed as a key part of the UK's ambition to remain a world leader in science and innovation. In this sense, HPC should be seen as strategic national infrastructure, as important to our economic future in an information age as steel was in the industrial age. Lack of investment in this vital infrastructure could make the UK less competitive in both academia and key industries like life sciences, aerospace, and financial services.

Furthermore, there is a real danger that lack of investment in HPC will undermine growth in other technology ecosystems like artificial intelligence and quantum, both of which have been identified in current government policy as key pillars of the UK's science and technology ecosystem.

With this in mind, techUK will present the key themes from its Future of Compute work, such as technology convergence, international collaboration, and sustainability, before looking at how the UK tech sector envision the future of HPC in the UK. In doing so, we will ultimately explore how the UK tech sector, policy makers and academia in HPC could work together to support the UK's ambition of becoming a science and technology super power.

Bio: Laura is techUK's Head of Programme for Technology and Innovation.

She supports the application and expansion of emerging technologies across business, including Geospatial Data, Quantum Computing, AR/VR/XR and Edge technologies.

Before joining techUK, Laura worked internationally in London, Singapore and across the United States as a conference researcher and producer covering enterprise adoption of emerging technologies. This included being part of the strategic team at London Tech Week.

Laura has a degree in History (BA Hons) from Durham University, focussing on regional social history. Outside of work she loves reading, travelling and supporting rugby team St. Helens, where she is from.

techUK
FOR WHAT COMES NEXT

Why is HPC integral to becoming a “science and technology superpower?”

techUK

1. How HPC can enable the UK to become a science and technology superpower
2. How to ensure HPC achieves this vision

techUK

Who we are

Our purpose

techUK champions technology's role in preparing and empowering the UK for what comes next, delivering a better future for people, society, the economy and the planet.

Championing tech

techUK

20+
Programmes exploring tech markets, policy and innovation

25+
Industry reports providing thought leadership each year

50
National trade association partners across the world

400+
Member events created by techUK each year

950
Member companies countrywide including 500 SMEs

2,500
Attendees for techUK events each month

71K
Visits to our website each month

700K
Employees represented across our network

techUK

How does high-performance compute help achieve the UK's vision of becoming a science and technology superpower?

What does HPC enable?

techUK

Unlocking the value of data

Enabling access to innovation for SMEs

Pushing forward emerging technologies

Sustainablity

Industry, academia and Government working together!

techUK's Future of Compute

To unleash Britain's potential

Panel of four academics and industry specialists appointed to support review led by Professor Zoubin Ghahramani.

Four experts have been appointed to support an external review of the UK's advanced computing capabilities, as part of plans to turn the country into a science and technology superpower.

d during London Tech Week, is being led by the Professor of Information Engineering at oubin Ghahramani, who is a world-leading expert in the field of machine learning.

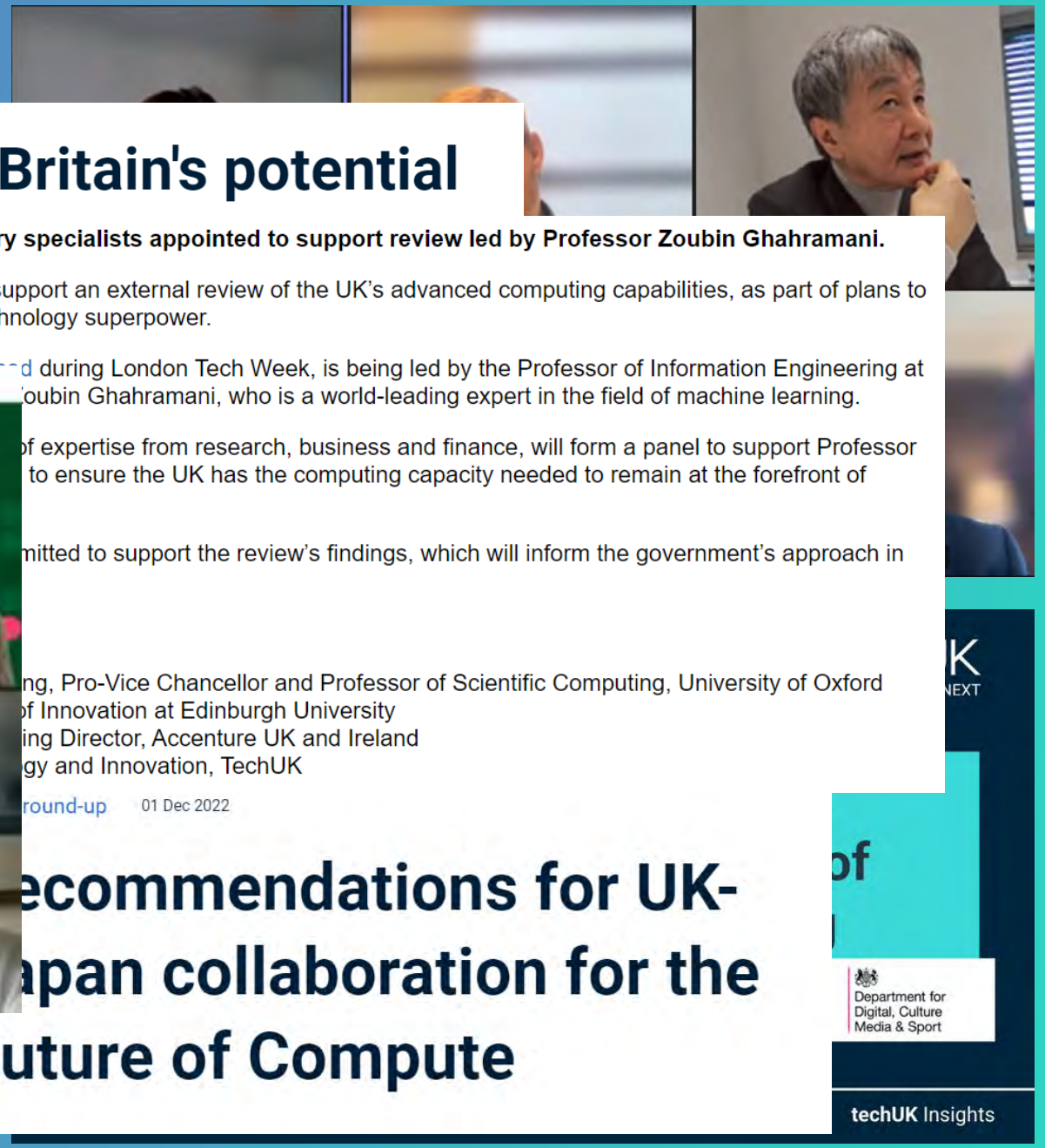
f expertise from research, business and finance, will form a panel to support Professor to ensure the UK has the computing capacity needed to remain at the forefront of

mitted to support the review's findings, which will inform the government's approach in

ng, Pro-Vice Chancellor and Professor of Scientific Computing, University of Oxford
f Innovation at Edinburgh University
ing Director, Accenture UK and Ireland
gy and Innovation, TechUK

round-up 01 Dec 2022

ecommendations for UK-apan collaboration for the Future of Compute



techUK

Thank you!

CIUK 2022 Cluster Challenge

The third edition of the CIUK Student Cluster Challenge saw the highest entry so far with six teams taking part. Teams from the Universities of Birmingham, York, Durham, Bristol (two teams) and UCL took on a series of online challenges leading up to the conference, followed by a series of face-to-face challenges during CIUK. The first ever cluster challenge champions from Durham were back to try to regain their title from last year's winners Bristol.

The three online challenges saw Team ClusDur from Durham take three straight victories for a maximum of thirty points. The two Bristol teams occupied second and third place in the first two challenges with York, UCL and Birmingham all picking up points.

The face-to-face challenges during CIUK saw a change of fortune with both York and Birmingham picking up wins in the first two challenges and Team ClusDur coming in last place both times! This meant that the overall leader board was now extremely close going into the final two challenges. However, any hopes that the chasing pack had of catching Team ClusDur disappeared as they found their form again at the right time to take the wins in both of the last two challenges and with it the overall title of CIUK 2022 Student Cluster Champions.

Massive congratulations go to all of our teams for making the competition so exciting and, of course, to our partner companies for engaging with the competition, providing access to their systems and for mentoring the teams throughout. The competition would not be possible without them so thank you Alces Flight, Boston, Graphcore, OCF and Q Associates.

Well done to Team ClusDur from Durham University who will now go forward to represent CIUK in the ISC'23 Student Cluster Challenge Competition.

alcesflight

BOSTON
Servers | Storage | Solutions

GRAPHCORE

OCF

Q ASSOCIATES
Specialist IT Services

ENERGETIC: A workshop regarding Energy Benchmarking on Heterogeneous Systems

Organisers: Oliver Thomson Brown, Michael Bane, Teymoor Ali, Jamie Quinn, Deepayan Bhowmik, David Stansby

On Friday 2nd December 2022, the ENERGETIC project held a discussion workshop collocated with Computing Insight UK (CIUK) in Manchester on 1-2 December 2022 [CIUK]. There were around 25 attendees to the ENERGETIC workshop, with UKRI Digital Research Infrastructure (DRI) users, system administrators, and hardware vendors represented. After a short introduction, the workshop divided in to three groups, each focused on one discussion topic.

The three discussion prompts presented to participants (to discuss one per group) were:

- Energy advantage.
  - Some applications may consume less energy overall when run on one or more accelerators (i.e. GPUs, FPGAs). Which types of applications? Which accelerators?
- Fair comparison across architectures.
  - Energy benchmarking frameworks should be standardised to compare fairly across architectures.
- User-level energy benchmarking.
  - Energy usage is highly dependent on specific job, so users should be able to benchmark their own jobs' energy usage across UKRI DRI.

In spite of the disparate discussion topics, common themes appeared across all groups.

1. Standards for energy measurement across platforms are highly desirable.
2. Better tools are needed to support those standards, as well as support less experienced users.
3. There is a need for better education for UKRI DRI users on energy efficient computing.

There was broad agreement that standards for energy measurement across platforms are highly desirable. Since there are various valid choices concerning the specific implementation of energy measurement, it is very important that the chosen methodology is clearly documented. It was broadly agreed across discussion groups that better tools are needed. There are high-quality tools which support specific architectures, or even multiple architectures, but only one at a time. The user is responsible for implementing whole-node energy measurements (or as close to it as the available tools allow). The final theme which occurred across discussion groups was the need for education on all aspects of energy efficient computing. Users who wish to make their own scientific computing 'greener' may be unsure how to make appropriate measurements of their energy usage (in part due to the complexity of tools), how to interpret the data they may gather, or what to do about it once they have the data.

Once again we thank the organisers of CIUK 2022 for hosting the ENERGETIC workshop, and to all attendees for contributing to the valuable discussion.

CIUK'S FIRST WOMEN IN HPC BREAKFAST STARTS WITH HEART

For the first time the Computing Insight UK (CIUK) conference welcomed WHPC volunteers and chapter leaders together in what hopes to become an annual event. Over two hours CIUK badge holders networked, listened to talks, and engaged in lively discussion around the positive changes needed in order to bring High Performance Computing (HPC) to the next level in diversity and inclusion.

This year's distinguished speaker, Cristin Merritt of Alces Flight, framed the conference theme of sustainability in terms of people, skills, and change. Rather than pointing to a specific process of 'solving' the diversity and inclusion problem in HPC (and in the wider community of STEM) she utilised community input as well as her experience in attending and facilitating WHPC events to come up with a 'wall of ideas.'

"People are weird," Cristin said, "They can't have a patch installed on their network, or have their code modified. People will learn differently, behave sometimes irrationally, and cope with change and ideas in ways that defy any type of logic. This is why there isn't one specific program to learn, one key metric to gather data on, or one path to follow. You have to constantly be assessing what works and always keep edging towards improvement."

The second portion of the event was a panel and group discussion talking about everything from how chapters and affiliates work, to what it is like to volunteer. As there was a 50% attendance from our male supporters and allies, a robust and meaningful conversation around supporting women and underrepresented groups also took place. Many of our male supporters spoke of personal experiences including thoughts and feelings on the topic of diversity and inclusion with openness and honesty. We are incredibly grateful for their insight.

This event was possible due to the efforts of Georgina Ellis, Marion Weinzierl, Karen Stoneham, Cristin Merritt, and Elisabetta Boella. The breakfast was underwritten by Alces Flight and UKRI STFC. We are also grateful to Damian Jones for putting the foundation of this event in place.

You can download a copy of Cristin's presentation here: CIUK22_WHPC_CristinMerritt_Sustainability. Supporting community insight can be found here: WHPC – CIUK – Defining Sustainability. Please note, there are some statements within the insight that people may view as offensive. We have chosen to leave them in as a testament that there is still work to be done in progressing the field of HPC and STEM.

CoSeC Conference 2022

The Computational Science Centre for Research Communities (CoSeC) held its annual meeting in 2022 as part of the wider CIUK conference. This took place over the course of day 1 of CIUK and featured 9 submitted talks around diverse areas of computational science and engineering, as well as a chaired panel discussion featuring CoSeC staff. The conference was organised by a 9-person committee formed of CoSeC staff as well as those from the university sector directly involved in CoSeC activities.

The event was split into three main topics: Data Science and Machine Learning, Coupling and Data Workflows and High Performance and Future Computing. It saw presentations on various topics ranging from *"A Natural Language Processing (NLP)-Based Deep Learning Approach to Predict Solubility Parameters for Drug Discovery"* through to *"Fast pattern detection in kinetic Monte Carlo simulations of heterogeneous catalysis"*. The event was available to join remotely but all talks were delivered in-person. The recordings of these, as well as the chaired panel session can be found on the main CoSeC website (www.cosec.stfc.ac.uk) or as a playlist on STFC's main YouTube channel.

Stephen Longshaw (STFC) Introduces the CoSeC Conference organising committee at the CoSeC@CIUK 2022 event as part of the opening introduction

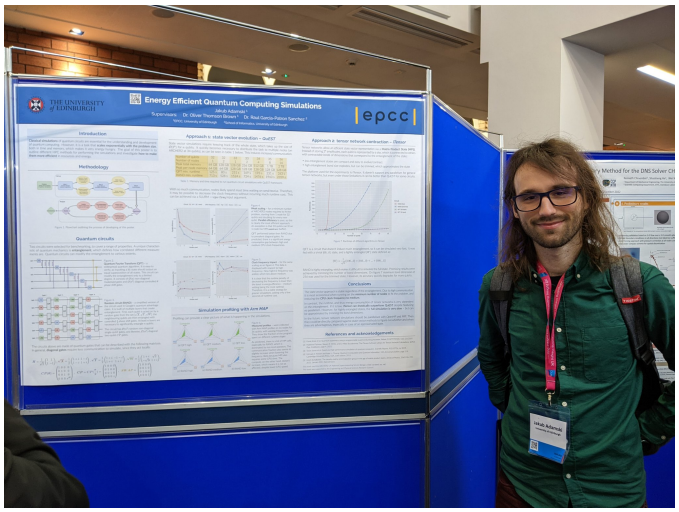
The event was hugely successful, with a great response from the communities that CoSeC supports and some thought-provoking conversation from those who attended. The work presented was all of a very high standard and really highlighted the diverse but inter-related nature of the UKS computational science communities. The event will return for 2023.

CIUK 2022 Poster Competition Winner

Energy Efficient Quantum Computing Simulations

Jakub Adamski
University of Edinburgh

As we are entering the era when quantum advantage becomes viable, it is especially important to push the boundaries of classical simulations of quantum computing. It involves running exponentially complex algorithms, so the use of high-performance computing is essential and entails huge energy consumption. The simulation can be performed via state vector evolution or by contracting a tensor network of matrix product states and operators. Each method offers different advantages, and allows potential optimisations to save energy. Various benchmarks have been set up and run on ARCHER2 to determine the most economical approach. It was found that by downclocking the CPU, a state vector simulation can consume up to 30% less energy. On the other hand, tensor networks proved exponentially more efficient when the entanglement was limited. The goal of this poster is to present and explain the benchmarking results, and encourage greener HPC use when simulating quantum computing.



Energy Efficient Quantum Computing Simulations

Jakub Adamski [1]

Supervisors: Dr. Oliver Thomson Brown [1] Dr. Raul Garcia-Patron Sanchez [2]

[1]EPCC, University of Edinburgh [2]School of Informatics, University of Edinburgh

Introduction

Classical simulations of *quantum circuits* are essential for the understanding and development of *quantum computing*. However, it is a task that **scales exponentially with the problem size**, both in *time and memory*, which makes it very energy hungry. The goal of this poster is to outline different HPC methods for performing the simulations and investigate **how to make them more efficient** in resources and energy.

Methodology

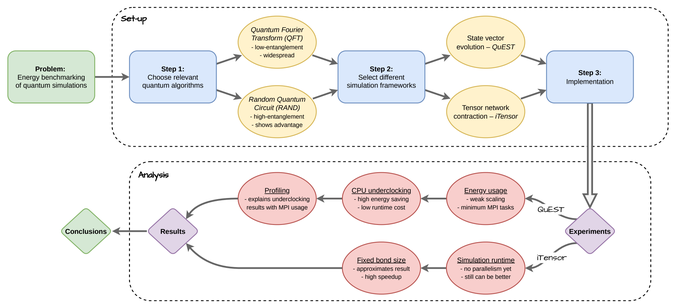


Figure 1. Flowchart outlining the process of developing of this poster.

Quantum circuits

Two circuits were selected for benchmarking, to cover a range of properties. A unique characteristic of quantum mechanics is **entanglement**, which defines how correlated different measurements are. Quantum circuits can modify the entanglement to various extents.

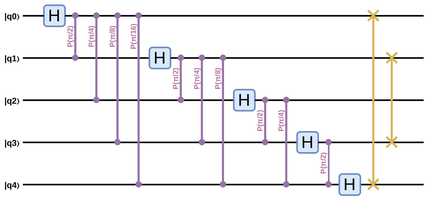


Figure 2.
Quantum Fourier Transform (QFT) – a widespread quantum algorithm. It is easy to verify, as inputting a $|0\rangle$ state should output an equal superposition of all states. This circuit can modify the entanglement only to a limited degree. It consists of $\mathcal{O}(n)$ non-diagonal *Hadamard gates* and $\mathcal{O}(n^2)$ diagonal controlled θ *phase shift gates*.

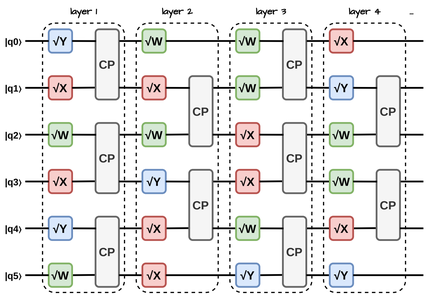


Figure 3.
Random circuit (RAND) – a simplified version of the circuit used in Google's quantum advantage claim. It is built of multiple layers that create entanglement. First, each qubit is acted on by a random gate from the set $[\sqrt{X}, \sqrt{Y}, \sqrt{W}]$; the neighbouring qubits are then entangled with controlled $\frac{\pi}{2}$ *phase shift gates*. At least n layers are necessary to significantly entangle n qubits.

The circuit has $\mathcal{O}(n^2)$ random non-diagonal single-qubit gates, and likewise, $\mathcal{O}(n^2)$ diagonal two-qubit CP gates.

The circuits above are made of quantum gates that can be described with the following matrices. In general, **diagonal gates** require less communication to simulate, since they *act locally*.

$$\boldsymbol{H} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}, \boldsymbol{\sqrt{X}} = \frac{1}{2}\begin{pmatrix}1+i & 1-i\\ 1-i & 1+i\end{pmatrix}, \boldsymbol{\sqrt{Y}} = \frac{1}{2}\begin{pmatrix}1+i & -1-i\\ 1+i & 1+i\end{pmatrix}, \boldsymbol{\sqrt{W}} = \frac{1}{2}\begin{pmatrix}1+i & \sqrt{2}\\ -\sqrt{2}i & 1+i\end{pmatrix},$$

$$\boldsymbol{CP(\theta)} = \begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&e^{i\theta}\end{pmatrix}, \boldsymbol{CP} = \boldsymbol{CP}(\frac{\pi}{2}) = \begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&i\end{pmatrix}, \boldsymbol{SWAP} = \begin{pmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{pmatrix}$$

Approach 1: state vector evolution – *QuEST*

State vector simulations require keeping track of the whole state, which takes up the size of $\mathcal{O}(2^n)$ for n qubits. It quickly becomes necessary to distribute the task to multiple nodes (on ARCHER2 at 34 qubits), as can be seen in table 1 below. This induces *increased communication*.

Number of qubits	32	32	33	33	34	35	36
Number of nodes	1	2	1	2	4	8	16
Peak total memory	64 GB	128 GB	128 GB	256 GB	512 GB	1 TB	2 TB
Peak per-node memory	64 GB	64 GB	128 GB	128 GB	128 GB	128 GB	128 GB
QFT min. runtime	125 s	80 s	251 s	169 s	193 s	231 s	263 s
RAND min. runtime	713 s	528 s	1524 s	724 s	1476 s	1943 s	2359 s

Table 1. Memory and time required to run quantum circuit simulations with QuEST framework.

With so much communication, nodes likely spend most time waiting on *send/receive*. Therefore, it may be possible to decrease the clock frequency without incurring much runtime cost. This can be achieved via a SLURM `--cpu-freq` input argument.

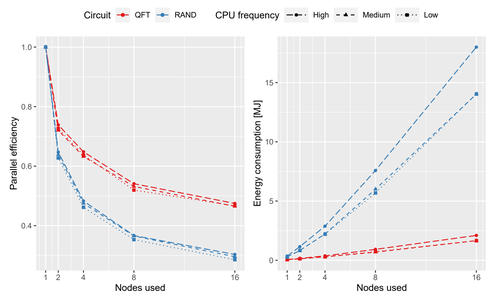


Figure 4.
Weak scaling – for a minimum number of ARCHER2 nodes required to fit the problem, starting from 1 node for 32 qubits and doubling for every new qubit. **Parallel efficiency** is poor, so this is clearly the most efficient approach. An exception is that 33 qubits can fit on 1 node (no MPI `sendrecv` buffer).

QFT performed better then RAND due to prevalent diagonal gates. As predicted, there is a significant *energy consumption gap* between high and medium CPU clock frequencies.

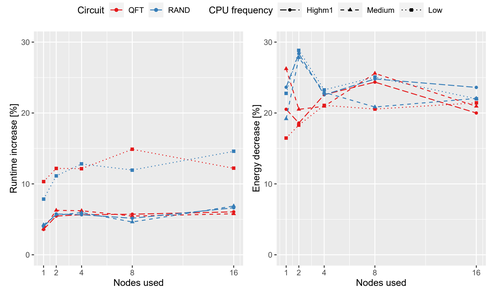


Figure 5.
Clock frequency impact – for the same scaling as on figure 4. The data is displayed with respect to high frequency. New highm1 frequency was added, which lies above medium.

It is clear that the *runtime penalty* of decreasing the frequency is *lower than the boost in energy efficiency* – medium setting being the most optimal. Therefore, it is a *viable strategy for greener simulations*, adding only a few seconds of runtime cost.

Simulation profiling with *Arm MAP*

Profiling can provide a clear picture of what is happening in the simulations.

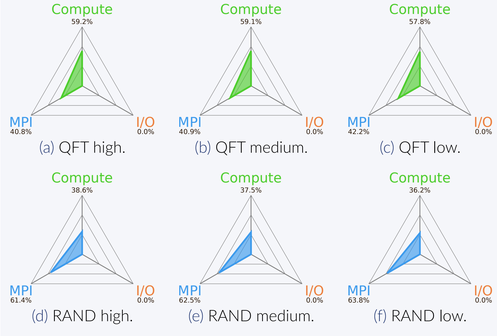


Figure 6.
Measured profiles – were collected with Arm MAP profiler on 16 nodes for 36 qubits, with varying frequencies. They show the fraction of the program spent on different *runtime stages*.

As predicted, there is a lot of *MPI calls*, especially for RAND, which is dominated by *non-local operators*. The communication fraction also seems to slightly increase when lowering the frequency, likely because MPI also requires some CPU time. The compute, on the other hand, doesn't increase, which means it is not as affected, despite lower CPU speed.

Approach 2: tensor network contraction – *iTensor*

Tensor networks allow *an efficient state vector representation* via a **Matrix Product State (MPS)**. Instead of storing 2^n amplitudes, each qubit is represented by a site, which is connected to others with *contractable bonds* of dimensions that correspond to the entanglement of the state:

- *low entanglement* states are compact and easy to evolve/contract
- *high entanglement* bond size explodes, but can be *trimmed*, which approximates the state

The platform used for the experiments is iTensor. It doesn't support any parallelism for general tensor networks, but even under those limitations it can be better than QuEST for some circuits.

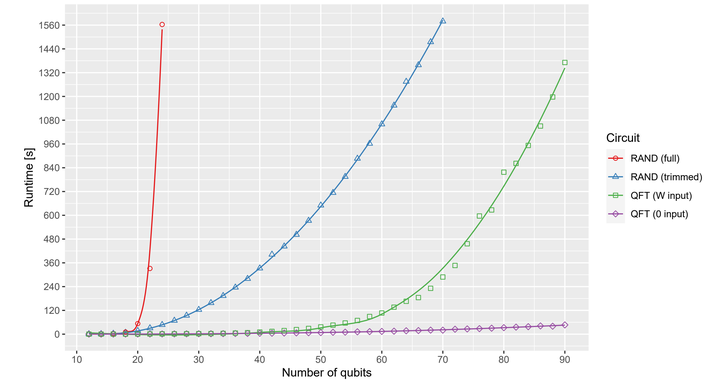


Figure 7. Runtimes of different algorithms in iTensor.

QFT is a circuit that doesn't induce much entanglement, so it can be simulated very fast. It was fed with a trivial $|00...0\rangle$ state, and a lightly entangled $|W\rangle$ state defined as:

$$|W\rangle = \frac{1}{\sqrt{n}}(|100...0\rangle + |010...0\rangle + ... + |000...1\rangle)$$

RAND is *highly entangling*, which makes it difficult to simulate the full state. Promising results were achieved by trimming the number of bond dimensions. On figure 7 *maximum bond dimension of 256* was used for the trimmed state. However, its accuracy quickly degrades for more qubits.

Conclusions

The state vector approach is *stable* regardless of the entanglement. Due to *high communication*, it is most economical when running on the **minimum number of nodes** to fit the problem, and reducing the **CPU clock frequency to medium**.

In contrast, the runtime, and thus energy consumption of tensor networks is very *dependent on the entanglement*. If it is low, **iTensor can drastically outperform QuEST** despite featuring *no parallelism*. However, for highly entangled states, the **full simulation is very slow** – but can be approximated by *trimming the bond dimensions*.

In the future, tensor network simulations should be *parallelised with OpenMP and MPI*. Then they could be directly compared against state vector methods to figure out whether and when they are advantageous, especially in case of an approximated state.

References and acknowledgements

[1] Frank Arute et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, Oct 2019.

[2] Matthew Fishman, Steven R. White, and E. Miles Stoudenmire. The ITensor Software Library for Tensor Network Calculations. *SciPost Phys. Codebases*, page 4, 2022.

[3] Tyson Jones et al. Quest and high performance simulation of quantum computers. *Scientific Reports*, 9(1):10736, Jul 2019.

[4] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*, page 219. Cambridge University Press, USA, 10th edition, 2011.

[5] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96–192, 2011. January 2011 Special Issue.

This work used the ARCHER2 UK National Supercomputing Service (`https://www.archer2.ac.uk`).

This work was supported by the Engineering and Physical Sciences Research Council.

Applications of quantum computing for quantum chemistry

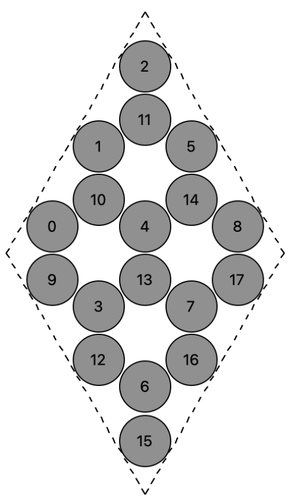
Bruno Camino
University College London

Quantum chemistry has been predicted to be one of the first fields to benefit from the development of quantum computing. In this work we explore applications of quantum annealing for the study of solid solutions. These materials are of great interest for energy store applications and simulating their properties with classical computers is particularly challenging because of the large configuration space to explore. Using vacancies of graphene as a model system, we show how quantum annealers can be used to tackle these type of problems.

Quantum Annealing Applications for Quantum Chemistry

Goal	Method	Test model
Predict lowest energy configuration of solid solutions	Quantum annealing, D-Wave machines	Simulate vacancies in graphene

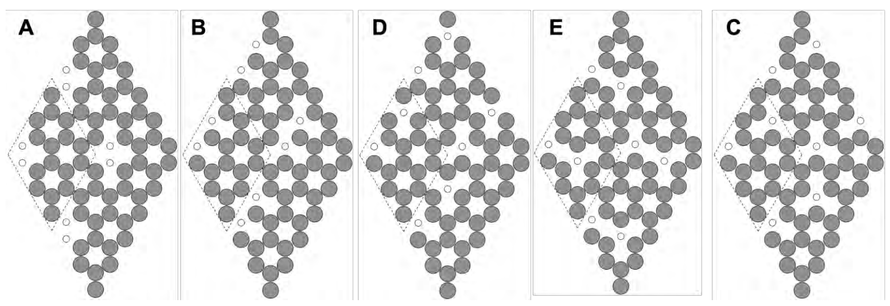
Binary Quadratic Model (QUBO)



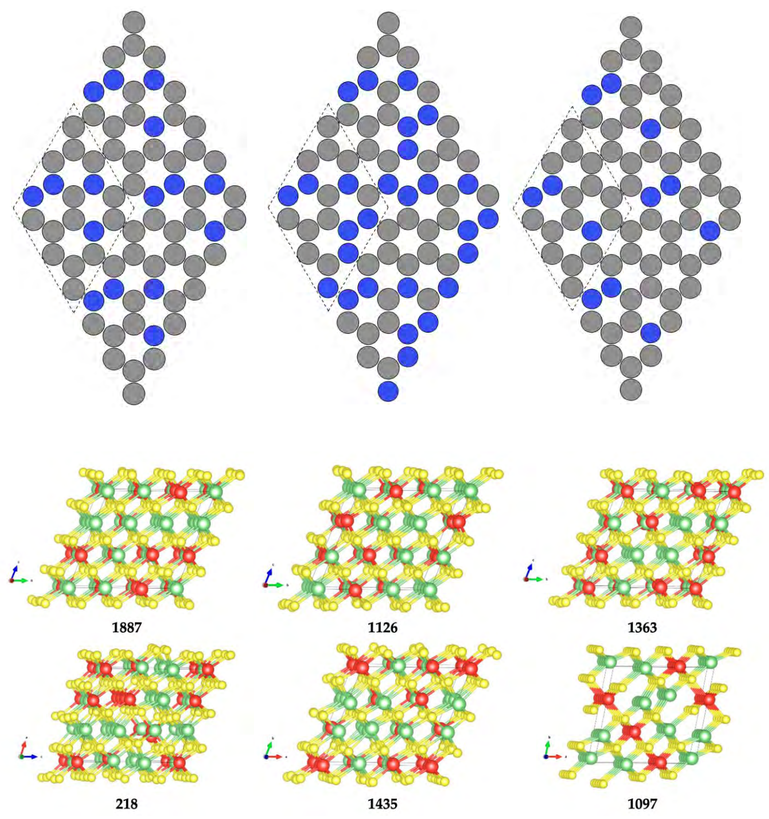
$$\alpha \underbrace{\sum_{i}^{N_{sites}} \sum_{j>i}^{N_{sites}} -A_{i,j}x_i x_j}_{\text{broken bonds objective}} + \beta \underbrace{\sum_{i}^{N_{atoms}} \sum_{j>i}^{N_{atoms}} -B_{i,j}x_i x_j}_{\text{coord objective}} +$$

$$+ \underbrace{\lambda \left(\sum_{i}^{N_{sites}} \left(1 - 2N_{atoms}\right) x_i + \sum_{i}^{N_{sites}} \sum_{j>i}^{N_{sites}} 2x_i x_j \right)}_{\text{constraint}}$$

$$\binom{N_{sites}}{N_{vac}} = \frac{N_{sites}!}{N_{vac}!(N_{sites} - N_{vac})!}$$



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	% broken chains	Energy	% occurence
A	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0.284900	-20.0	75.0
B	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0.793651	-19.0	3.6
C	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0.000000	-19.0	10.7
D	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0.000000	-19.0	3.9
E	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1.010101	-19.0	2.0



DFT calculated interaction energies

D-Wave Quantum annealing

Configurational analysis + Thermodynamic properties

SCAN ME!

QEVEC

D-wave demo

simulating
atomic vacancies
In graphene

Quantum Enhanced Verified Exascale Computing

ExCALIBUR
CROSS-CUTTING PROJECT

UKRI

Quantum Computing & Simulation Hub

ExCALIBUR 10

Validation and Application of Lagrangian Stochastic Methods for Indoor Air Quality

Harriet Jones
STFC / University of Chester

This STFC Air Quality Network (SAQN) project uses EDF's computational fluid dynamics software Code_Saturne to model the dispersion of a key hazardous aerial pollutant, particulate matter (PM), during cooking experiments within a test house. This is done via the implementation of Code_Saturne's Lagrangian Particle Tracking module. The basis for the model is the EPSRC funded DOMestic Systems Technology InCubator (DOMESTIC) test house, a controlled environment designed to simulate a full-scale kitchen/diner and bathroom. Early results indicate that the model appears to effectively replicate the evolution of PM2.5 (particulate matter with a diameter of 2.5 microns or less) during cooking episodes, and further experimental validation results are pending.

Validation and Application of Lagrangian Stochastic Methods for Indoor Air Quality

Harriet Jones, Gregory Cartland-Glover and Stefano Rolfo
Computational Engineering Group, Scientific Computing Department, STFC Daresbury Laboratory, Daresbury, Warrington, WA4 4AD, UK

The project

- Computational Fluid Dynamics (CFD) simulations are used to replicate fine particulate matter ($PM_{2.5}$) dispersion from a cooking pot in the DOMESTIC air quality test house. Various ventilation scenarios are tested. The validated CFD model will ultimately feed data back into the experimental work at DOMESTIC.

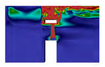
Tracking $PM_{2.5}$ dispersion

2.5 µm or less in diameter

- It may be invisible to the naked eye, but $PM_{2.5}$ is a serious health hazard. Regular exposure to concentrations of over **15 µg m^{-3}** is correlated with an increased risk of chronic obstructive pulmonary disorder (COPD), coronary heart disease, stroke, and lung cancer [1]. But cooking activities can cause local $PM_{2.5}$ concentrations of more than **350 µg m^{-3}** [2].

- Lagrangian particle tracking is highly computationally intensive, but enables individual particle trajectories to be calculated. Hence it is very suitable for air quality studies.

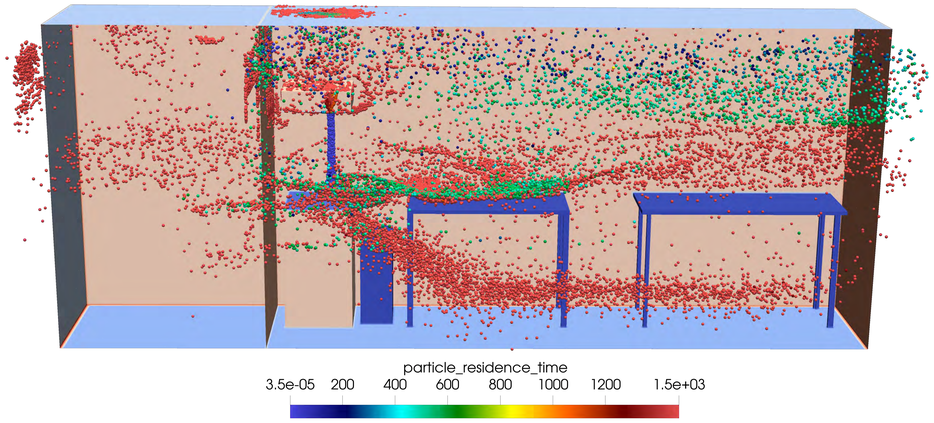
Code_Saturne



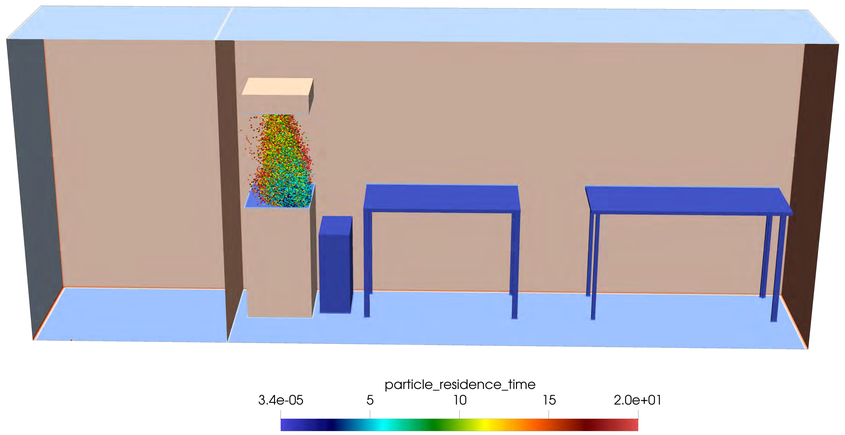
- Calculations have been performed using the open source CFD solver *Code_Saturne,* which is based on a finite volume discretisation. Second order central differencing is used for the convective terms, whereas an implicit Euler scheme is used for the time advance. The solver also includes a Lagrangian Particle Tracking (LPT) module [3].

Initial results – impact of ventilation on kitchen $PM_{2.5}$ concentrations during cooking

No ventilation…

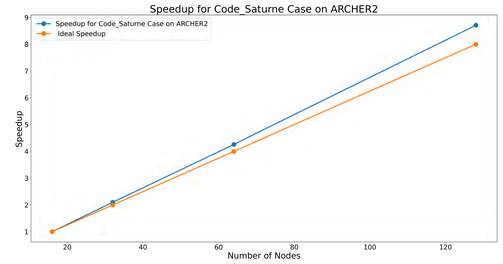


Versus hood extractor fan on (375 m^{3}hr^{-1})

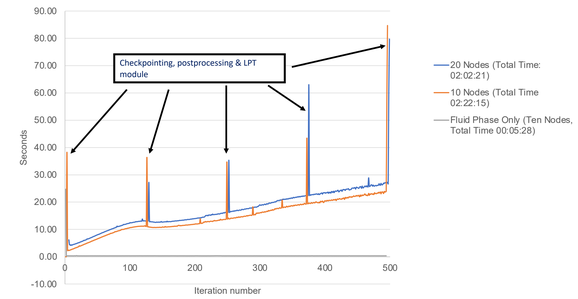


Cases use a ten million cell mesh, the EBRSM turbulence model and one-way fluid-particle coupling. 10,000 particles were injected per second, with a statistical weighting of 100. Simulations were run for 25 in-simulation minutes on twenty 32-core AMD nodes on STFC's Scientific Computing Application Resource for Facilities (SCARF) cluster.

Efficiency



- *Code_Saturne* scales excellently, with project cases showing a super-ideal speedup over a 16 -128 node range on ARCHER2. (Time step: 0.005 s, mesh size: 130 million cells)



- Efficient parallelization relies on the number of particles in the domain. Over the first 500 iterations of the case above, 20 nodes gave a 20-minute reduction in compute time, but the majority of iterations were slightly slower than for the 10-node case.

- It may be more energy efficient to initially run on fewer nodes, restarting on a higher number once the particle load is sufficiently high.

Acknowledgements

This project is funded by the STFC Air Quality Network and is a collaboration between the Science and Technology Facilities Council and the University of York. DOMESTIC is funded by the EPSRC

References

[1] World Health Organisation, **2021**, *WHO Air Quality Guidelines.* [2] Y. Omelekhina et.al., *Environmental Science: Processes & Impacts* **2020**, *22*, 1382-1396. [3] *Code_Saturne,* **2022, https://www.code-saturne.org/cms/web/**

Introducing Incoherence to Artificial Neural Networks

Asa Hopkins
University of Strathclyde

Artificial neural networks (NNs) at their core are an attempt to emulate the biological NNs found in the brains of animals, and can accomplish tasks with lower energy consumption than more traditional computing methods.However, there are still ways that artifical NNs fall short of their biological counterparts. Most artificial NNs are made up of layers of nodes, with edges only being formed between adjacent layers. This kind of strict ordering is not seen in nature and the existence of edges connecting non-adjacent layers is important to the stability of larger natural systems, such as food chains and metabolic pathways. The extent to which this strict layering is broken is known as the trophic incoherence.This work investigates methods of adding trophic incoherence to artificial NNs, and the effectsdoing so has on the convergence speed during training (fast convergenceis more energy efficient) and the accuracy after training is completed.

Introducing Trophic Incoherence to Traditional Neural Network Architectures

Asa Hopkins, Supervised by Dr. Samuel Johnson

UNIVERSITY OF BIRMINGHAM

Contact: asa.hopkins@strath.ac.uk
Github: https://github.com/Asa-Hopkins/Trophic-Networks

Artificial neural networks (NNs) at their core are an attempt to emulate the biological NNs found in the brains of animals, and can accomplish certain tasks more efficiently than more traditional computing methods. However, there are still ways that artifical NNs fall short of their biological counterparts. Most artificial NNs are made up of layers of nodes, with edges only being formed between adjacent layers. This kind of strict ordering is not seen in nature and the existence of edges connecting non-adjacent layers is important to the stability of larger natural systems, such as food chains and metabolic pathways. The extent to which this strict layering is broken is known as the trophic incoherence. This work investigates methods of adding trophic incoherence to artificial NNs, and the effects doing so has on the convergence speed during training (fast convergence is more energy efficient) and the accuracy after training is completed.

What is Trophic Incoherence?

Trophic incoherence is a way of measuring the amount of disorder in a directed graph, and it stems from the idea of a trophic level. We first define the trophic level of a node as the average of the trophic levels of the nodes that it takes inputs from, plus one. To completely define the trophic levels of the systems, nodes with no inputs are defined as having a trophic level of one.

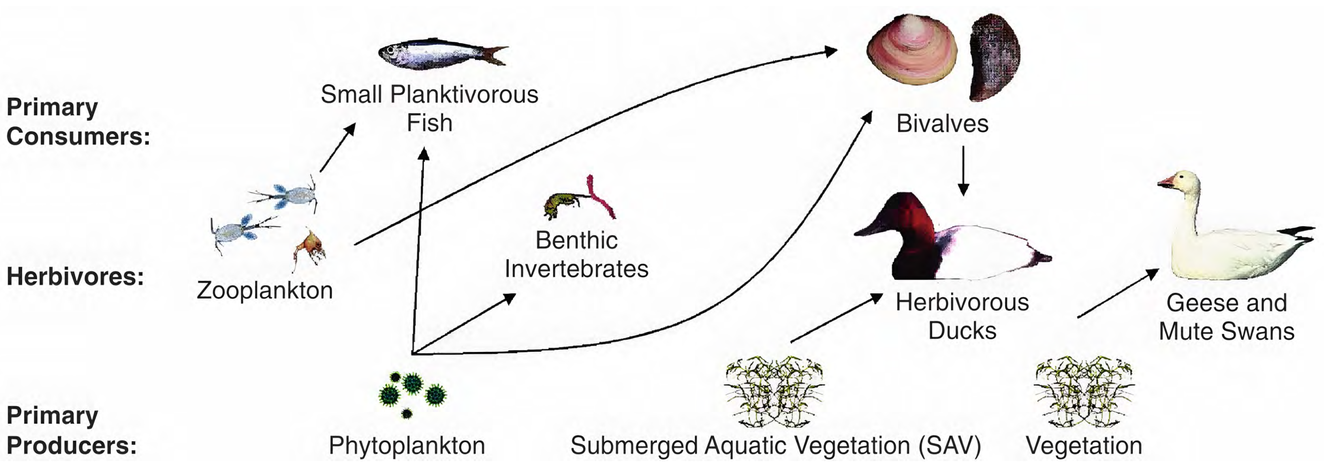
As an example, let's calculate the trophic level of the Herbivorous Ducks.

$$S_{Phyto}, S_{SAV} := 1, S_{Zoo} := 2$$

$$S_{Bivalve} = \frac{1}{2}\left(S_{phyto} + S_{Zoo}\right) + 1 = 2.5$$

$$S_{Ducks} = \frac{1}{2}\left(S_{SAV} + S_{Bivalve}\right) + 1 = 2.75$$

This places them somewhere between Herbivores and Primary Consumers

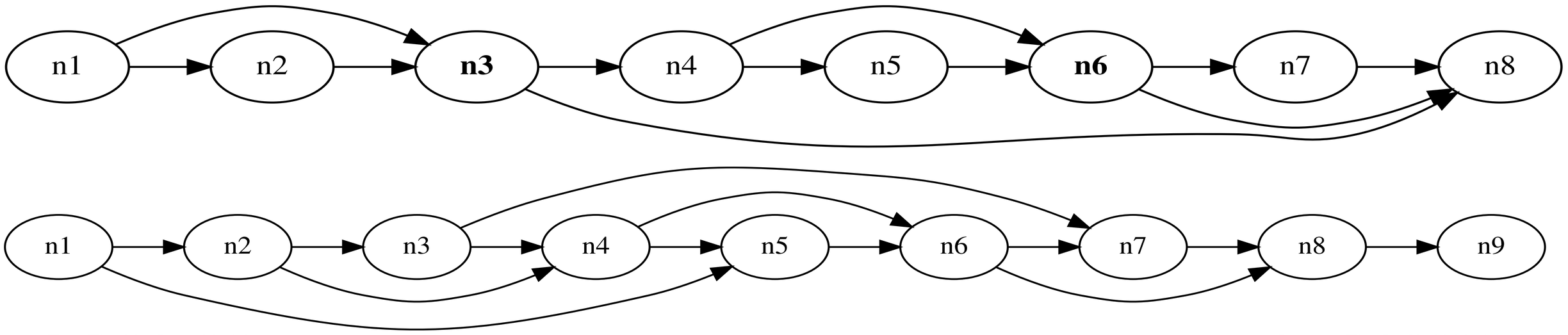


Consider the the "Herbivorous Ducks" in this image. They eat both primary producers and primary consumers, which makes it unclear whether they should be placed in herbivores or in secondary consumers. In this strictly layered diagram, there is no good option.

Connectivity

There are many ways in which incoherence can be introduced, and an early idea was to use randomly generated directed acyclic graphs, but this loses the advantage of efficient matrix multiplication for calculating node values. The approach that was eventually taken was to still have strict layers of nodes, but to allow connections to appear between non-adjacent layers. In the diagrams below, each circle represents an entire layer of nodes, and each connection represents a matrix of weights.

Even with this additional restriction, there are many possible choices of ways to add connections between layers. This is called the connectivity, and multiple connectivities have been tested.



Top: The MaxDist2 Model

Bottom: The ResNetX Model

Why Might Incoherence Help?

When calculating the derivative for gradient descent, each step in the connectivity graph represents one application of the chain rule in calculating the derivative with respect to that layer, and there is a tendency for this value to decrease as more applications of the chain rule are needed. The two connectivity models introduced, MaxDist1 and MaxDist2, ensure that a route to the output layer exists with only one or two steps respectively, meaning that this issue of vanishing gradients is mitigated.

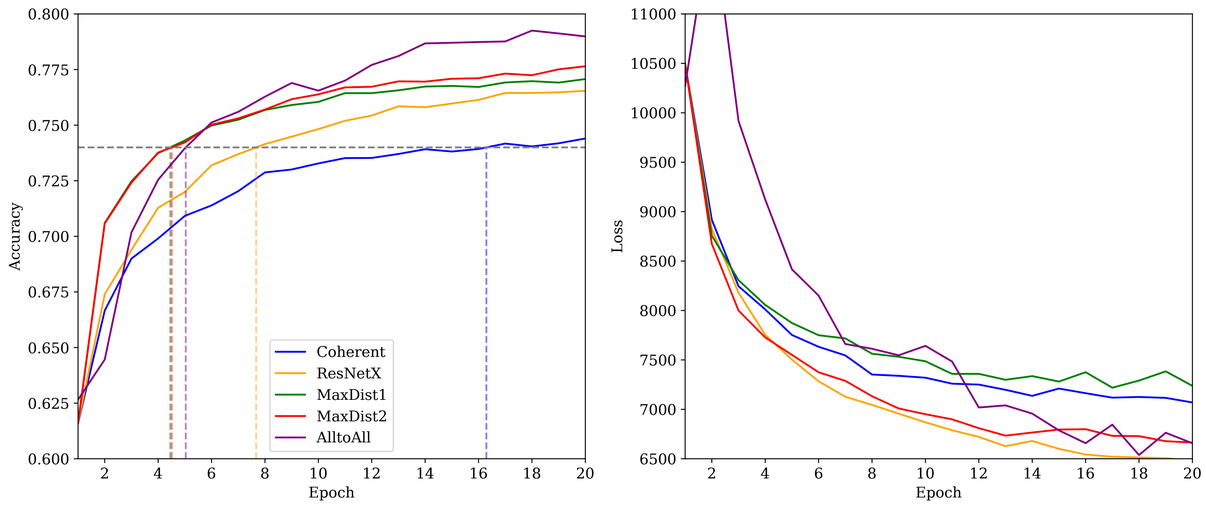
Neural Network Basics

Feed forward neural networks are a class of vector functions defined by the iteration

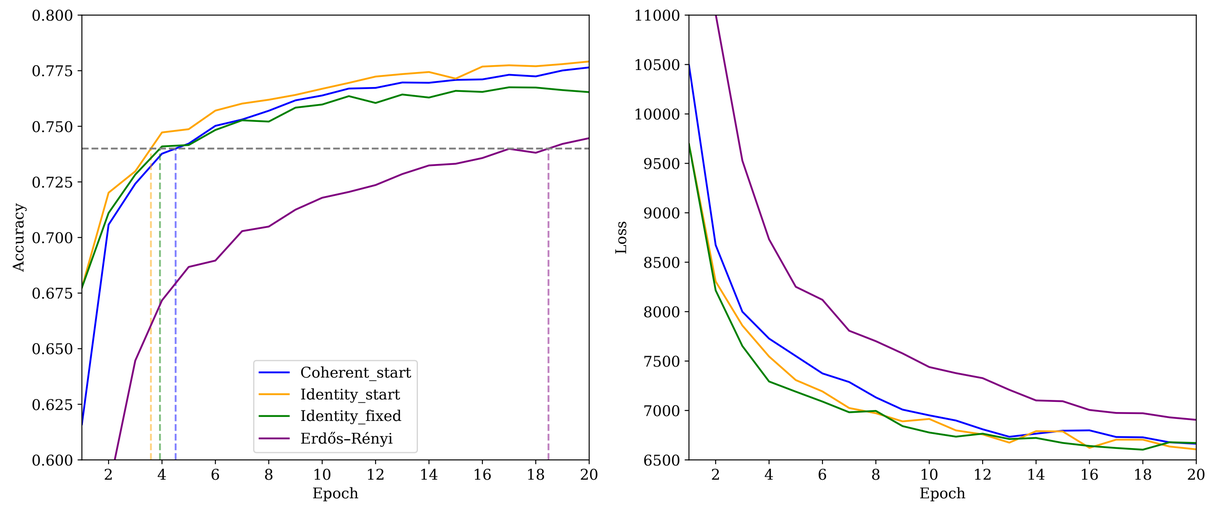
$$\mathbf{p}_{i+1} = \mathbf{W}_i \mathbf{n}_i, \qquad \mathbf{n}_{i+1} = f_i(\mathbf{p}_{i+1}).$$

Here $\mathbf{n}_0$ represents the input, $\mathbf{n}_l$ represents the output, each $\mathbf{W}$ is a matrix and each f is a (typically elementwise) nonlinear function. Initially, the elements of the matrices are random, but by inputting values which have a known output, an error can be calculated. The derivative of this error is then calculated, and a small step is taken in the direction of steepest descent to try and reduce it

Results for a 20 Layer Network on EMNIST

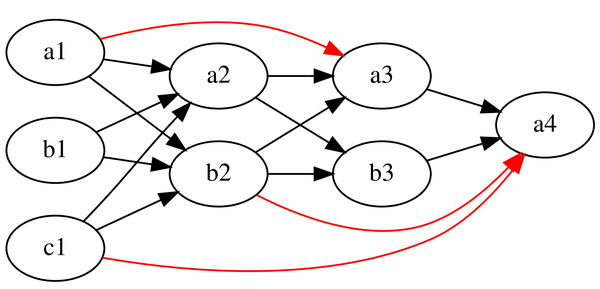


For this set of results, different connectivity models were tested. MaxDist1 and MaxDist2, show accelerated convergence and therefore require less training. In theory, an all-to-all connectivity should be ideal, as allowing more options for placing connections will never increase the global minimum, but in practise this makes the solution space harder to search.



For this set of results, MaxDist2 connectivity was used and different choices of inital connections were tried. It can be seen here that a good choice of initialisation can improve convergence by an epoch, although it is not obvious beforehand what choice is optimal.

University of Strathclyde Glasgow



UK Research and Innovation

Using Machine Learning Techniques to Determine Photometric Redshifts for Gravitational wave Cosmology

Lara Janiurek
University of Strathclyde

The inference of the Hubble constant using gravitational waves has allowed for a new way for the expansion of the universe to be probed, which may shed light on the current Hubble tension. Galaxy redshift surveys are a required for the application of these dark sirens. Photometric redshift surveys contain significant errors and spectroscopic redshifts are much more energy intensive than simply using an algorithm to estimate these values. Here, the random forest (RF) algorithm GALPRO is implemented to generate photometric redshift posteriors. GALPRO is calibrated using a truth dataset, which is successful, meaning it is useful when presented with an incomplete survey with missing redshift values. Analysis suggests that the redshift posterior distributions are non-Gaussian. Tests were run which determined that training and testing datasets must overlap by least 90% in range to give accurate results. However, the algorithm failed when the training and testing datasets came from different surveys meaning there is some underlying fundamental difference in galaxy surveys that must be recognised when using RFs.

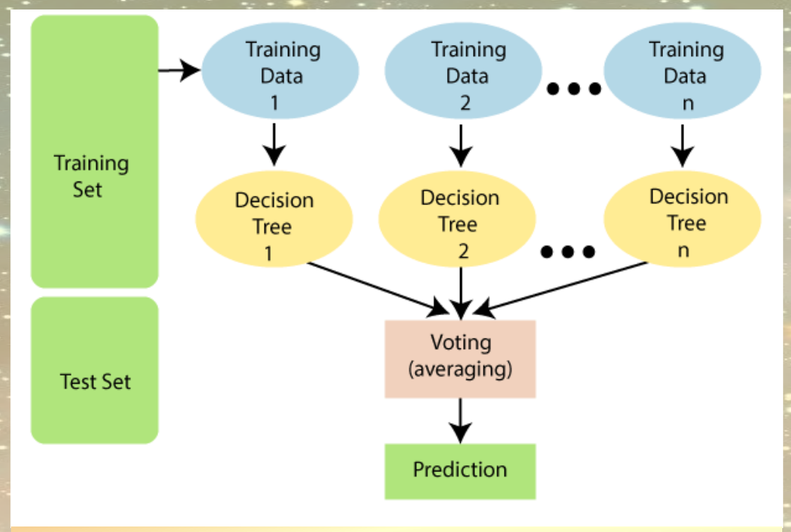
Using Machine Learning Techniques to Determine Photometric Redshifts for Gravitational wave Cosmology

Lara Janiurek , Supervised by Prof Martin Hendry

The inference of the Hubble constant using gravitational wave data has allowed for a new way for the expansion of the universe to be probed. The use of binary black hole merger events to measure the Hubble constant may shed considerable light on the current Hubble tension. Galaxy redshift surveys are a key ingredient for the application of these dark sirens in the measurement of H_0.

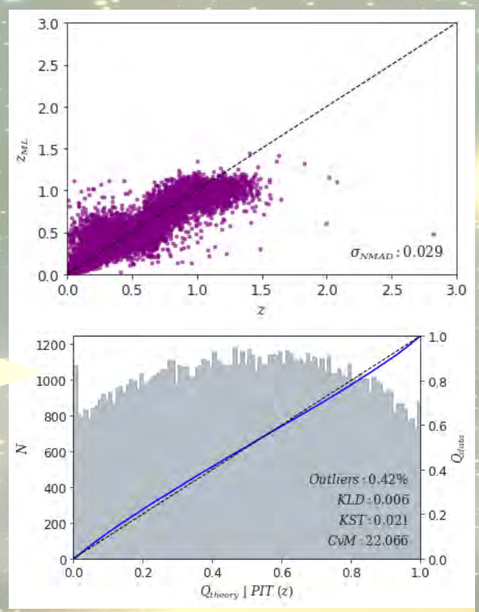
- Improving the performance of dark sirens requires a better understanding of the photometric redshift errors.
- Current redshift values used by LIGO for cosmological inference are assumed to have an associated Gaussian error.
- Quantification of the redshift posteriors would give a more accurate result in the overall.
- Photometric redshift surveys often contain significant errors.
- Spectroscopic surveys are expensive and rely on cosmological models
- Machine learning techniques are advantageous in that they don't rely on these models.

Contact
lara.Janiurek@strath.ac.uk
Supervisor: Martin Hendry
Martin.Hendry@glasgow.ac.uk



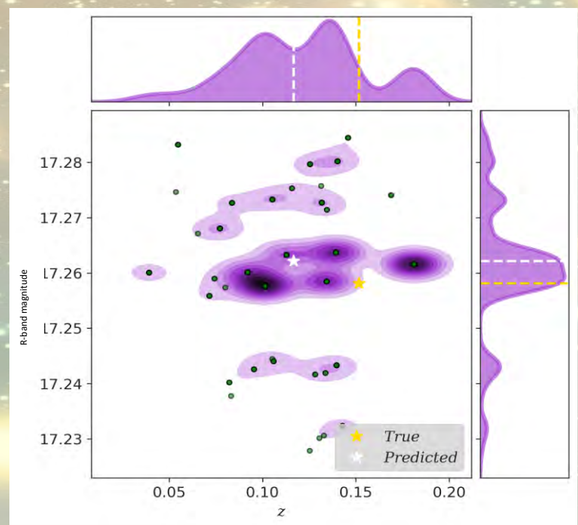
A simplified diagram of how a random forest algorithm learns mapping and predicts outcomes [1].

The random forest algorithm learns the mapping between the photometry data and spectroscopic redshift of each galaxy. This learnt mapping can then be applied to a new dataset which only has photometry data.



The photometric versus spectroscopic redshift and PIT plot of GALPRO when trained and tested using subsets of the Zhou et al dataset.

In this work, the random forest algorithm GALPRO is implemented to generate photometric redshift posteriors. It is initially calibrated and optimized using a truth dataset compiled by Zhou et al.

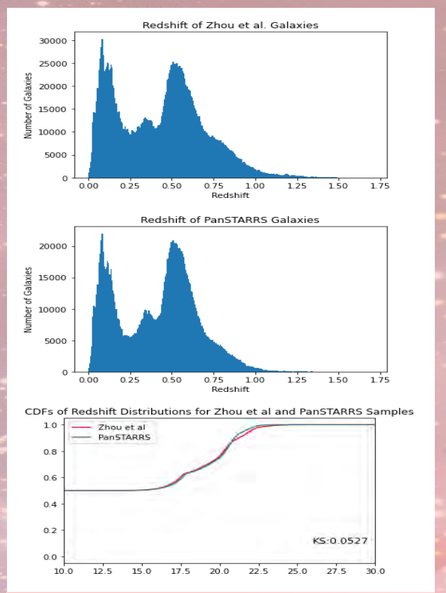


A joint posterior distribution of a randomly selected galaxy from the Zhou et al dataset. The redshift posterior is shown on the top and the r-band magnitude posterior is shown on the right.

The analysis suggests that the redshift posterior distributions are largely non-Gaussian, reinforcing the need for a reliable method to generate redshift posteriors to better represent these photometric errors.

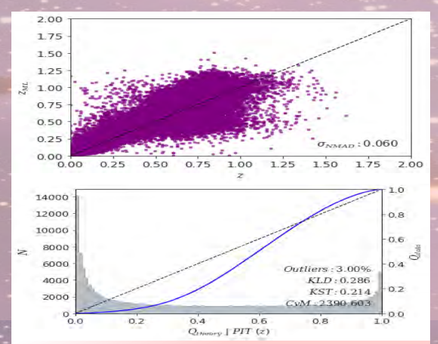
The initial calibration showed that GALPRO is very useful when provided with an incomplete galaxy survey with missing spectroscopic values, as missing redshift can be accurately predicted.

- Tests were run using the Zhou et al. dataset to determine how statistically similar the training and testing datasets must be for GALPRO to be applicable.
- The training and testing datasets were found to require similar redshift distributions and overlap by at least 90% in the band ranges to give accurate results.
- GALPRO was then trained using the Zhou et al. dataset and applied to a sample from the PanSTARRS survey to explore if GALPRO could be trained using a trusted dataset and applied to a general, new survey.



The redshift distributions and their compared CDFs for the PanSTARRS and Zhou datasets.

Despite this, application of the algorithm still resulted in a catastrophic failure, indicating that there must be some underlying fundamental difference between the two surveys that cause the program to not learn the correct mapping. This serves as a cautionary tale in the application of random forests to new surveys when generating photometric redshift posteriors.



The photometric versus spectroscopic redshift plot and PIT of the PanSTARRS sample when trained using the Zhou et al dataset. The scatter and PIT demonstrate a failure in the learnt mapping of the algorithm.

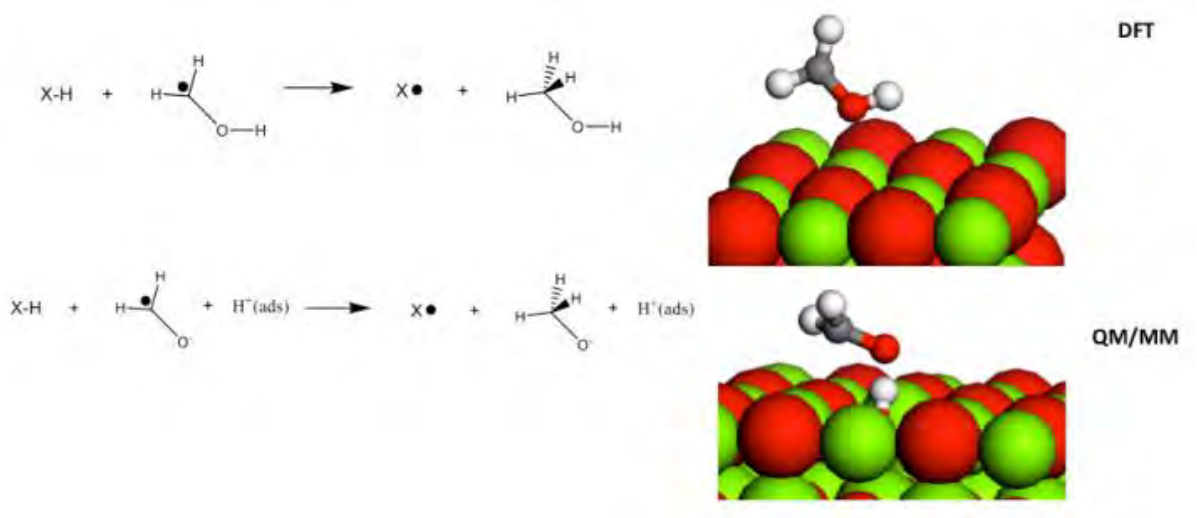
References
[1]https://www.javatpoint.com/machine-learning-random-forest-algorithm
[2] R Gray. Gravitational wave cosmology: measuring the Hubble constant with dark standard sirens. PhD thesis, University of Glasgow, 2021.

- Producing galactic spectrums is expensive and uses a lot of energy.
- These posteriors are used in many areas of astronomy.
- Being able to produce them using machine learning algorithms would decrease the energy cost of production and therefore help astronomy and the planet!

A combined periodic DFT and QM/MM approach to understand the radical mechanism of the catalytic production of methanol from glycerol

Mala Alhaji Sainna

The production of methanol from glycerol over a basic oxide, such as MgO, using high reaction temperatures (320 °C) is a promising new approach to improving atom efficiency in the production of biofuels. The mechanism of this reaction involves the homolytic cleavage of the C3 feedstock, or its dehydration product hydroxyacetone, to produce a hydroxymethyl radical species which can then abstract an H atom from other species. Obtaining a detailed reaction mechanism for this type of chemistry is difficult due to the large number of products present when the system is operated at high conversions. In this contribution we show how DFT based modelling studies can provide new insights into likely reaction pathways, in particular the source of H atoms for the final step of converting hydroxymethyl radicals to methanol. We show that water is unlikely to be important in this stage of the process, C-H bonds of C2 and C3 species can give an energetically favourable pathway and that the disproportionation of hydroxymethyl radicals to methanol and formaldehyde produces a very favourable route. Experimental analysis of reaction products confirms the presence of formaldehyde. The calculations presented in this work also provides new insight into the role of the catalyst surface in the reaction showing that the base sites of the MgO(100) are able to deprotonate hydroxymethyl radicals but not methanol itself. In carrying out the calculations we also show how periodic DFT and QM/MM approaches can be used together to obtain a rounded picture of molecular adsorption to surfaces and homolytic bond cleavage which are both central to the reactions studied.



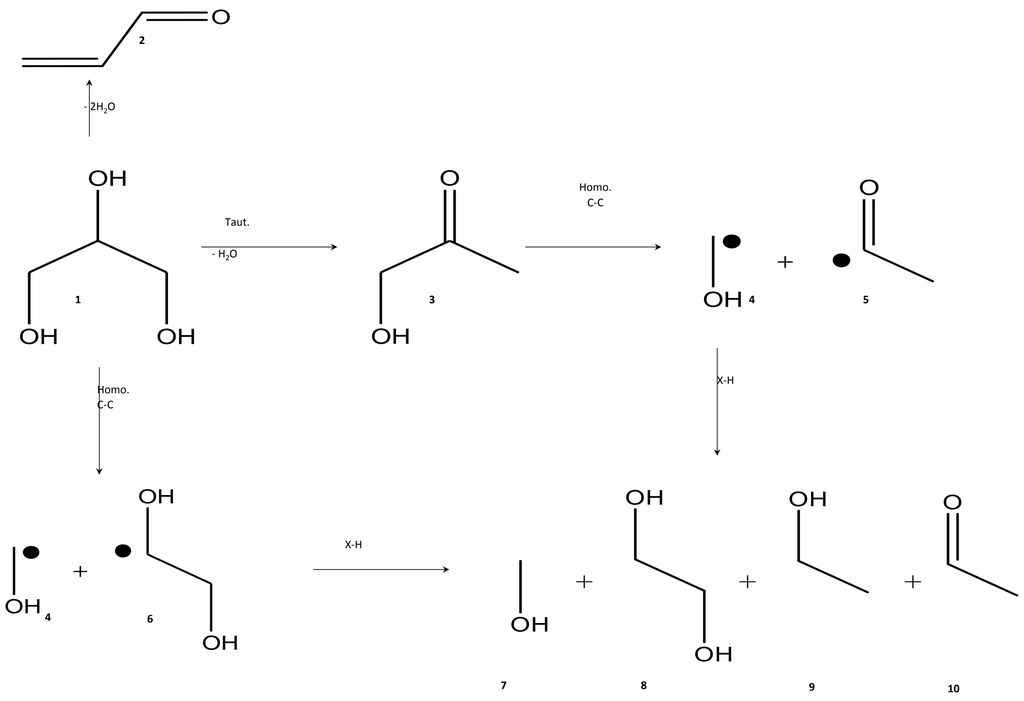
Reference:
• Sainna, MA; et al. Faraday Discuss. 229, 108 (2020)
• E. Lotero, Y. et al. Ind. Eng. Chem. Res., 44, (2005)

A combined periodic DFT and QM/MM approach to understand the radical mechanism of the catalytic production of methanol from glycerol.

Mala A. Sainna, Sachin Nanavati, Constance Black, Louise Smith, Karl Mugford, Harry Jenkins, Mark Douthwaite, Nicholas F. Dummer, C. Richard A. Catlow, Graham J. Hutchings, Stuart H. Taylor, Andrew J. Logsdail, and David J. Willock.

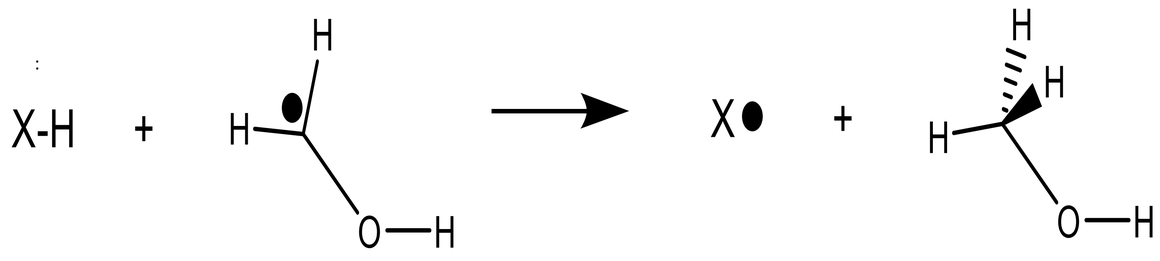
Introduction:

The production of methanol from glycerol over a basic oxide, such as MgO, using high reaction temperatures (320 °C) is a promising new approach to improving atom efficiency in the production of biofuels. The mechanism of this reaction involves the homolytic cleavage of the C_3 feedstock, or its dehydration product hydroxyacetone, to produce a hydroxymethyl radical species which can then abstract an H atom from other species. Obtaining a detailed reaction mechanism for this type of chemistry is difficult due to the large number of products present when the system is operated at high conversions. In this contribution we show how DFT based modelling studies can provide new insights into likely reaction pathways, in particular the source of H atoms for the final step of converting hydroxymethyl radicals to methanol. We show that water is unlikely to be important in this stage of the process, C-H bonds of C_2 and C_3 species can give an energetically favourable pathway and that the disproportionation of hydroxymethyl radicals to methanol and formaldehyde produces a very favourable route. Experimental analysis of reaction products confirms the presence of formaldehyde. The calculations presented in this work also provides new insight into the role of the catalyst surface in the reaction showing that the base sites of the MgO(100) are able to deprotonate hydroxymethyl radicals but not methanol itself. In carrying out the calculations we also show how periodic DFT and QM/MM approaches can be used together to obtain a rounded picture of molecular adsorption to surfaces and homolytic bond cleavage which are both central to the reactions studied.

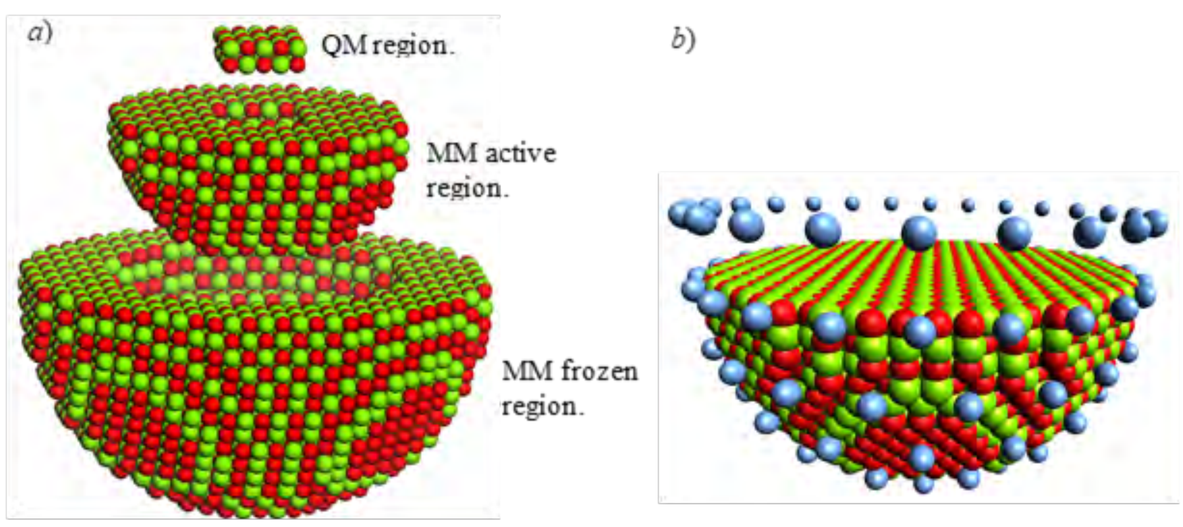


Scheme 1: Simplified reaction scheme for the production of methanol from glycerol.

Results and Discussion:



Scheme 2. Hydrogen transfer to produce methanol from hydroxymethyl radical



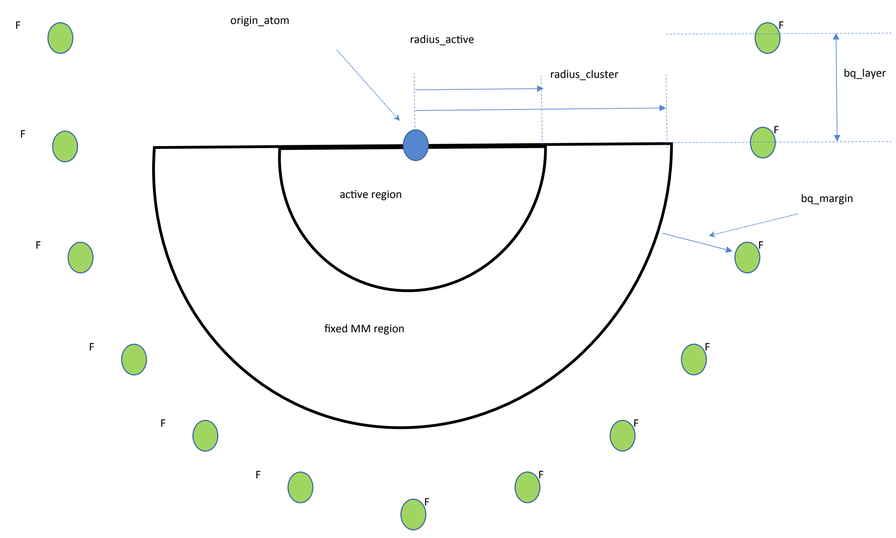
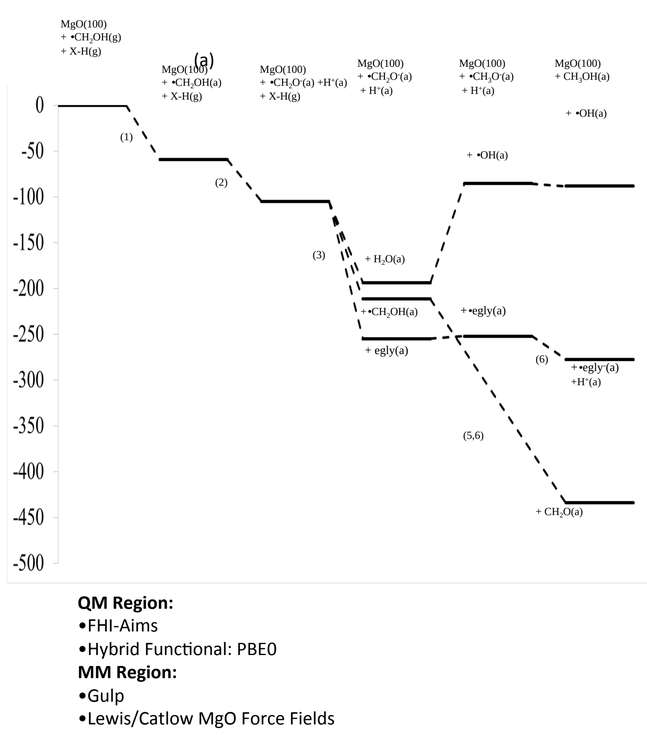
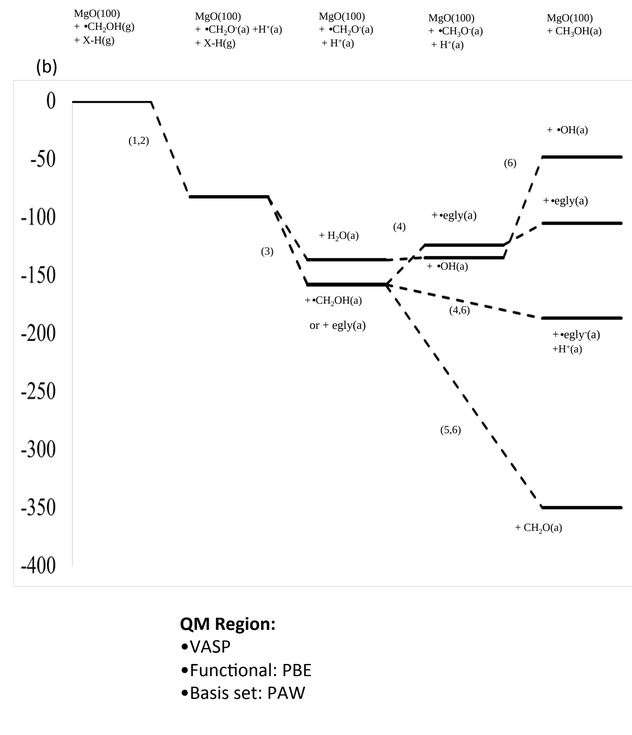


Figure 1: An expanded view of the QM/MM simulation regions used





Computational Methods:

VASP – Vienna Ab-Initio Simulation Package:

- projector augmented-wave (PAW) method reprented for core states,
- PBESol Functional, with Grimmes D3 Dispersion**,**
- Forces with a acceptance criterion of 0.01 eV Å-1
- planewave cut off of 800 eV .

QM/MM Technique – Chemshell Interface:

- QM region consisting of 2 Layered 50 ions, FHI-aims implemented,
- PBEsol Functional,
- TZVP equivalent Basis sets implemented,
- Non-coulombic repulsion/dispersion interactions are represented with Buckingham potentials,
- MM Region are treated using ***GULP*** Package software,
- Pseudopotential is positioned on each Mg^{2+} ion in the active MM region,
- ***CHEMSHELL*** interface is used to link QM and MM calculations.

QM Region:
- •FHI-Aims
- •Hybrid Functional: PBE0

MM Region:
- •Gulp
- •Lewis/Catlow MgO Force Fields

QM Region:
- •VASP
- •Functional: PBE
- •Basis set: PAW

Conclusion:

- we have used DFT calculations both with periodic boundary and with an embedded QM/MM approach ,
- We find that alcohol groups at C-centred radicals are deprotonated by the weak O_{5c} base sites on the surface, whereas alcohol groups of normal alcohols are not,
- we have shown how periodic and QM/MM approaches can be used in unison to arrive at a consistent model of surface processes,

Reference:

Sainna, MA; *et al*. *Faraday Discuss*. 229, 108 (2020),

Long-range dispersion-inclusive machine learning potentials for hybrid organic-inorganic interfaces

Shayantan Chaudhuri
University of Warwick

The computational prediction of the structure and stability of hybrid organic–inorganic interfaces provides important insights into the measurable properties of electronic thin film devices and catalyst surfaces, and plays an important role in their rational design. However, the rich diversity of molecular configurations and the important role of long-range interactions in such systems make it difficult to use machine learning potentials (MLPs) to facilitate structure exploration that would otherwise require computationally expensive electronic structure calculations. We present an ML approach that enables fast, yet accurate, structure optimisations by combining two different types of deep neural networks trained on high-level electronic structure data for gold nanoclusters on diamond (110) surfaces.

Long-Range Dispersion-Inclusive Machine Learning Potentials for Hybrid Organic-Inorganic Interfaces

Julia Westermayr[1], Shayantan Chaudhuri[1,2], Andreas Jeindl[3], Oliver T. Hofmann[3], Reinhard J. Maurer[1]

Department of Chemistry[1], Centre for Doctoral Training in Diamond Science and Technology[2], University of Warwick, Coventry, CV4 7AL, United Kingdom

Institute of Solid State Physics, Graz University of Technology, 8010 Graz, Austria[3]

1. Introduction

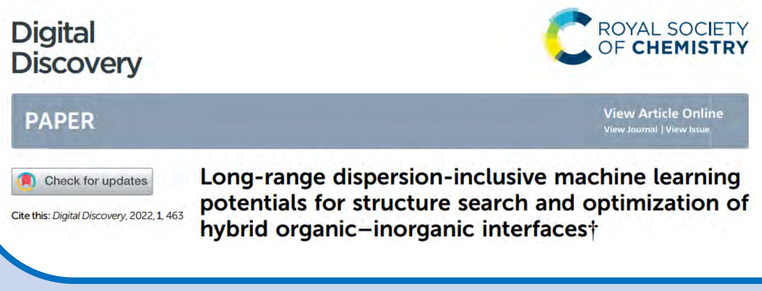
- Understanding how nanoclusters (NCs) form is crucial to controlling their final morphology and catalytic reactivity
- Machine learning potentials (MLPs) offer high computational efficiency and can retain the accuracy of electronic structure theory methods
- However, machine learning potentials are often based on local descriptors and therefore often incapable of efficiently learning long-range interactions e.g. dispersion (vdW) effects
- Is there a way to include long-range vdW effects with short-ranged machine learning potentials?

YES!

2. Solution

- Learn short-range effects from density functional theory (DFT)
- Add long-range vdW effects using the open-source Libmbd library
- Connect both via Hirshfeld atoms-in-molecules partitioning

If you're interested, you can read our paper in *Digital Discovery* now!


Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

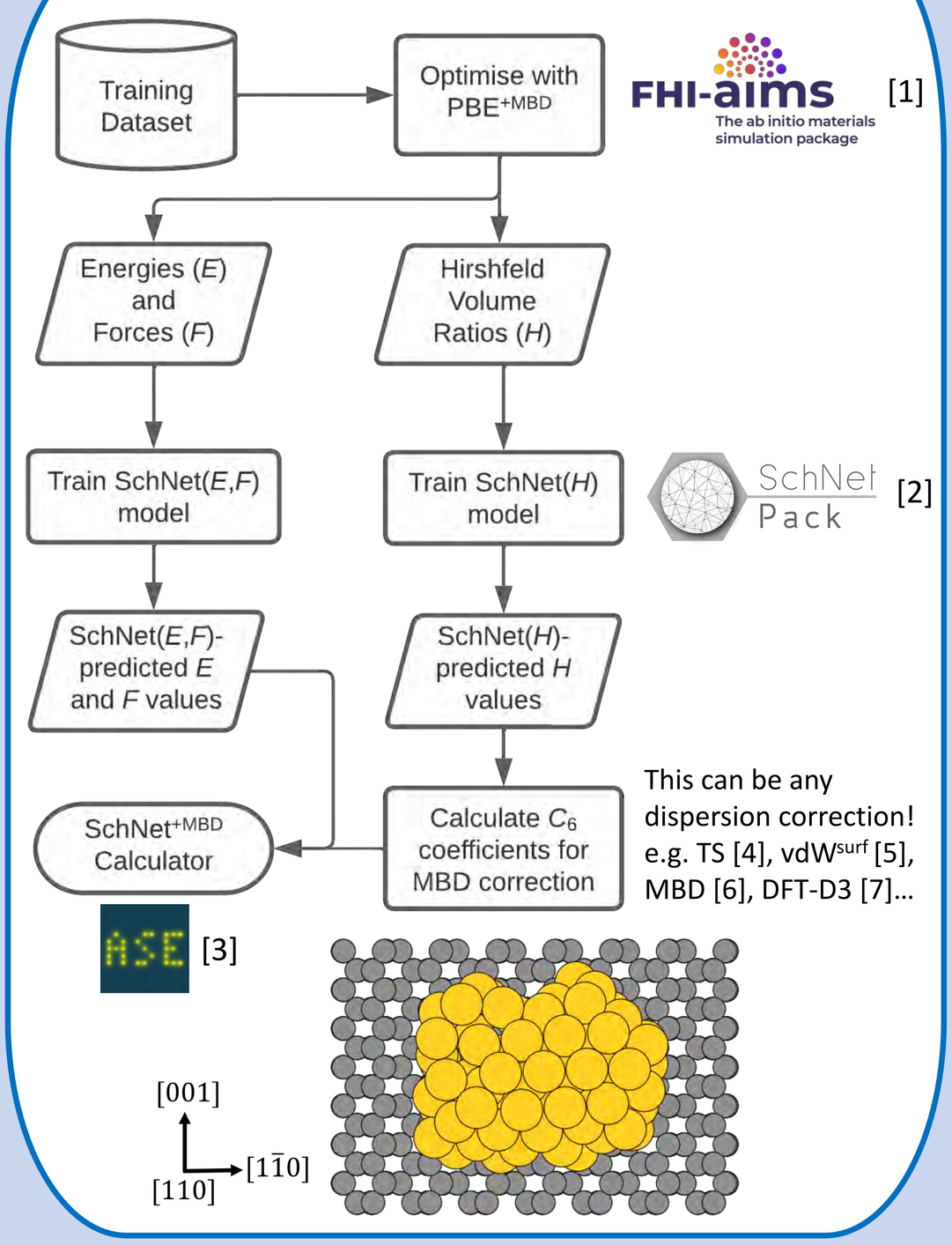
Check for updates

Cite this: *Digital Discovery*, 2022, 1, 463

Long-range dispersion-inclusive machine learning potentials for structure search and optimization of hybrid organic–inorganic interfaces†

Check out our open-access Python-based code on GitHub!

3. Training: Gold Nanoclusters on Diamond



FHI-aims The ab initio materials simulation package [1]

SchNet Pack [2]

This can be any dispersion correction! e.g. TS [4], vdW^{surf} [5], MBD [6], DFT-D3 [7]…

[3]

4. Results

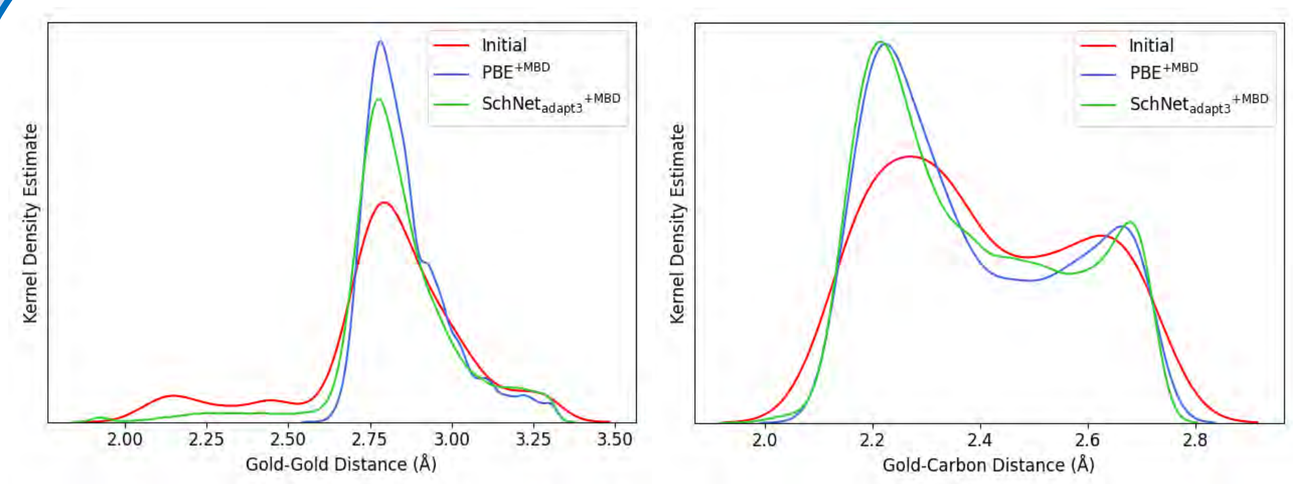


Fig. 1: Radial atom distributions for gold NCs after DFT- and ML-optimisation

$ML_{init.}$: Trained on initial training set

Adaptive sampling [8] done to generate data: $ML_{init.}$ → $ML_{adapt.1}$ → $ML_{adapt.2}$ → $ML_{adapt.3}$

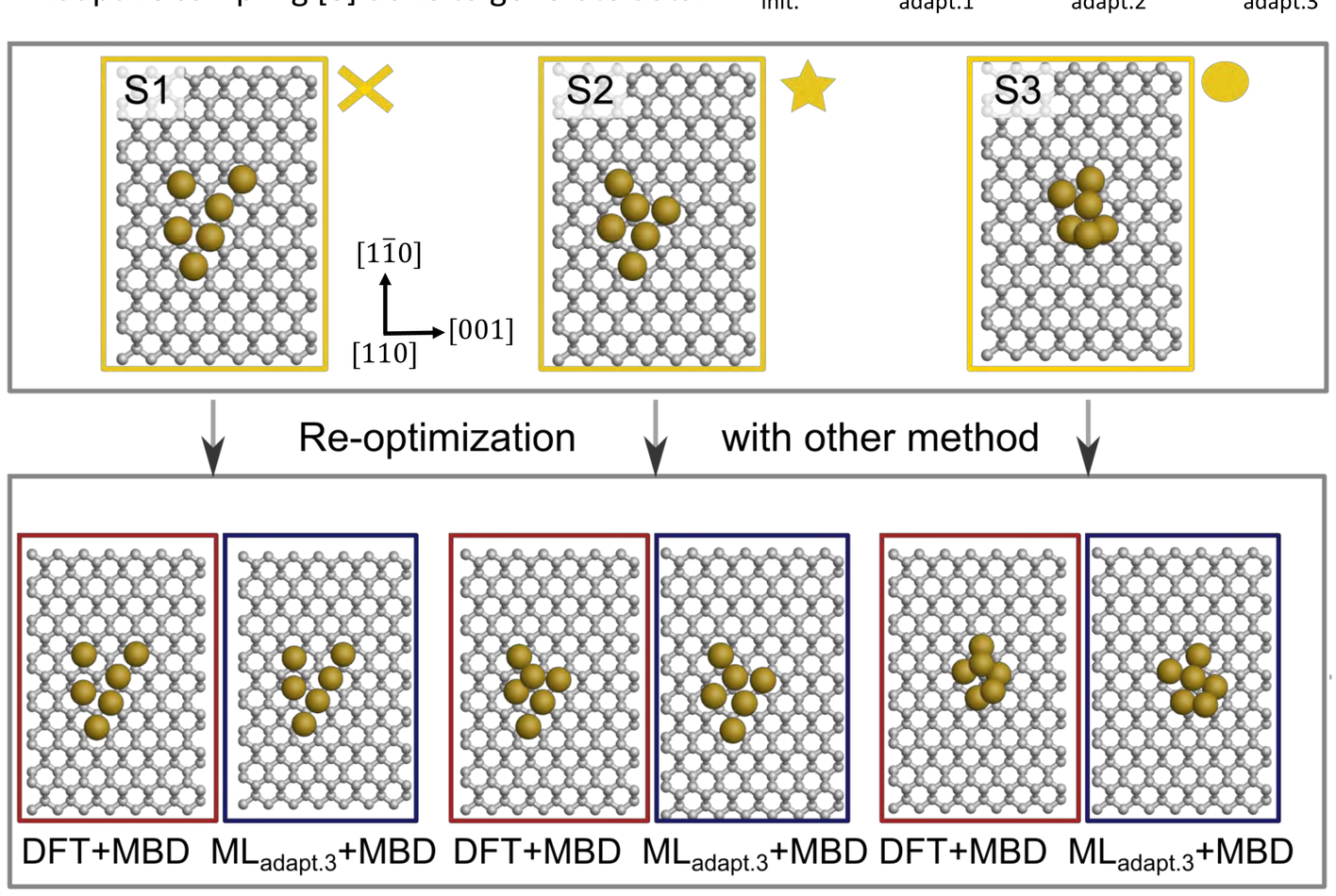


Fig. 2: Comparison of optimisations using DFT and our method. S1–S3 are minima obtained from a basin-hopping algorithm [9,10] with $ML_{init.}^{+MBD}$

Method	Computational Cost (kCPUh)
PBE^{+MBD}	72.71
$ML_{init}^{+MBD} + PBE^{+MBD}$	59.56
$ML_{adapt1}^{+MBD} + PBE^{+MBD}$	36.35
$ML_{adapt2}^{+MBD} + PBE^{+MBD}$	27.07
$ML_{adapt3}^{+MBD} + PBE^{+MBD}$	12.38
ML_{adapt3}^{+MBD}	7.83×10^{-4}

Table 1: Computational costs of a single geometry optimisation using various methods, as recorded with the ARCHER2 supercomputer

5. Conclusions

- We have developed a framework to combine short-range MLPs with long-range vdW effects
- Our method can be used for fast (pre-) relaxations of complex systems
- Our method links Libmbd to the Atomic Simulation Environment, as well as SchNetPack
- Our method has also been tested on diverse organic molecules adsorbed onto metal surfaces

References

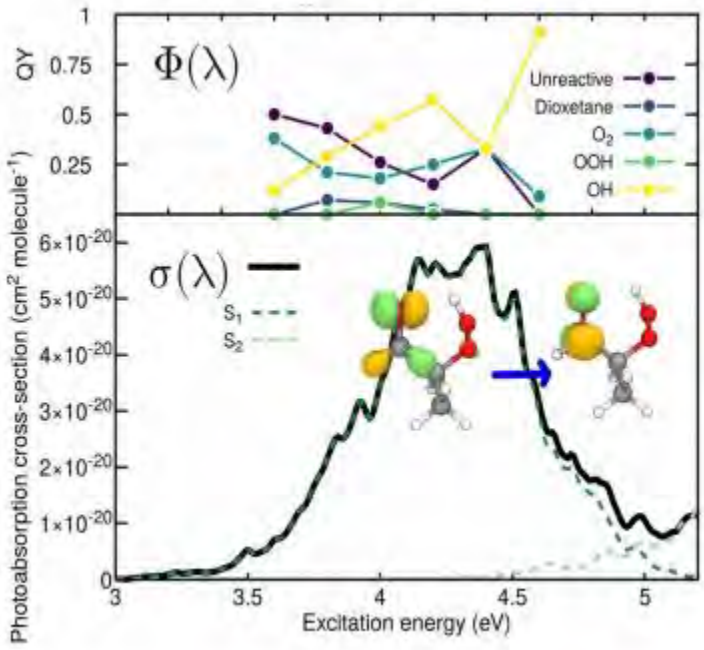
1. Schütt *et al.* (2019) *J. Phys. Theory Comput.* **15**: 448
2. Blum *et al.* (2009) *Comput. Phys. Commun.* **180**: 2175
3. Larsen *et al.* (2017) *J. Phys. Condens. Matter* **29**: 273002
4. Tkatchenko and Scheffler (2009) *Phys. Rev. Lett.* **102**: 073005
5. Ruiz *et al.* (2016) *Phys. Rev. B* **93**: 035118
6. Tkatchenko *et al.* (2012) *Phys. Rev. Lett.* **108**: 236402
7. Grimme *et al.* (2010) *J. Chem. Phys.* **132**: 154104
8. Behler *et al.* (2015) *Int. J. Quantum Chem.* **115**: 1032
9. Wales and Doye (1997) *J. Phys. Chem. A* **101**: 5111
10. Wales and Scheraga (1999) *Science* **285**: 1368

HEC MCC UK · archer2 · NI-HPC Northern Ireland High Performance Computing · UKRI Future Leaders Fellowships · FWF Der Wissenschaftsfonds. · UKRI Engineering and Physical Sciences Research Council · DST Diamond Science Technology Centre · maurer group · WARWICK The University of Warwick

A Theoretical Perspective on the Actinic Photochemistry of 2-hydroperoxypropanal

Emanuele Marsili
University of Bristol

Determining the chemical composition of the Earth's troposphere and its evolution over time is crucial for shaping the political and societal decisions regarding global warming. Presently used chemical mechanism models - encompassing experimental and theoretical data for many ground-state reactions of volatile organic compounds (VOCs) - allow estimating the outcomes of VOCs reactions. Interestingly though, the role of light-induced, excited-state processes is still largely unexplored and photochemical reactions of transient VOCs are mostly neglected in predictive atmospheric models.



One important family of VOCs is the α-hydroperoxycarbonyls. Since experimental studies on these transient molecules are hardly feasible, we have employed high-level quantum chemical methods to fully characterize the photochemistry of the 2-hydroperoxypropanal (2-HPP) [1]. Using the nuclear ensemble approach we calculated the photo-absorption cross-section (σ(λ)) [2] while we resorted to nonadiabatic molecular dynamics to determine the wavelength-dependent photolysis quantum yield (Φ(λ)). These two ingredients, together with the solar actinic flux (F(λ)), allow us to predict the photolysis rate constant J, a crucial piece of information required by predictive chemical mechanism models.

[1] Marsili E., et al. The Journal of Physical Chemistry A 2022, 126, 5420–5433
[2] Prlj A., Marsili E., et al., ACS Earth and Space Chemistry 2022, 6, 207-217

A Theoretical Perspective on the Actinic Photochemistry of 2-Hydroperoxypropanal

University of BRISTOL

IN SILICO PHOTOCHEMISTRY GROUP

Emanuele Marsili, Antonio Prlj and Basile F.E Curchod

School of Chemistry University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom. E-mail: or22568@bristol.ac.uk

- **Volatile Organic Compounds** (VOCs) are important molecules in the atmosphere. Upon their release, they trigger a complex network of oxidative reactions.
- Atmospheric chemists have determined the reactivities of the most relevant VOC reactions by incorporating the knowledge of products and kinetics in atmospheric models to predict the time evolution of chemicals in the atmosphere.
- Oxidative reactions often involve **transient** VOCs containing one or more **chromophoric groups**. Interestingly, the photochemistry of these transient species is still largely unexplored.
- The **alpha-hydroperoxycarbonyl** is a family of transient VOCs that can undergo photolysis under solar irradiation with important consequences on the oxidative balance of the atmosphere. Unfortunately, their photochemistry is out of the reach of experimental studies.

CAN WE USE COMPUTATIONAL AND THEORETICAL CHEMISTRY TO COMPUTE THE OBSERVABLES REQUIRED BY ATMOSPHERIC MODELS?

Compute in silico the photolysis rate constant

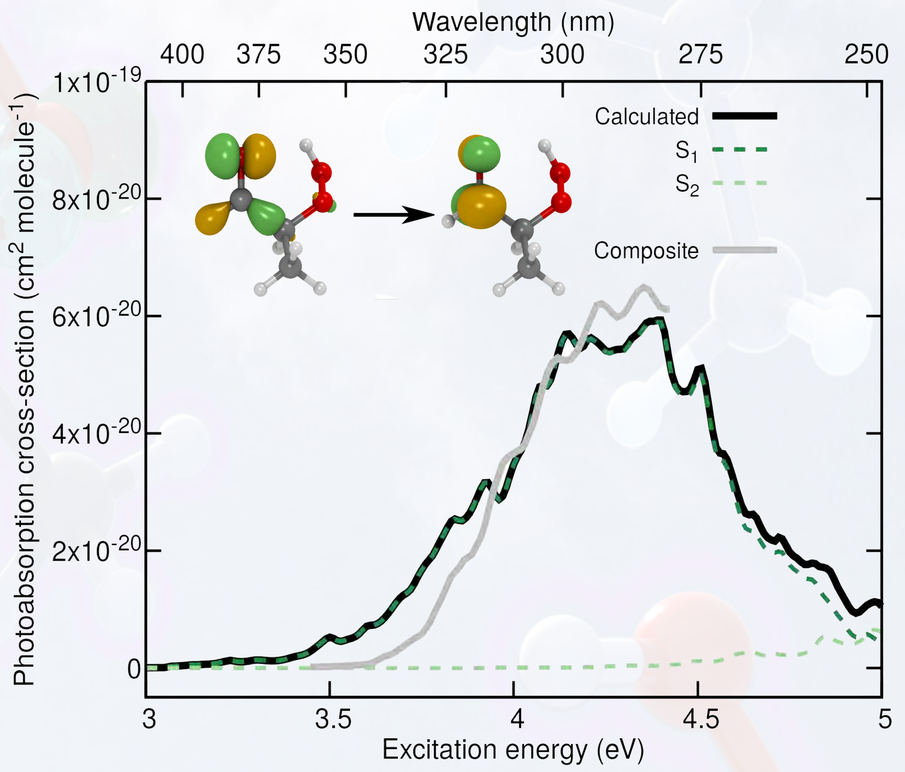
The photolysis process is characterized by its photolysis rate constant J - a first-order decay constant - defined as

$$J = \int_{\lambda_{min}}^{\lambda_{max}} F(\lambda)\ \sigma(\lambda)\ \phi(\lambda)\ d\lambda$$

F(λ) is the flux of the irradiation source
σ(λ) is the photoabsorption cross-section of the molecule
ϕ(λ) is the wavelength-dependent quantum yield

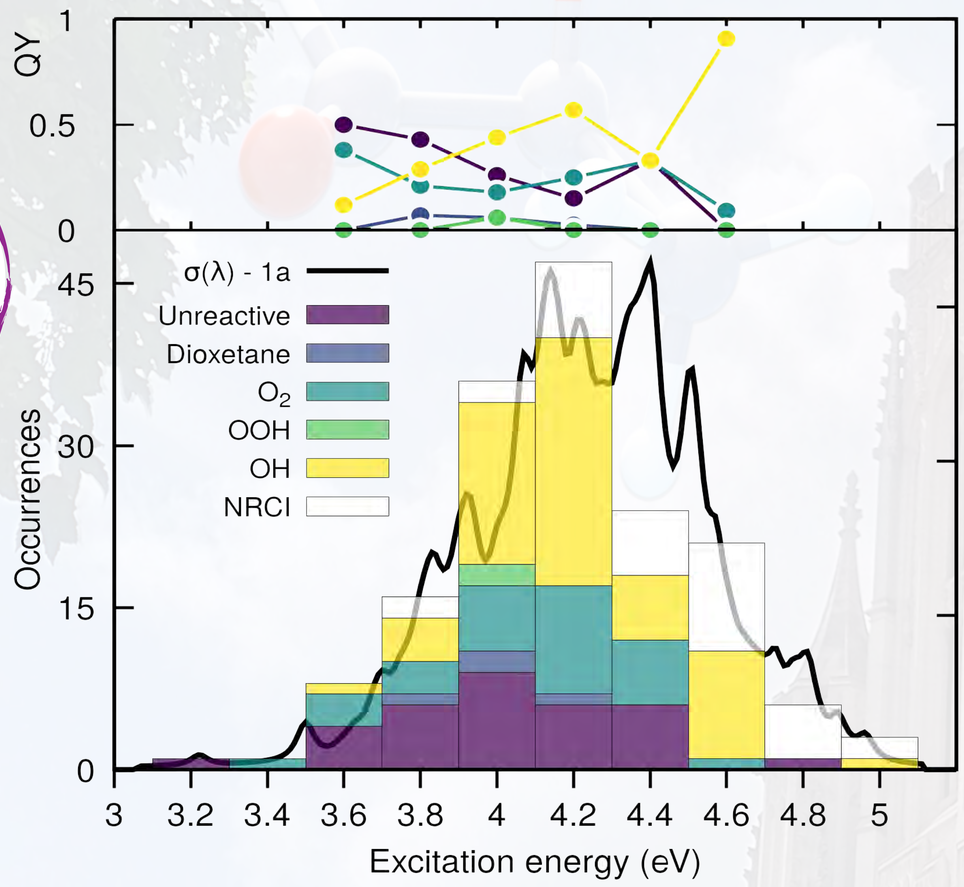
- σ(λ) is estimated by employing the **nuclear ensemble approach** (NEA). The geometries were optimized at the SCS-MP2/def2-SVP level of theory. Vertical transitions and oscillator strengths were evaluated with SCS-ADC(2)/def2-SVP.
- ϕ(λ) is predicted using **nonadiabatic molecular dynamics** simulations with the Tully surface hopping algorithm. We initiated the excited state dynamics with SCS-ADC(2)/def2-SVP and we switch to XMS(3)-CASPT2/cc-pVDZ to describe the S_1/S_0 nonadiabatic transitions and the following ground state dynamics.

- The main contribution to σ(λ) in the solar actinic region is from the $S_0 \rightarrow S_1$ transition. It corresponds to an $n(O) \rightarrow \pi^*(CO)$ character.
- The composite cross-section is based on the experimental photoabsorption cross-section of methylhydroperoxide and propanal.

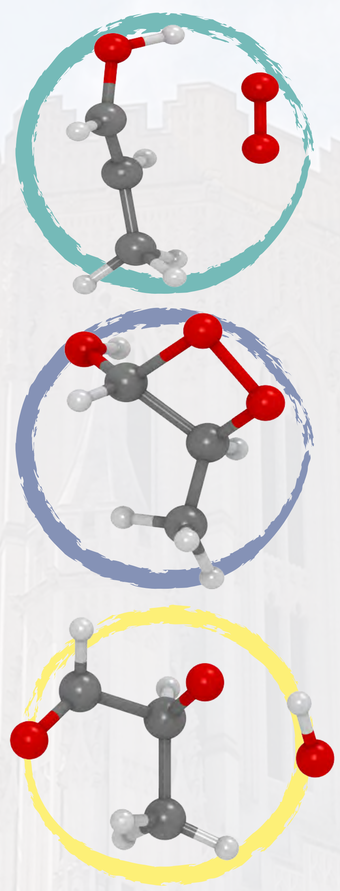


σ(λ)

- The most relevant process occurring at low excitation energies is the excited-state proton-coupled electron transfer. This leads to the release of $^1O_2$, return to the Frank-Condon region following the nonradiative pathway to S_0, or even more exotic photoproduct such as the formation of dioxetane ring.
- The OH photodissociation gains importance when exciting 2-HPP with higher-energy photons.



ϕ(λ)



Conclusions

The photolysis rates for many transient VOCs are hardly available based on experiments alone. We have used a fully in silico protocol to investigate the photolysis of 2-HPP, an atmospherically relevant multifunctional molecule from the family of alpha-hydroperoxycarbonyls.

We calculate σ(λ) and ϕ(λ), the key observables to evaluate J of the most important photolysis channels. We have shown the photolysis processes of 2-HPP have sizable rates with an impact on the chemical balance of the atmosphere although not as important as estimated by previous works.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 803718, project SINDAM) and the EPSRC Grant EP/V026690/1

Marsili Emanuele, Antonio Prlj, and Basile FE Curchod. "A Theoretical Perspective on the Actinic Photochemistry of 2-Hydroperoxypropanal." *J. Phys. Chem. A,* **2022**, 126, 5420-5433
Marsili Emanuele, Antonio Prlj, and Basile FE Curchod. "Caveat when using ADC(2) for studying the photochemistry of carbonyl-containing molecules." *Phys. Chem. Chem. Phys.*, **2021**, 23, 12945-12949.

An Immersed Boundary Method for the DNS Solver CHAPSIM

Kenneth Chinembiri
University of Sheffield

CHAPSim 2.0 is a Direct Numerical Simulation code developed by the Collaborative Computational Project – Nuclear Thermal Hydraulics (CCP-NTH) as an open-source UK nuclear community code. The solver is fast, efficient, and capable of simulating turbulent thermal flows with strong physical property variation. This paper discusses the methodology and validation of an Immersed Boundary Method (IBM) for complex geometries in the solver CHAPSim 2.0. When adopting this method, the effect of the solid body to flow field is mimicked by introducing a forcing term to the governing momentum equations of the CFD solver. The forcing term allows the user to impose a desired target velocity at the grid nodes of the complex solid boundary and is computed courtesy of the direct forcing approach. This function enables CHAPSim2 to simulate flow over arbitrary geometry without complicated grid generation process.

An Immersed Boundary Method for the DNS Solver CHAPSim

Kenneth Chinembiri[1], Shuisheng He[1], Wei Wang[2]

[1] Department of Mechanical Engineering, The University of Sheffield

[2] Scientific Computing Department, STFC, Daresbury Laboratory

Computing Insight UK 2022, Manchester, 1st - 2nd December 2022

1. Introduction

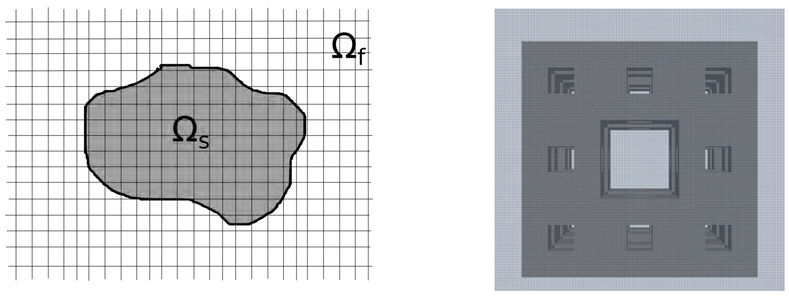


Fig.1 Complex geometries being represented on cartesian grids

- Immersed Boundary Methods are a set of techniques used to mimic the effect of a complex geometry in fluids on a structured grid.
- Being developed in CHAPSim 2.0 to study effect of roughness on heat transfer.
- CHAPSim 2.0 is an open-source DNS solver of high accuracy and scalability for turbulent thermal flows with significant property variation. (git@github.com:CHAPSim/CHAPSim2.git)

2. Aim & Objectives

- Implement IBM methods in CHAPSim 2.0
- Extend the IBM approach to account for Conjugate Heat Transfer.

3. Methodology

i. Geometry description

Steps for capturing the solid geometry are:

a. STL files to define Immersed surface.
b. Boundary nodes detected using locate → bounding box → ray-triangle intersection
c. Solid nodes detected using one-direction backfilling approach

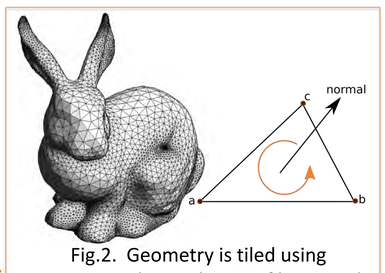


Fig.2. Geometry is tiled using unstructured triangles. STL files provide vertex and unit normal per facet

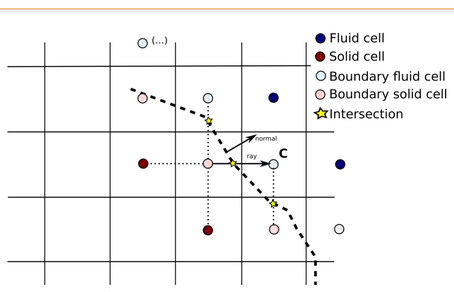


Fig.3. Checking the intersection between triangular element and ray. Three steps (i) check intersection occurs with plane, (2) check if intersect within facet & (3) determine inside or outside

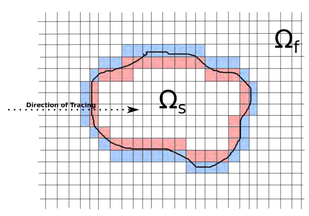


Fig.4. Solid nodes tagged by moving in one-direction

ii. Reconstructing the boundary

Focus on rigid body boundary reconstruction

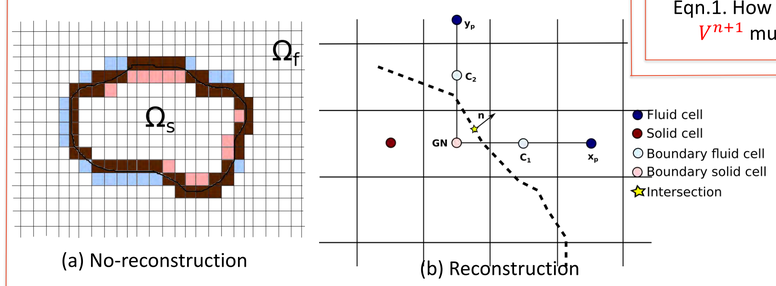
a. Direct forcing approach [1]
b. Ghost node approach with no forcing (not shown) [2]

$$\frac{\partial u}{\partial t} = -\nabla(uu) - \frac{\nabla p}{\rho} + v\nabla^2 u + f$$

$$\frac{u^{n+1} - u^n}{\Delta t} = \mathrm{RHS}^{n+\theta} + f^{n+\theta}$$

$$f^{n+\theta} = \frac{V^{n+1} - u^n}{\Delta t} - \mathrm{RHS}^{n+\theta}$$

Eqn.1. How to define the forcing term. Note, V^{n+1} must be determined beforehand



(a) No-reconstruction (b) Reconstruction

Fig.5. How to determine V^{n+1}. For method (b), we interpolate to image point along normal using inverse distance weighing. Then a Lagrange polynomial from image point to GN.

iii. Treatment of internal flow

a. Zero pressure correction of all solid nodes
b. Zero pressure correction for only solid boundary nodes
c. Allow pressure correction on all nodes

$$\frac{u^* - u^k}{\Delta t} = a_k F^k + b_k F^{k-1} - c_k \nabla p^k + c_k f^{k+1}$$

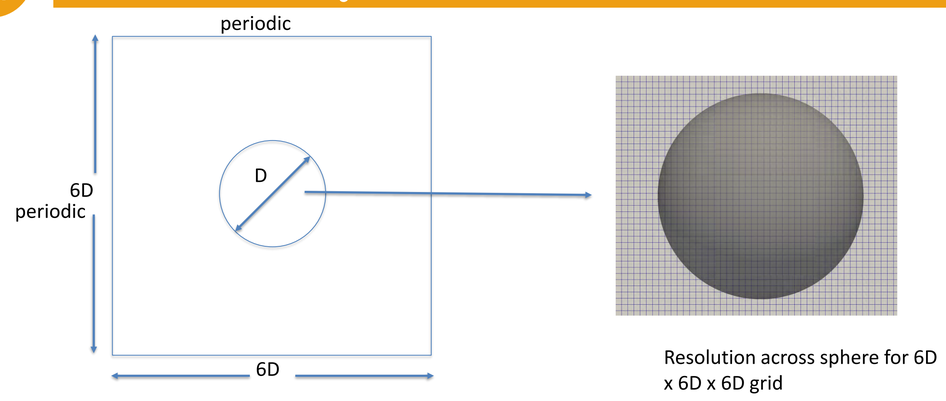
$$\frac{u^{**} - u^*}{\Delta t} = c_k \nabla p^k$$

$$\frac{u^{k+1} - u^{**}}{\Delta t} = -c_k \nabla p^{k+1}$$

In general: $u^{n+1} \neq V^{n+1}$

Eqn.2. How to treat pressure in the internal nodes ? Balance between stability and accuracy. Method (c) is more stable but alters target velocity. Methods (a) & (b) increased accuracy but prone to diverge or generating localized peaks.

4. Preliminary results



Resolution across sphere for 6D x 6D x 6D grid

Fig.6. Schematic of the computational domain ()left) and grid resolution across sphere (right)

- Preliminary validation based on 3-D flow over a 3-D smooth sphere.
- Periodic boundary conditions in all directions (not ideal but suffices for testing)
- Direct forcing approach with pressure correction at all nodes used in test.
- Sixth order compact scheme for spatial discretisation

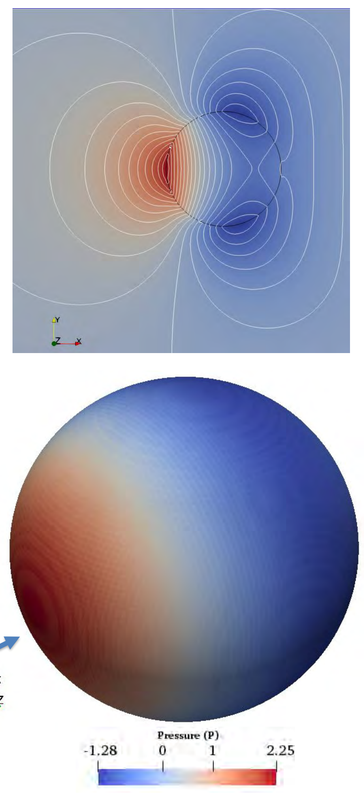


Fig.7. Pressure field at Re_D=30

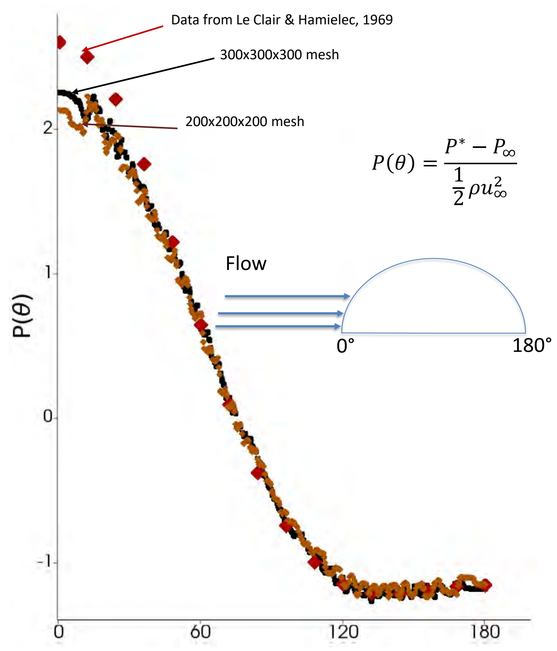


Fig.8. Comparison of pressure coefficient at Re_D = 5 against data by Le Clair and Hamielec, 1969.

5. Future Work

- Complete validation process using spheres but with more appropriate boundary conditions
- Complete pyramid roughness tests and verification (see Fig.9 and Fig.10)
- Complete conjugate heat transfer implementation

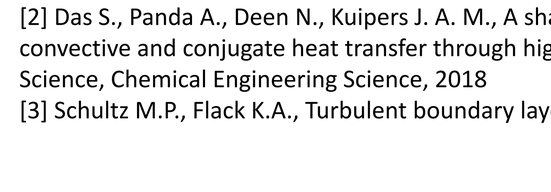


Fig.9. Pyramid roughness based on experiments by M. Schultz & K. Flack [3]

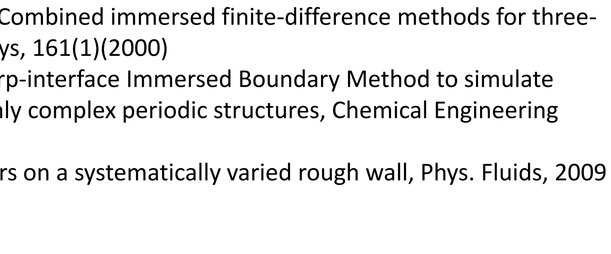


Fig.10. Present tests for Pyramid roughness with CHAPSim2. Note still testing and case yet to be fully verified

6. Acknowledgements

- This work made use of computational support by CoSeC, the Computational Science Centre for Research Communities, through CCP-NTH (Grant Ref: EP/T026685/1).
- This project received funding from the Euratom research and training program 2019-2020 under Grant agreement No 945324

References

[1] Fadlun E., Verzicco R., Orlandi P., Mohd-Yusof J., Combined immersed finite-difference methods for three-dimensional comples flow simulations, J Comput Phys, 161(1)(2000)

[2] Das S., Panda A., Deen N., Kuipers J. A. M., A sharp-interface Immersed Boundary Method to simulate convective and conjugate heat transfer through highly complex periodic structures, Chemical Engineering Science, Chemical Engineering Science, 2018

[3] Schultz M.P., Flack K.A., Turbulent boundary layers on a systematically varied rough wall, Phys. Fluids, 2009


COMPUTING
INSIGHT UK 2023
UKRI
Science and
Technology
Facilities Council
SAVE
THE
DATE
7-8 December 2023
Manchester Central, UK
SEE YOU IN 2023
www.stfc.ac.uk/ciuk