Science & Technology Facilities Council

Application Performance Evaluation Studies of Multi- Core Nodes and the Gemini Network

Mike Ashworth, Xiaohu Guo, Charles Moulinec, Stephen Pickles, Martin Plummer, Andrew Porter, Andrew Sunderland and Ilian Todorov

Computational Science & Engineering Department,
STFC Daresbury Laboratory,
Warrington WA4 4AD, UK

Outline

Introduction to the UK HECToR system

Applications

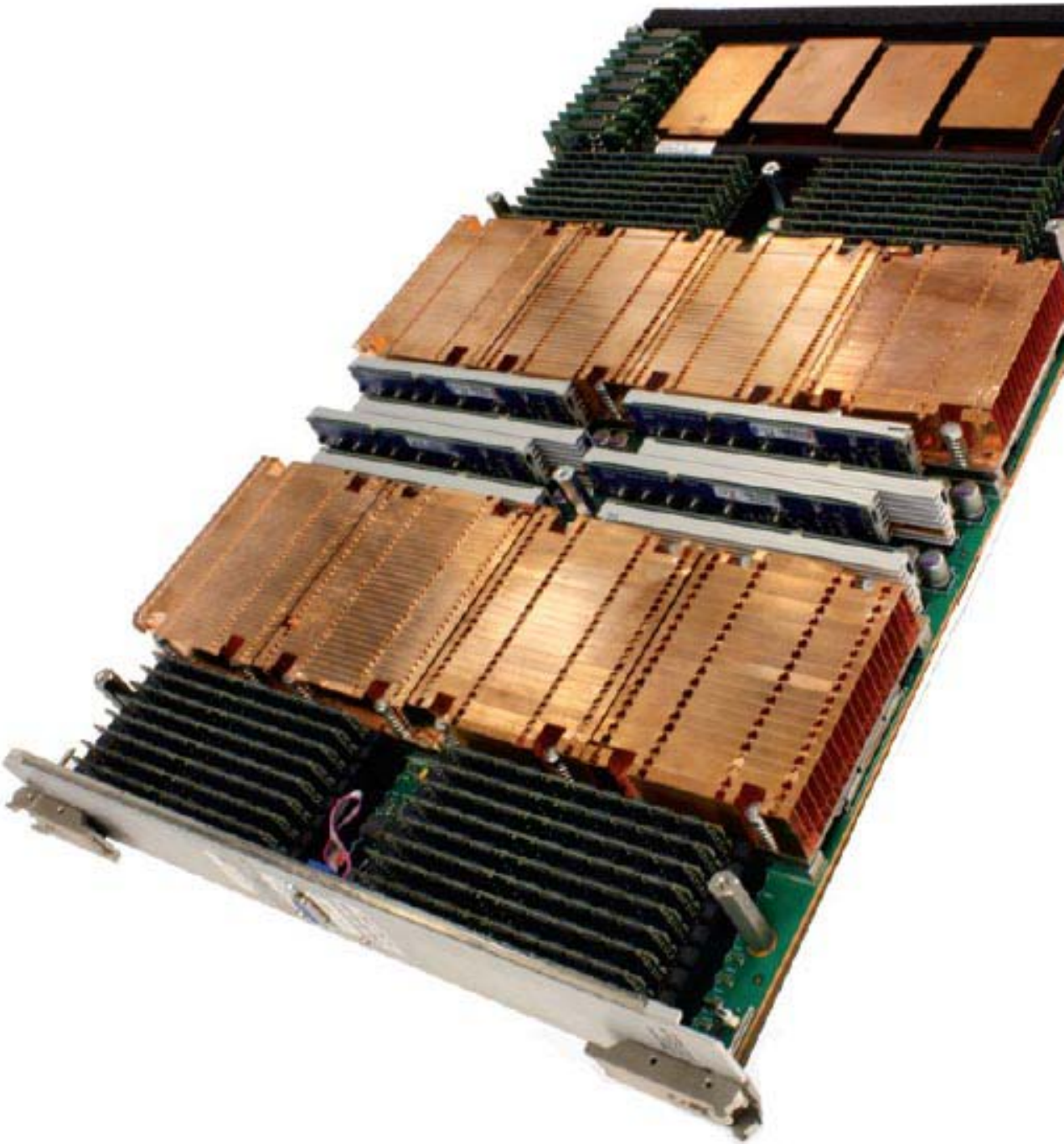
- DL_POLY_4 (see paper)
- fd3d
- Fluidity-ICOM
- PFARM
- POLCOMS
- ScaLAPACK
- Telemac
- WRF

Conclusions

HECToR -
High End Computing Technology Resource

UK National HPC service

http://www.hector.ac.uk/



HECToR XT4 vs. XE6

HECToR phase2a XT4

Processor
- AMD Barcelona quad-core

Core
- 2.3Ghz clock frequency
- SSE SIMD FPU (4flops/cycle = 9.2GF peak)

Memory
- 16 GB/node symmetric
- DDR2
- 12GB/s peak @ 800MHz

Interconnect
- SeaStar

HECToR phase2b XE6

Processor
- AMD Magny-Cours 24-core

Core
- 2.1Ghz clock frequency
- SSE SIMD FPU (4flops/cycle = 8.4 GF peak)

Memory
- 16 GB/node symmetric
- DDR3
- 85GB/s peak @ 1333MHz

Interconnect
- Gemini

HECToR 'interim' phase2b XT6

Same network as XT4, same processors as XE6

Science & Technology
Facilities Council

fd3d

Science & Technology Facilities Council

Large subduction earthquakes

On 19th Sep 1985 a large Ms 8.1 subduction earthquake occurred on the Mexican Pacific coast with an epicentre at about 340 km from Mexico City with about 30,000 deaths and losses of $7 billion.

On 12th May 2008 the Ms 7.9 Sichuan, China, earthquake produced about 70,000 deaths and $80 billion losses

On 11th March 2011 the Mw 9.0 Tohoku, Japan, earthquake resulted in about 15,000 deaths, $15-$30 billion losses

Therefore, there is a seismological, engineering and socio economical interest to model these types of events, particularly, due to the scarcity of observational instrumental data for them

Science & Technology Facilities Council

fd3d earthquake simulation code

Seismic wave propagation

3D velocity-stress equations

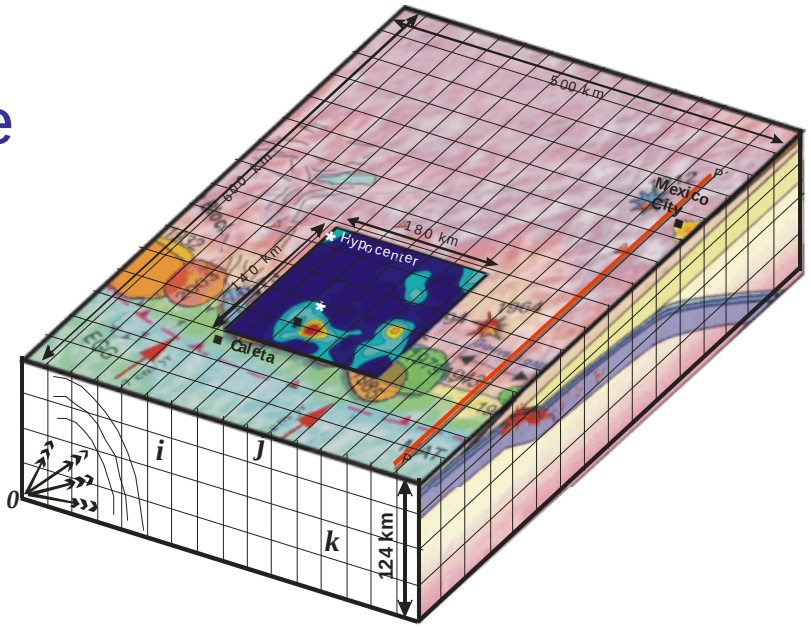
Structured grid

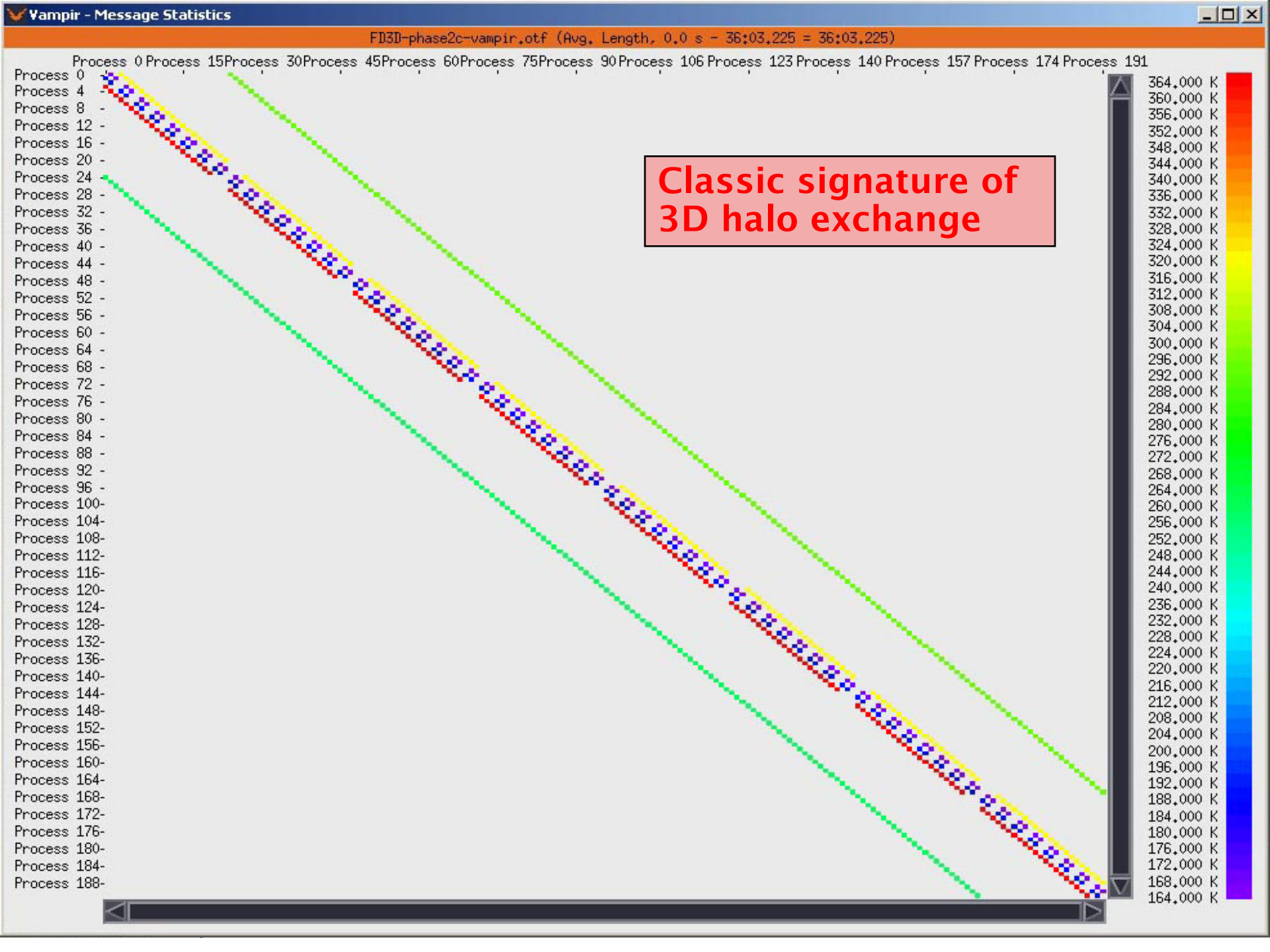
Explicit scheme

- 2nd order accurate in time
- 4th order accurate in space

Regular grid partitioning

Halo exchange



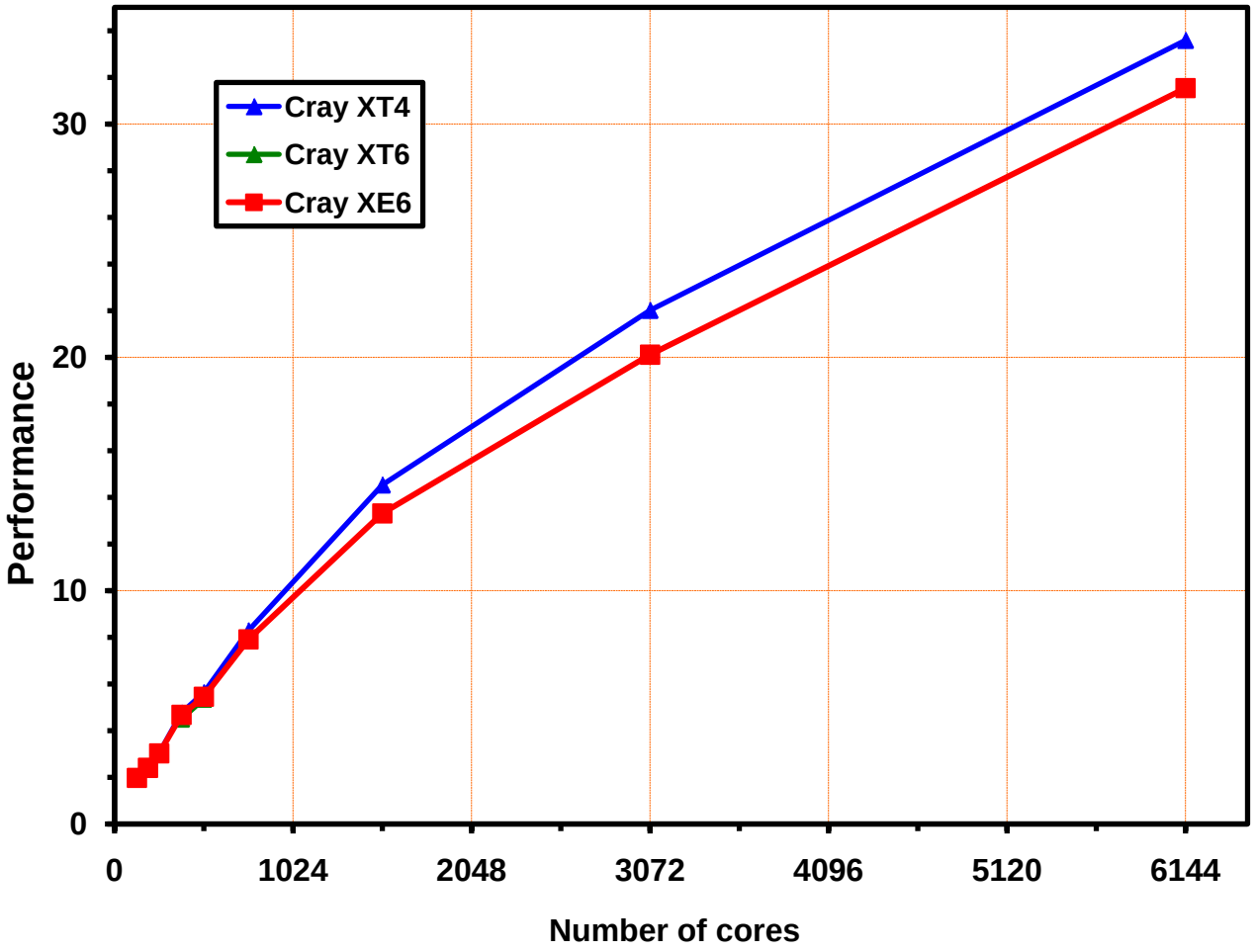

Vampir - Message Statistics
FD3D-phase2c-vampir.otf (Avg. Length, 0.0 s - 36:03.225 = 36:03.225)
Process 0 Process 15 Process 30 Process 45 Process 60 Process 75 Process 90 Process 106 Process 123 Process 140 Process 157 Process 174 Process 191
Process 0
Process 4
Process 8
Process 12
Process 16
Process 20
Process 24
Process 28
Process 32
Process 36
Process 40
Process 44
Process 48
Process 52
Process 56
Process 60
Process 64
Process 68
Process 72
Process 76
Process 80
Process 84
Process 88
Process 92
Process 96
Process 100
Process 104
Process 108
Process 112
Process 116
Process 120
Process 124
Process 128
Process 132
Process 136
Process 140
Process 144
Process 148
Process 152
Process 156
Process 160
Process 164
Process 168
Process 172
Process 176
Process 180
Process 184
Process 188
364.000 K
360.000 K
356.000 K
352.000 K
348.000 K
344.000 K
340.000 K
336.000 K
332.000 K
328.000 K
324.000 K
320.000 K
316.000 K
312.000 K
308.000 K
304.000 K
300.000 K
296.000 K
292.000 K
288.000 K
284.000 K
280.000 K
276.000 K
272.000 K
268.000 K
264.000 K
260.000 K
256.000 K
252.000 K
248.000 K
244.000 K
240.000 K
236.000 K
232.000 K
228.000 K
224.000 K
220.000 K
216.000 K
212.000 K
208.000 K
204.000 K
200.000 K
196.000 K
192.000 K
188.000 K
184.000 K
180.000 K
176.000 K
172.000 K
168.000 K
164.000 K
Classic signature of
3D halo exchange

fd3d performance on XT4, XT6 and XE6

62.5m resolution model of the Parkfield, CA, quake

Little performance difference

Comms speed-up on XE6 more than offset by memory contention



Fluidity- ICOM

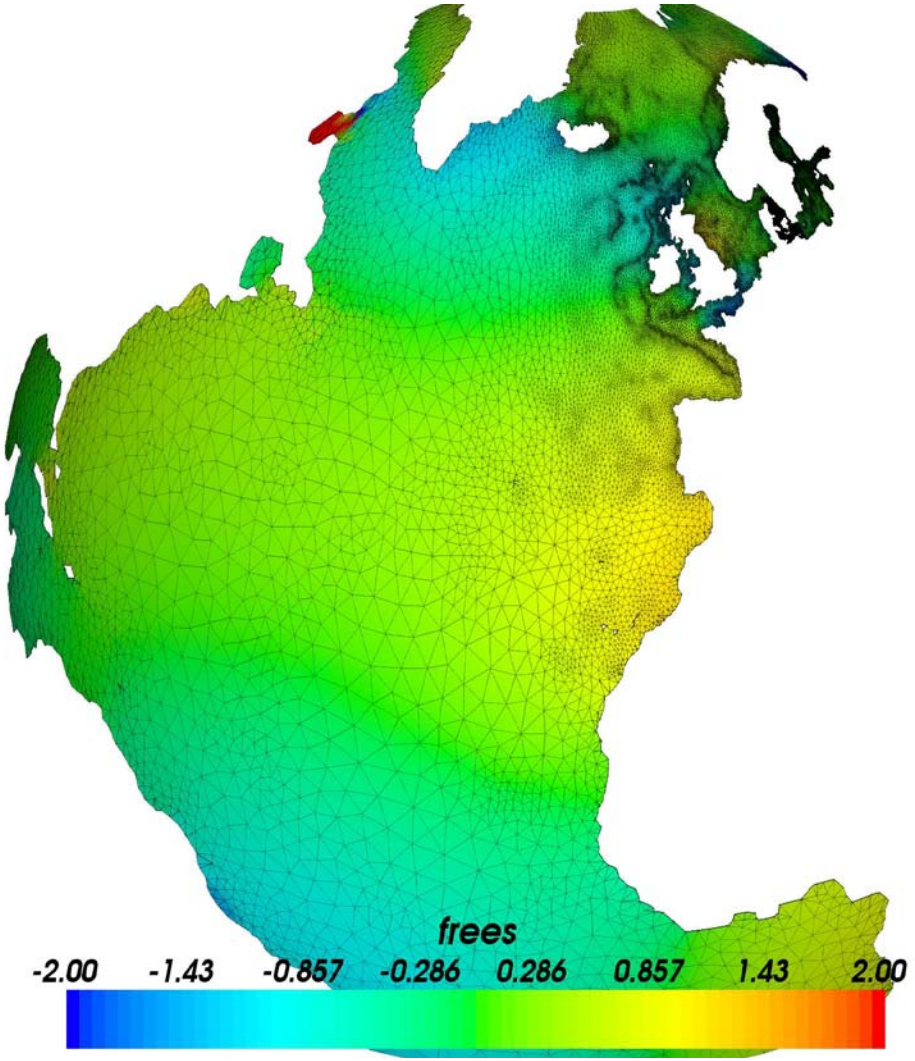
Science & Technology Facilities Council

Unstructured Mesh Ocean Modelling

Fluidity-ICOM is build on top of Fluidity, an adaptive unstructured finite element code for computational fluid dynamics

The Imperial College Ocean Model (ICOM) has the capability to efficiently resolve a wide range of scales simultaneously

This offers the opportunity to simultaneously resolve both basin-scale circulation and small-scale processes



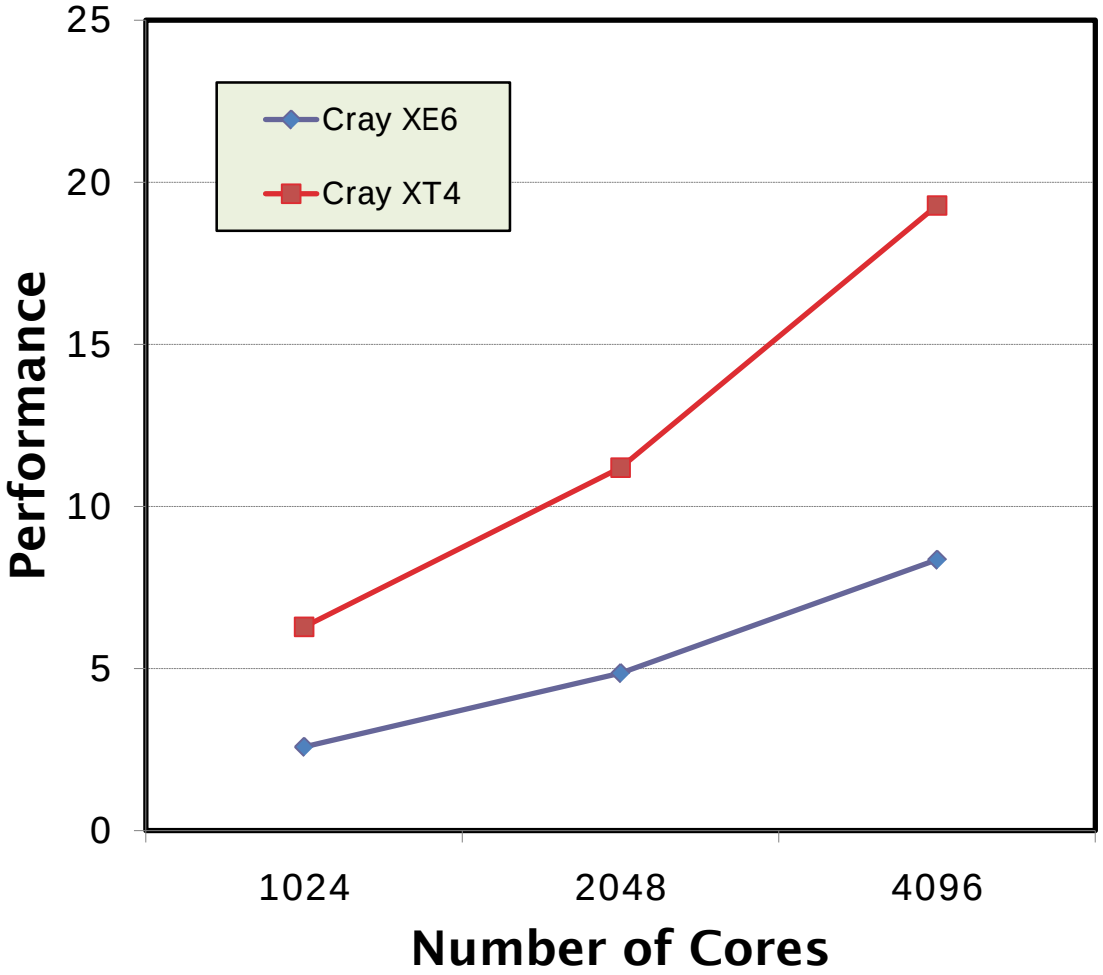
Fluidity- ICOM on the Cray XT4 and XE6

10 million vertex benchmark case

Performance of momentum-solve shows much worse performance on XE6

Presumed due to memory contention between 24 cores on a node vs. quad-core XT4

Part of ongoing performance investigations



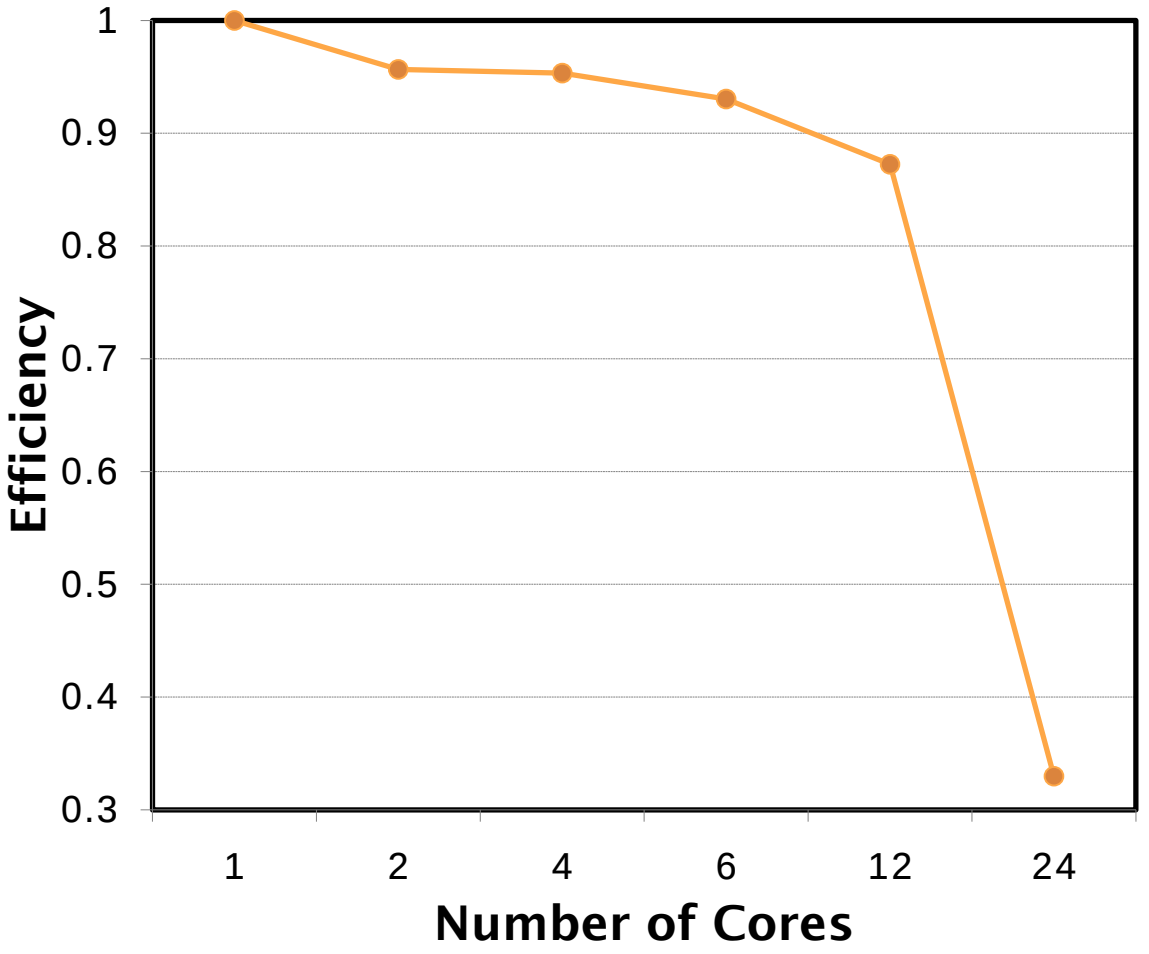
Fluidity- ICOM on the Cray XT4 and XE6

Current work focusing on hybrid MPI-OpenMP

Momentum matrix assembly

Efficiency is good out to 6 threads / 4 tasks per node

Allows us to reduce MPI tasks to 4 tasks per node and decrease memory footprint



Science & Technology
Facilities Council

PFARM

Science & Technology Facilities Council

Atomic Molecular and Optical Physics

Electron and photon collisions with atoms and ions

Applications in …

Astrophysics: understanding of scattering and excitation processes which power light emission from nebulae

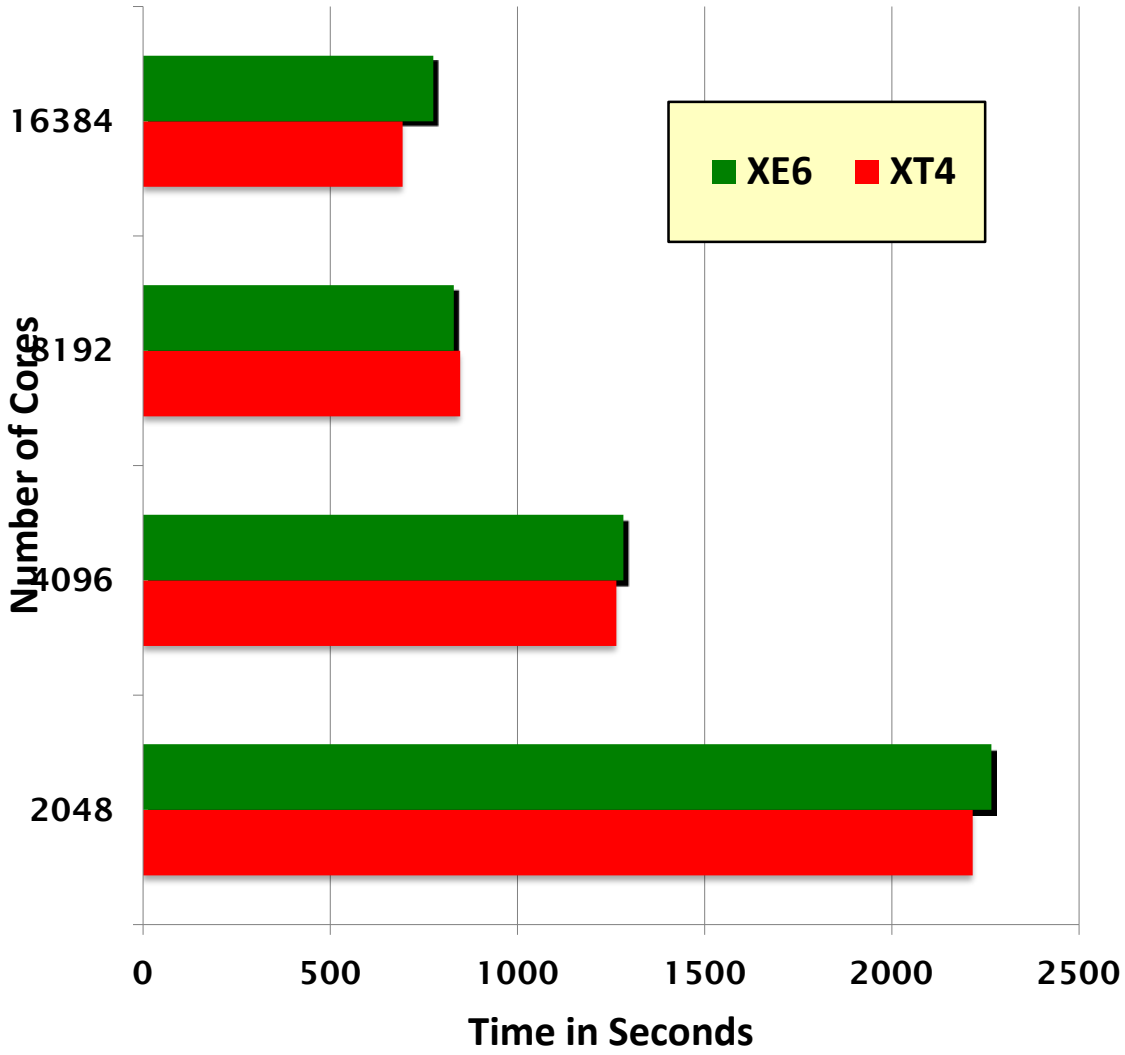
Lasers: exciting, new field of high-powered lasers. Short, very high intensity pulses of light can blow atoms apart. This process could one day be used to control the outcome of chemical reactions - among the many applications envisaged.

External region code EXAS on XT4 and XE6

FeIII scattering case, involving 21080 scattering energies

Timing reveals that initialization costs increase markedly on the XE6 and grow with core count

Subject for future optimization



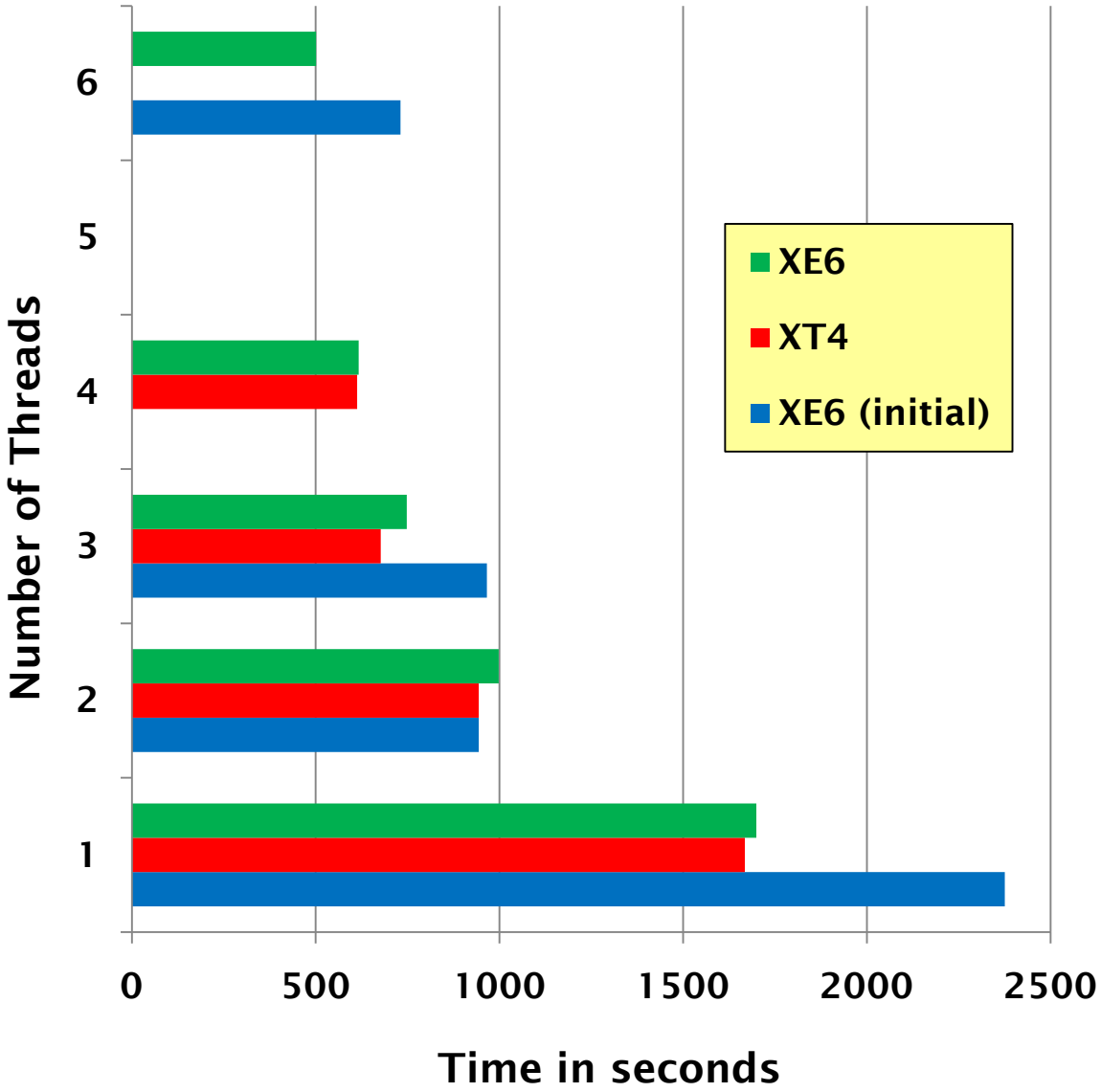
Internal region code RAD on XT4 and XE6

Electron-oxygen atom scattering case

OpenMP utilized for up to 6 threads per task (XE6), 4 threads (XT4)

Subject for current optimization project – initial improvement shown

XE6 slower by clock ratio 2.1/2.3 e.g. 3 threads



Science & Technology
Facilities Council

POLCOMS

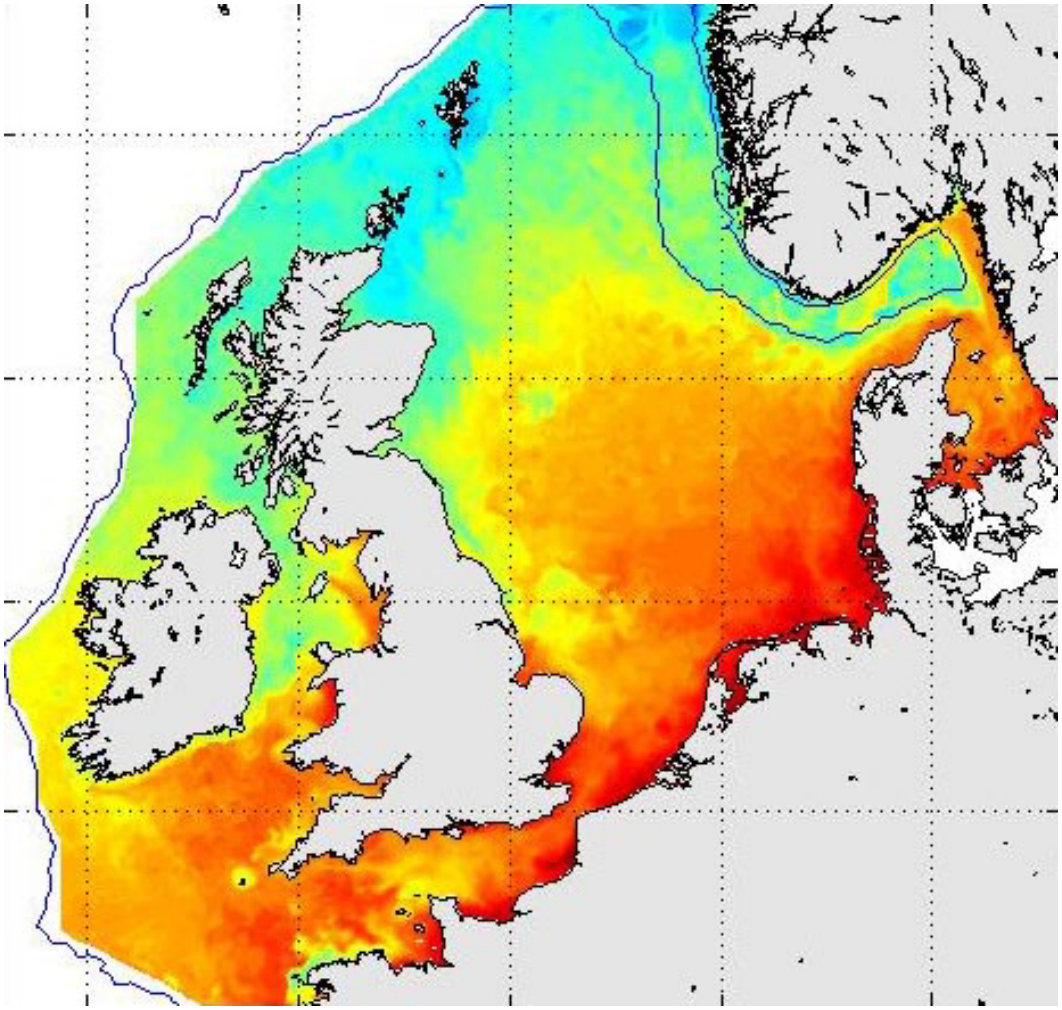
Science & Technology Facilities Council

High- Resolution Coastal Ocean Modelling

POLCOMS is the finest resolution model to-date to simulate the circulation, temperature and salinity of the Northwest European continental Shelf

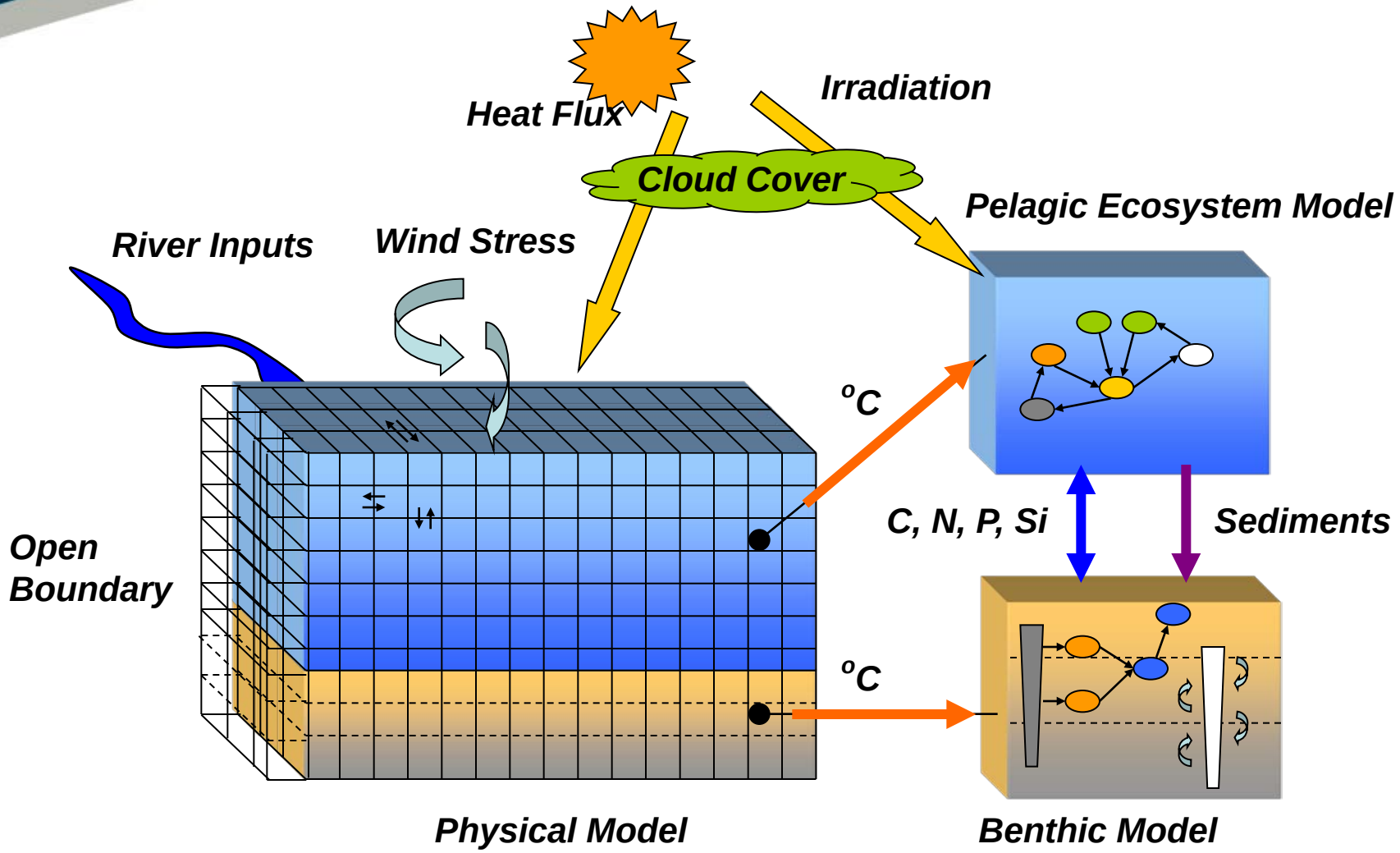
Important for understanding the transport of nutrients, pollutants and dissolved carbon around shelf seas

We have worked with POL on coupling with ERSEM, WAM, CICE, data assimilation and optimisation for HPC platforms



Summer surface temperature, 2km resolution

Coupled Marine Ecosystem Model



Science & Technology Facilities Council

POLCOMS Halo Exchange on XT4, XT6, XE6

Performance – high is good

	240 cores			360 cores		
Operation	XT4	XT6	XE6	XT4	XT6	XE6
2D	6818	2700	36913	7366	2272	30628
3D	3273	1174	6451	3841	1229	7552
Mixed-D	3250	1171	6032	3670	1194	7292

Pure MPI, one task per core

XT6 performance poor – network poorly matched to 24-way nodes
XE6 MUCH improved

2D latency limited
XE6/XT6 10x speed- up

3D bandwidth limited
XE6/XT6 5x speed- up

See paper for multi-core aware partitioning (Pickles, CUG 2010)

Science & Technology
Facilities Council

ScaLAPACK

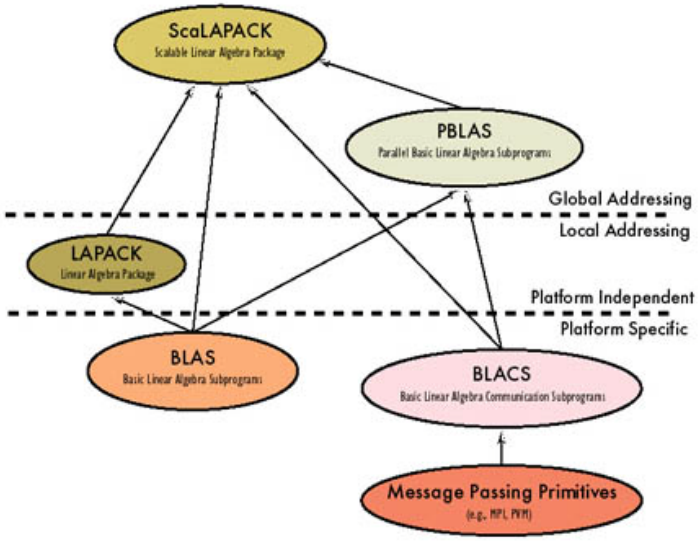
Subset of LAPACK routines redesigned for distributed memory MIMD parallel computers

Widely used in a range of STFC applications include PRMAT, CRYSTAL, GAMESS-UK, KPPW and CASTEP

dependent upon efficient parallel symmetric diagonalizations

ScaLAPACK

A Software Library for Linear Algebra Computations on Distributed-Memory Computers



AVAILABLE SOFTWARE:

Dense, Band, and Tridiagonal Linear Systems
- general
- symmetric positive definite

Full-Rank Linear Least Squares

Standard and Generalized Orthogonal Factorizations

Eigensolvers
- SEP: Symmetric Eigenproblem
- NEP: Nonsymmetric Eigenproblem
- GSEP: Generalized Symmetric Eigenproblem

SVD

Prototype Codes
- HPF interface to ScaLAPACK
- Matrix Sign Function for Eigenproblems
- Out-of-core solvers (LU, Cholesky, QR)
- Super LU
- PBLAS (algorithmic blocking and no alignment restrictions.)

DOCUMENTATION:

ScaLAPACK Users' Guide
http://www.netlib.org/scalapack/slug/scalapack_slug.html

Future Work
- Out-of-core Eigensolvers
- Divide and Conquer routines
- C++ and Java Interfaces

Commercial Use

ScaLAPACK has been incorporated into the following software packages:
- NAG Numerical Library
- IBM Parallel ESSL
- SGI Cray Scientific Software Library

and is being integrated into the VNI IMSL Numerical Library, as well as software libraries for Fujitsu, HP/Convex, Hitachi, and NEC.

http://www.netlib.org/scalapack/

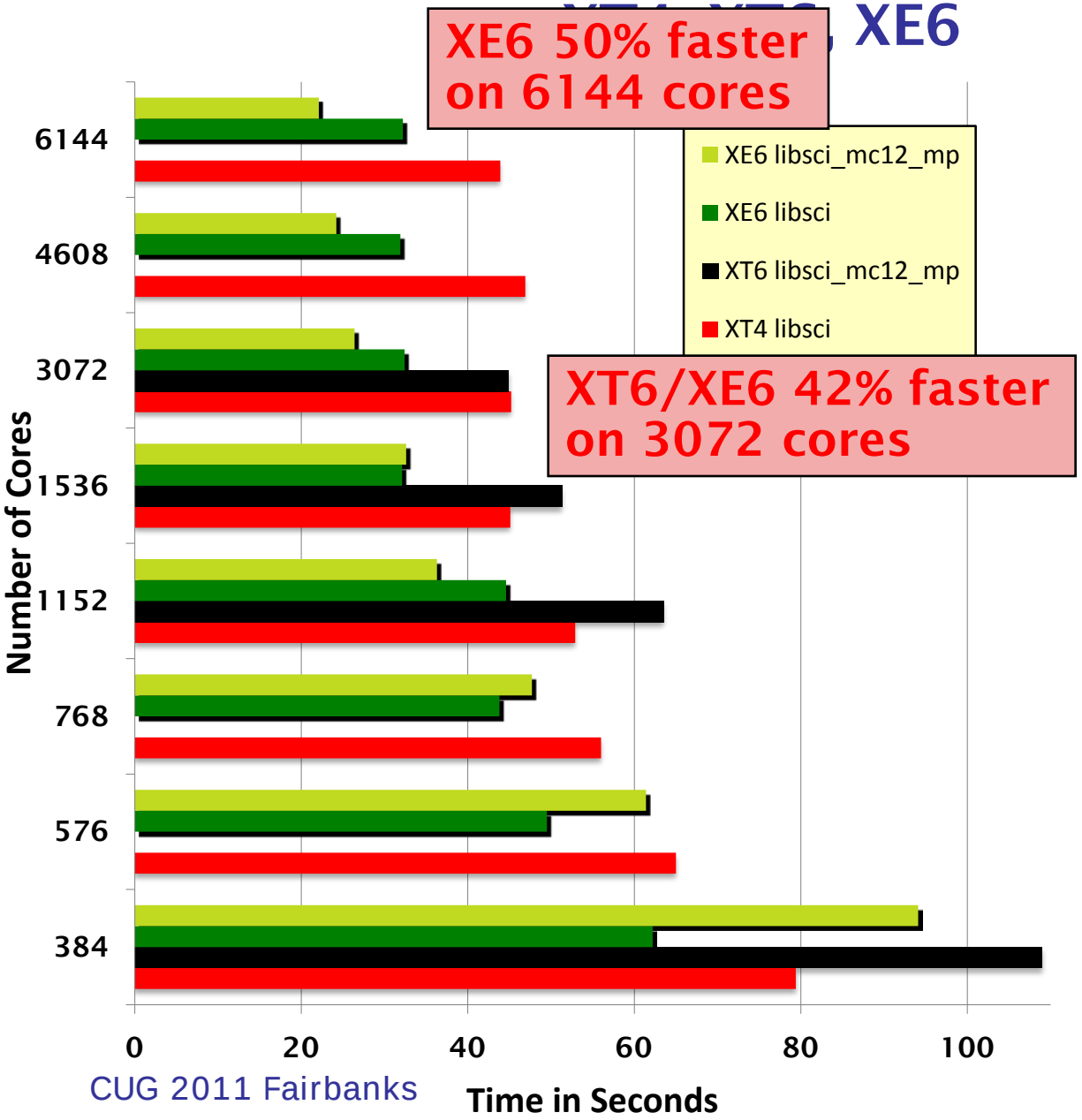
The University of Tennessee Oak Ridge National Laboratory

ScaLAPACK: Timings XE6

Timings for parallel PDSYEVD-based eigensolves for the CRYSTAL 20480 matrix on Cray XT4, XT6, XE6 platforms

"mp" indicates hybrid multi-threading 2 MPI tasks per 12-core processor

XE6 faster on high core counts (lower MPI overheads) and for hybrid execution



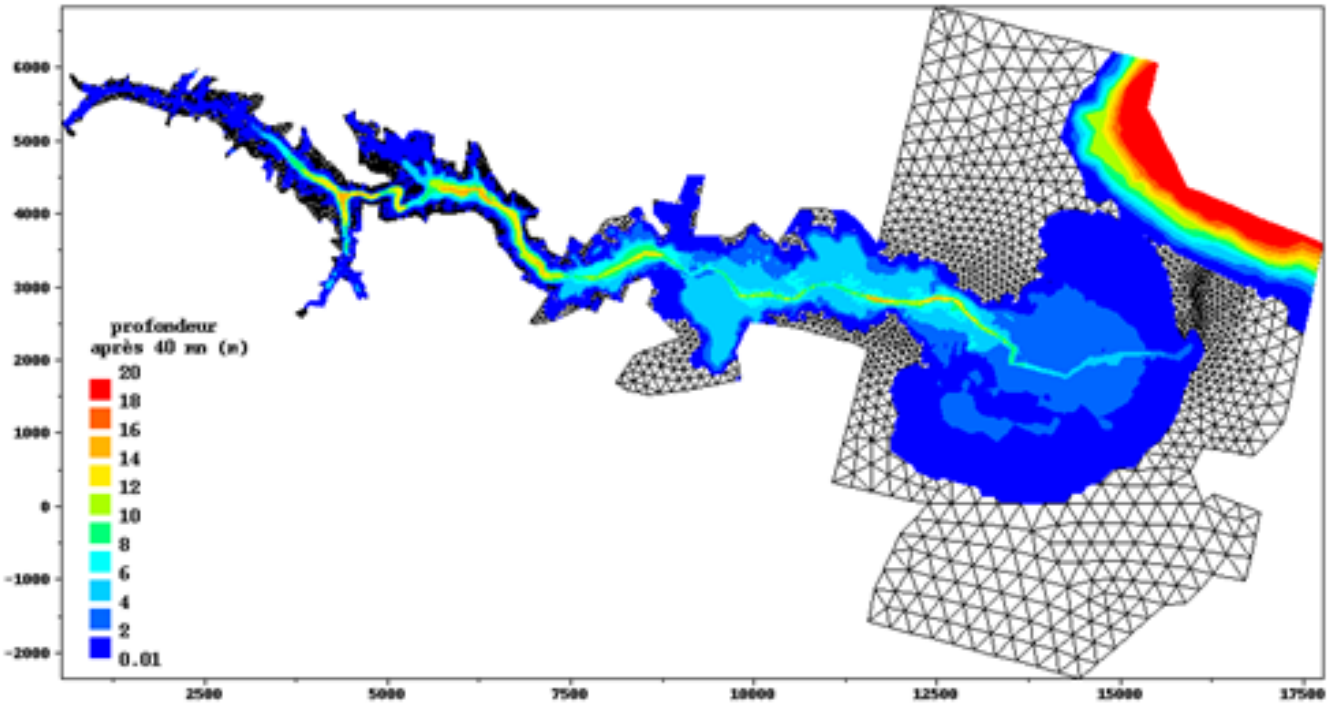
Science & Technology
Facilities Council

Telemac

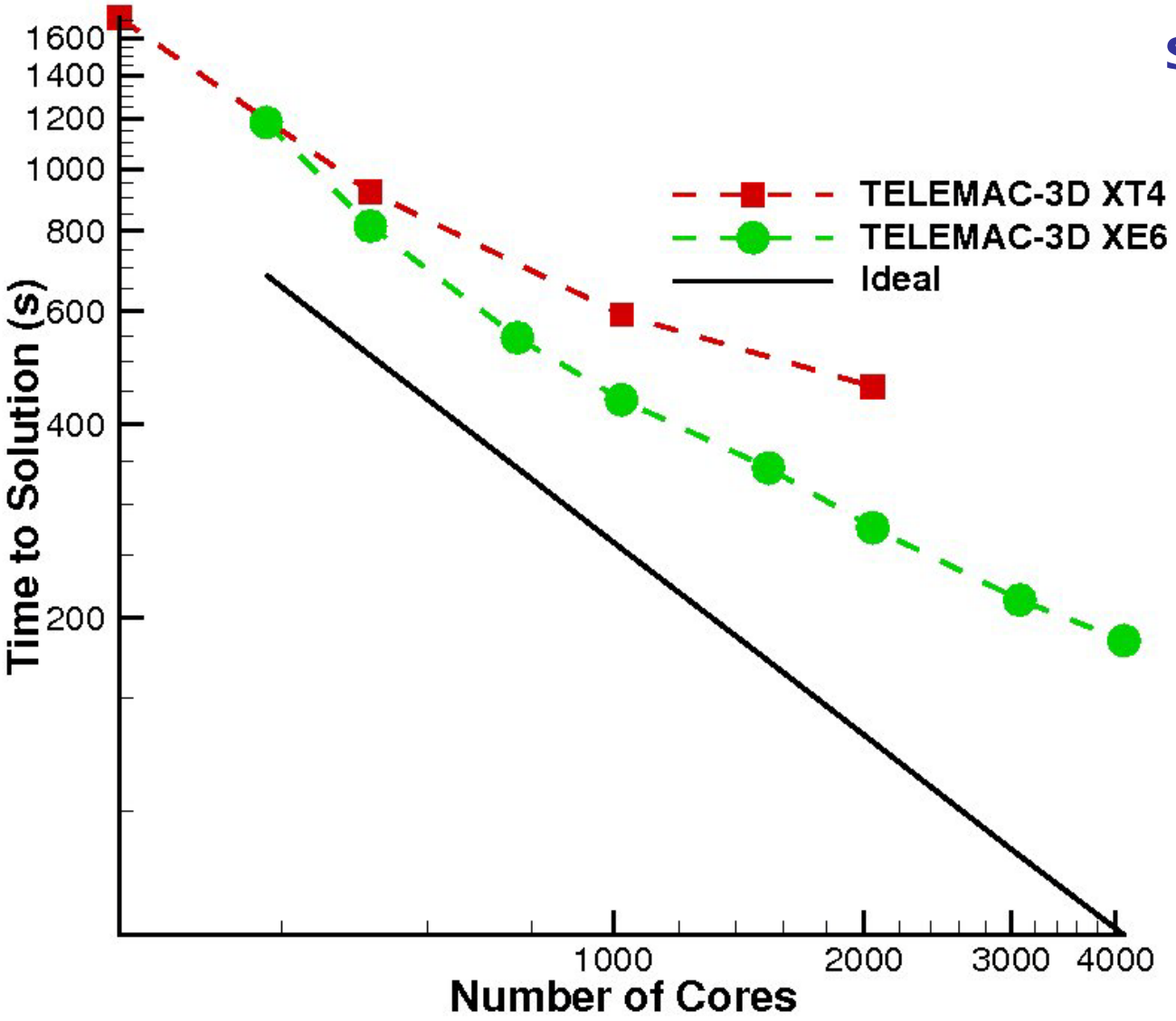
Telemac: free surface flows

The software suite Telemac, dedicated to free surface flows, has seen a growing success since 1993 and has been widely distributed throughout the world, with more than 200 licences and several hundreds of users.

Simulation of the Malpasset dam break flood wave in 1959, with a 26000 elements mesh (the run, 1000 time steps of 4 s, takes 10 s on an 8-core desktop computer)



Telemac: time to solution



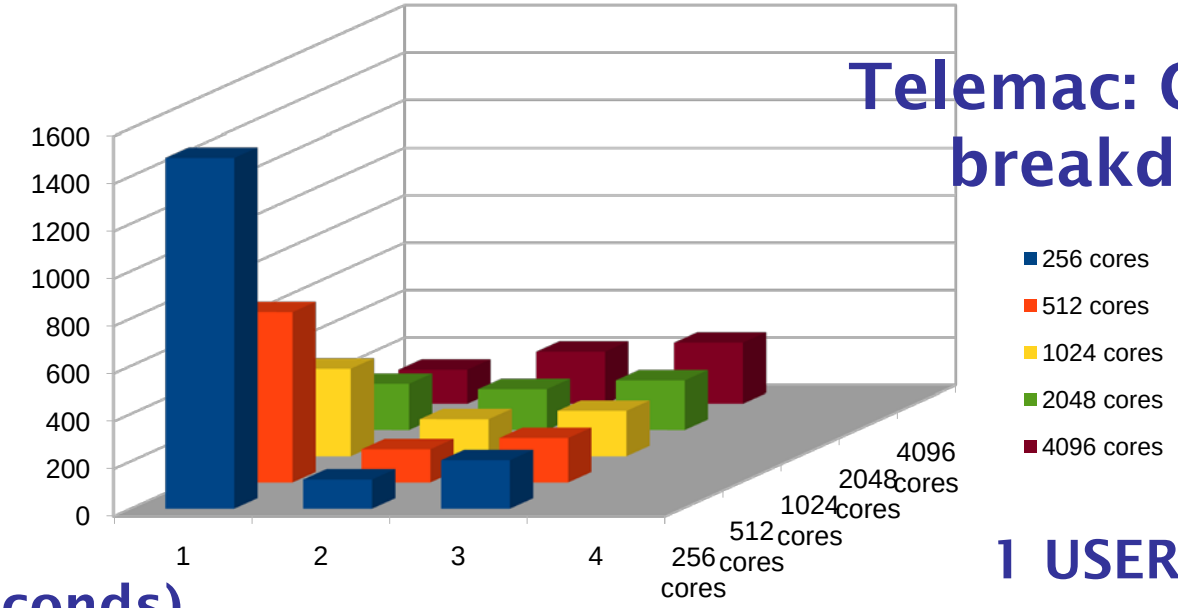
Model from a study the impact of fresh water release from a hydro-electric power plant in the Berre lagoon (in the south of France)

3-D model based on 0.4 M 2-D triangles

31 layers would yield 12M triangles

Telemac: Craypat breakdown

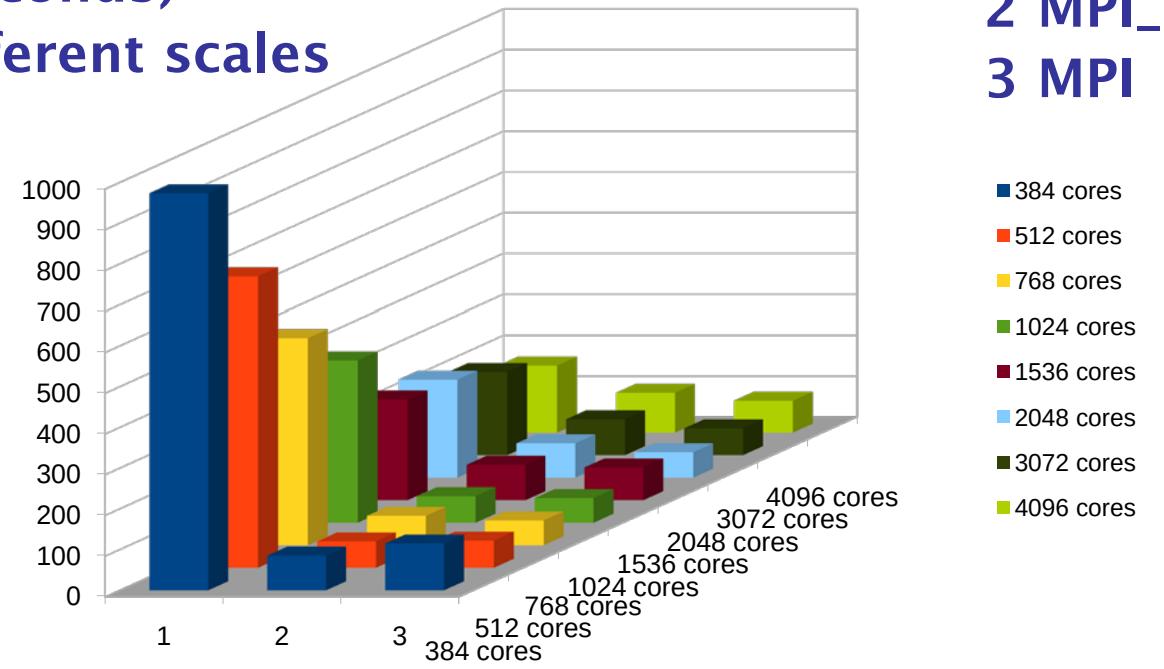
Cray XT4



Time (seconds)
Note different scales

1 USER
2 MPI_SYNC
3 MPI

Cray XE6



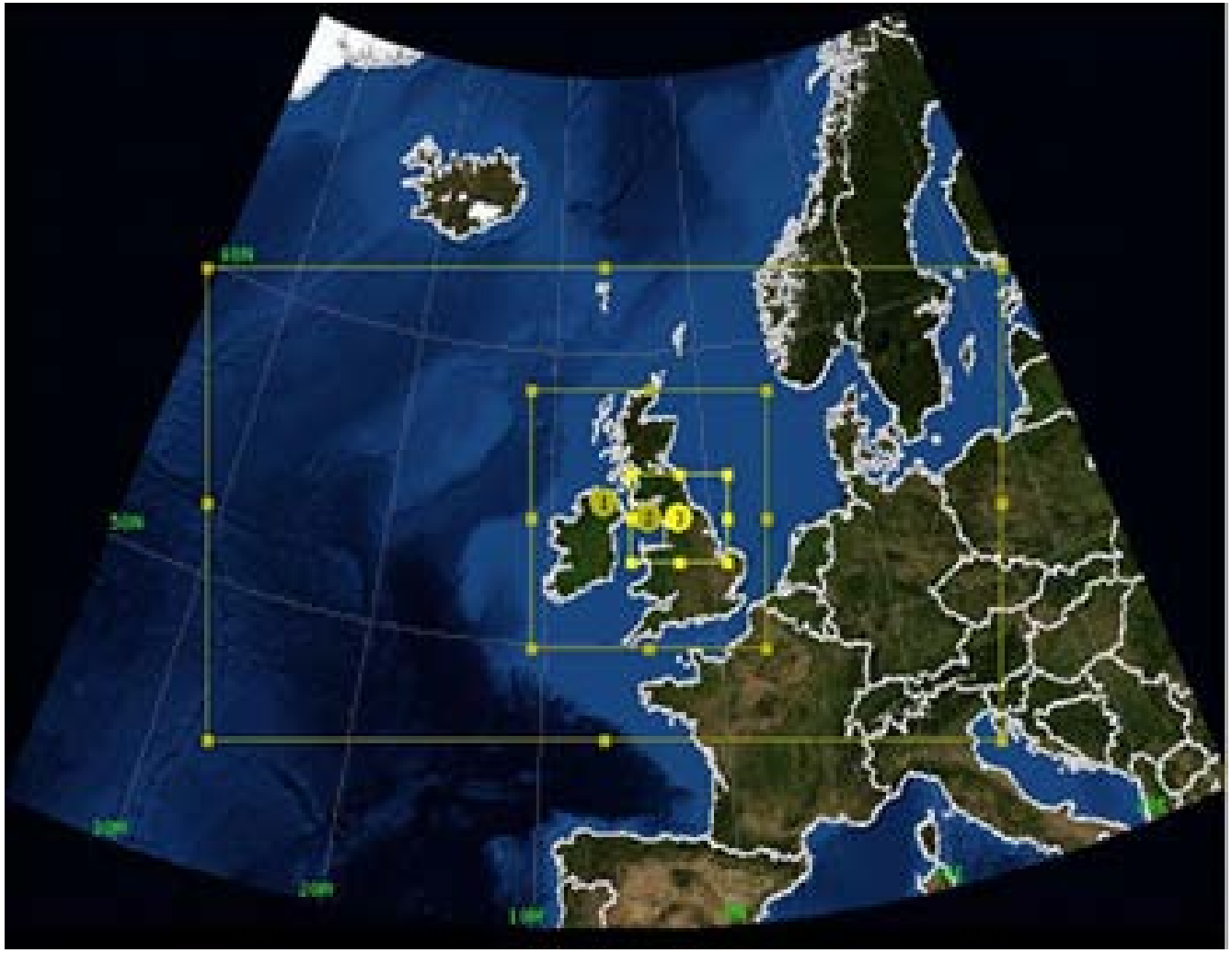
WRF

WRF Weather Model

Craypat timings for Great North Run, nested model of three grids

Great North Run nested grids for regional climate modelling

69k, 103k, 128k points resp.

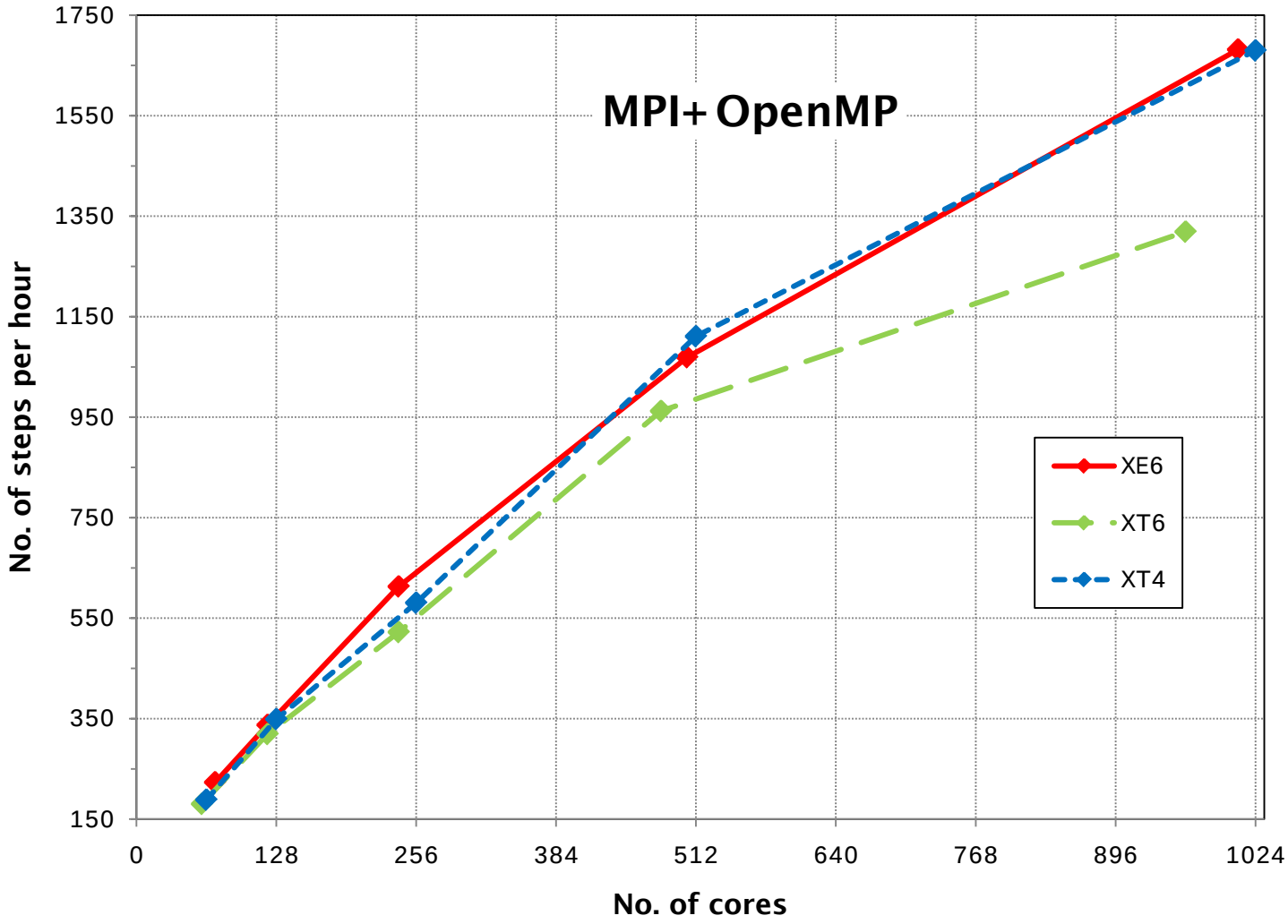


WRF: performance on XT4, XT6 and XE6

Performance for Great North Run, nested model of three grids

Lose out on performance from XT4 to XT6

Regained from XT6 to XE6

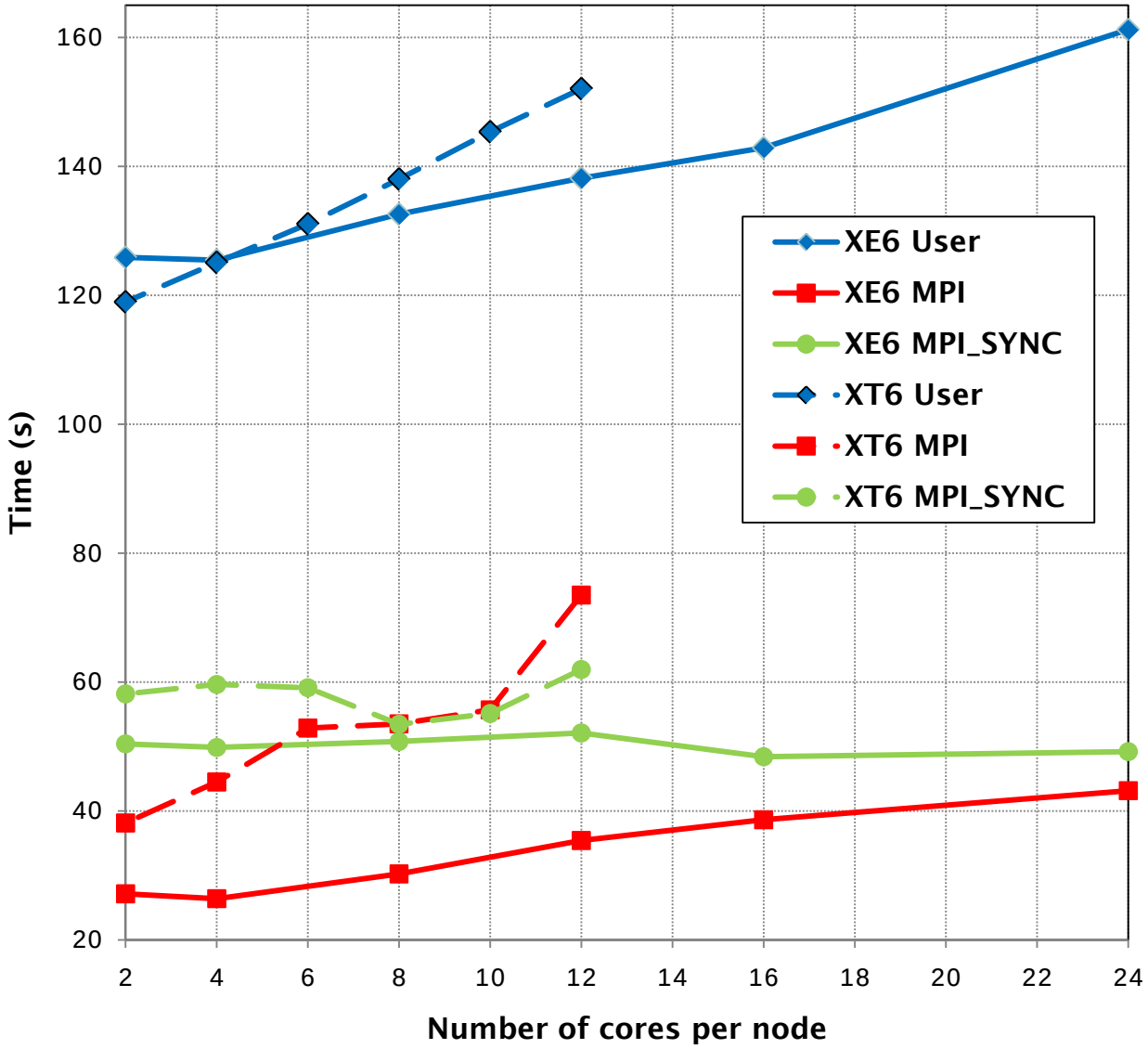


WRF: Craypat timings on XT6 and XE6

Craypat timings for Great North Run, nested model of three grids

Pure MPI runs with varying cores per node; 480 cores total

MPI time shows good reduction



Conclusions

We have looked at a range of applications from different areas of science comparing performance on the Cray XT4, XT6 and XE6 systems

Focuses on change from quad-core to 24-core nodes and from SeaStar to Gemini interconnects

Some apps (POLCOMS, ScaLAPACK, Telemac, WRF) show some/good benefit; others (DL_POLY_4, fd3d, Fluidity-ICOM, PFARM) do not

Need to learn from the good guys and re-engineer the sluggards

Science & Technology Facilities Council

Acknowledgements

Ian Bush, NAG Ltd, for DL_POLY_4 results

This work made use of the facilities of HECToR, the UK's national high-performance computing service, which is provided by UoE HPCx Ltd at the University of Edinburgh, Cray Inc and NAG Ltd, and funded by the Office of Science and Technology through EPSRC's High End Computing Programme.
http://www.epsrc.ac.uk/about/progs/rii/hpc/

This work was performed as part of the project “Computational Science and Engineering Core Support at STFC Daresbury Laboratory 2010-11” funded by EPSRC.

DL_POLY and PFARM are developed through Collaborative Computational Projects (CCPs) which bring together the major UK groups in a given field of computational research to tackle large-scale scientific software development projects, maintenance, distribution, training and user support. http://www.ccp.ac.uk/