# High Performance Computing : Code_Saturne in the PRACE project


Andy SUNDERLAND

Charles MOULINEC

**STFC Daresbury Laboratory, UK**


*Code_Saturne User Meeting*

*Chatou – 1st-2nd Dec 2008*

# STFC Daresbury Laboratory

HPC service provider to the UK academic community for > 25 yrs

- research, development & support centre for leading edge academic engineering and physical science simulation codes
- STFC-DL / EPCC continue to run National UK HPC service


Science & Technology Facilities Council

## The Hartree Centre at STFC Daresbury Laboratory

Strategic science themes incl. energy, biomedicine, environment, functional materials

€60M from Large Facilities Capital Fund to build and equip:

10,000 sq ft machine room

10 MW power

€12M systems / two year cycle


The Hartree Centre will be a new kind of Computational Sciences institute for the UK that will:

- stimulate a step change in modelling capabilities for strategic science themes – Grand challenge projects
- multi-disciplinary, multi-scale, effective and efficient simulation
- **have at its heart the collaborative development, support and exploitation of scientific applications software**

Science & Technology Facilities Council

# *Contents*

- **PRACE**
- **Blue Gene & Green Top 500**
- **Porting on Different Machines (Code_Saturne 1.3.2)**
- **Current Performance (Code_Saturne 1.3.2)**
- **Conclusions – Future Work**

# What is the relevance of PRACE to this meeting ?

# What is the relevance of PRACE to this meeting ?

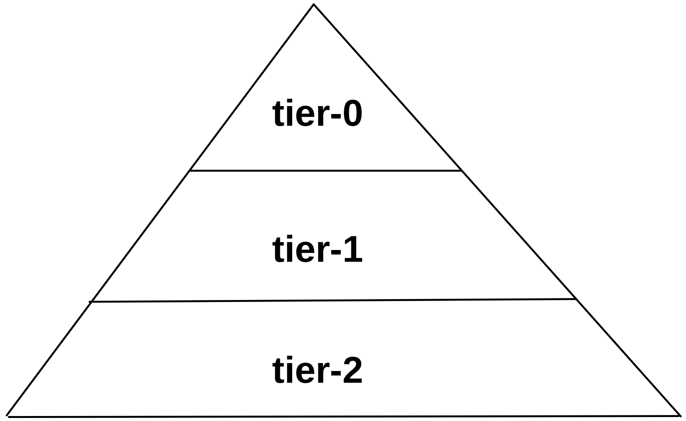
**Code_Saturne** has been chosen as one of the principal applications benchmarks for the Partnership for Advanced Computing in Europe (PRACE) project

- The CFD benchmark for the project

Science & Technology Facilities Council

# The ESFRI* Vision for a European HPC service

- European HPC-facilities at the top of an HPC provisioning pyramid
  - Tier-0: 3-5 European facilities
  - Tier-1: National facilities
  - Tier-2: Regional/University Centres




- Creation of a European HPC ecosystem involving all stakeholders
  - HPC service providers on all tiers
  - Grid Infrastructures
  - **Scientific** and **industrial** user communities
  - The European HPC hard- and software industry

**PRACE – The Partnership for Advanced Computing in Europe**
**An Initiative created to implement the ESFRI vision of a European HPC service**

*European Strategy Forum on Research Infrastructures

Science & Technology Facilities Council

# PRACE – Project Facts

- Objectives of the PRACE Project:
  - Prepare the contracts to establish the PRACE permanent Research Infrastructure as a single Legal Entity in 2010 including governance, funding, procurement, and usage strategies.
  - Perform the technical work to prepare operation of the Tier-0 systems in 2009/2010 including deployment and **benchmarking** of prototypes for Petaflop/s systems and **porting, optimising, peta-scaling of applications**
- Project facts:
  - Partners: 16 Legal Entities from 14 countries
  - Project duration: January 2008 – December 2009
  - Project budget: 20 M € ,    EC funding: 10 M €

CAPACITIES

e-infrastructure

Science & Technology Facilities Council

# PRACE – Project Consortium

Science & Technology Facilities Council

| Participant no. | Participant organisation name | Part. short name | Country |
|---|---|---|---|
| 1 (Coordinator) | Forschungszentrum Juelich GmbH | FZJ | Germany |
| 2 | Universität Stuttgart - HLRS | USTUTT-HLRS | Germany |
| 3 | Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften | BADW-LRZ | Germany |
| **4** | **Grand Equipement national pour le Calcul Intensif** | **GENCI** | **France** |
| **5** | **Engineering and Physical Sciences Research Council (EPCC, STFC DL)** | **EPSRC** | **United Kingdom** |
| 6 | Barcelona Supercomputing Center | BSC | Spain |
| 7 | CSC Scientific Computing Ltd. | CSC | Finland |
| 8 | Eidgenössische Technische Hochschule Zürich - CSCS | ETHZ | Switzerland |
| 9 | Netherlands Computing Facilities Foundation | NCF | Netherlands |
| 10 | Institut fuer Graphische und Parallele Datenverarbeitung der Johannes Kepler Universitaet Linz | GUP | Austria |
| 11 | Swedish National Infrastructure for Computing | SNIC | Sweden |
| 12 | CINECA Consorzio Interuniversitario | CINECA | Italy |
| 13 | Poznan Supercomputing and Networking Center | PSNC | Poland |
| 14 | UNINETT Sigma AS | SIGMA | Norway |
| 15 | Greek Research and Technology Network | GRNET | Greece |
| 16 | Universidade de Coimbra | UC-LCA | Portugal |

Science & Technology Facilities Council

# The next tasks: … growing into a persistent Research Infrastructure

- Identify architectures and vendors capable of delivering Petaflop/s systems by 2009/2010
- Install prototypes at partner sites to verify viability
- Define consistent operation models and evaluate management software
- Capture application requirements and create a benchmark suite
- Port, optimize and scale selected applications
- Define an open, permanent procurement process
- Define and implement a strategy for continuous HPC technology evaluation and system evolution within the RI
- Foster the development of components for future multi-petascale systems in cooperation with European and international HPC industry
- ***Start a process of continuous development and cyclic procurement of technology, software and systems for the PRACE Research Infrastructure***

# Prace Work Packages

- WP1 - Management of the Contract
- WP2 - Organisational Concept of the Research Infrastructure
- WP3 - Dissemination, Outreach and Training
- WP4 - Distributed System Management
- WP5 - Deployment of Prototype Systems
- WP6 - Software Enabling of Petaflop/s systems
- WP7 - Petaflop/s Systems for 2009/2010
- WP8 - Future Petaflop/s Computer Technologies beyond 2010

# Prace Work Package 6

- WP6 – Software for petaflop/s systems
- Largest workpackage in PRACE
- All 16 partners involved
- 280 person-months of effort
- Up to 46 person-months of effort from UK

- Goal:
- *To identify and understand the software libraries, tools, benchmarks and skills required by users to ensure that their application can use a Petaflop/s system productively and efficiently*

# Overview of Tasks in WP6

- T6.1: Identification & categorisation of applications
  - Identification of applications & initial benchmark suite
- T6.2: Application requirements capture
  - Analysis of identified applications
- T6.3: Performance analysis & benchmark selection
  - Selection of tools and benchmarks
- T6.4: Petascaling (up to 6 pms from DL)
  - Petascaling of algorithms
- T6.5: Optimisation (up to 6 pms from DL)
  - Establishment of best-practice in optimisation
- T6.6: Software libraries & programming models
  - Analysis of libraries and programming models

COMPLETED

FUTURE

# Task 6.1: Applications

- Purpose is to understand what applications will be brought to any PRACE systems
- Having identified and categorised the set of applications an initial representative benchmark suite is proposed
  - Attempt to span a broad field of research and the ‘Seven Dwarfs’ computational methods
  - Computational engineering/CFD code choice between **Trio_u, Open_Foam & Code_Saturne**
- Supplies results to T6.2 and WP5
- D6.1: list of representative applications (M6)

# Code Coverage (example)

| Area / Dwarf | Dense linear algebra | Spectral methods | Structured grids | Sparse linear algebra | Particle methods | Unstructured grids | Map reduce methods |
|---|---|---|---|---|---|---|---|
| Astronomy and Cosmology | 0 | 0.62 | 4.91 | 3.59 | 5.98 | 2.99 | 0 |
| Computational Chemistry | 15.35 | 26.09 | 1.80 | 3.45 | 7.49 | 0.53 | 12.98 |
| Computational Engineering | 0 | 0 | 0.53 | 0.53 | 0 | 0.53 | 2.8 |
| Computational Fluid Dynamics | 0 | 1.70 | 7.37 | 3.05 | 0.32 | 3.00 | 0 |
| Condensed Matter Physics | 9.10 | 15.07 | 1.62 | 0.73 | 1.76 | 0.28 | 5.70 |
| Earth and Climate Science | 0 | 2.03 | 5.83 | 1.33 | 0 | 0.26 | 0 |
| Life Science | 0 | 4.72 | 0.94 | 0.13 | 0.94 | 0.28 | 3.46 |
| Particle Physics | 12.50 | 0 | 4.59 | 0.92 | 0.10 | 0 | 89.27 |
| Plasma Physics | 0 | 0 | 1.33 | 1.33 | 3.55 | 0.42 | 0.63 |
| Other | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Science & Technology Facilities Council

Final List of Principal Application Benchmark Codes

| Application | Field | BCO |
|---|---|---|
| NAMD | Molecular Dynamics | EPCC - UK |
| CPMD | Material Science | BSC - SPAIN |
| VASP | Computational Chemistry | BSC - SPAIN |
| QCD | Particle Physics | FZJ - GERMANY |
| GADGET | Cosmology | LRZ - GERMANY |
| CODE_SATURNE | CFD | DL - UK |
| TORB | Plasma & Nuclear Physics | BSC - SPAIN |
| NEMO | Oceanography | SARA - NL |
| ECHAM5 | Atmospheric Physics | CSCS - CH |

Science & Technology Facilities Council

# Benchmark Code Owner

- One Benchmark Code Owner (BCO) all the way through tasks 6.3 to 6.5
  - Clear single interface to the application owner/developers
  - Responsible for creating the benchmark code from the application (input set, reference output, etc.)
  - Responsible for porting, petascaling and optimization
  - Recruits other people/resources to work on specific parts of the benchmark code
  - Reports back to 6.3/6.4/6.5 task leaders

**STFC Daresbury is BCO for Code_Saturne**

Science & Technology Facilities Council

# Task 6.3: Benchmarks

- Task focuses on
  - Identifying performance analysis tools
  - Preparing a representative benchmark suite
    - Initial code preparation
    - Determination of suitable datasets
    - Porting to target hardware platforms
- Relies on work of T6.4 and T6.5 to prepare optimised codes for benchmarking
- WP5 is consumer of the benchmarks

- D6.3.1: Report on available performance analysis & benchmark tools, representative benchmark suite(M11)
- D6.3.2: Final benchmark suite (M24)

# Tasks 6.4 & 6.5 : Petascaling & Optimisation

- Understanding how best to petascale & optimise applications identified in T6.1
- Provides petascaled, optimised benchmarks to T6.3
- Results used by WP5 via T6.3

- D6.4: Report on approaches to petascaling (M22)
- D6.5: Report on porting and optimisation of applications (M22)
- Petascaled, optimised codes for benchmark suite

Science & Technology Facilities Council

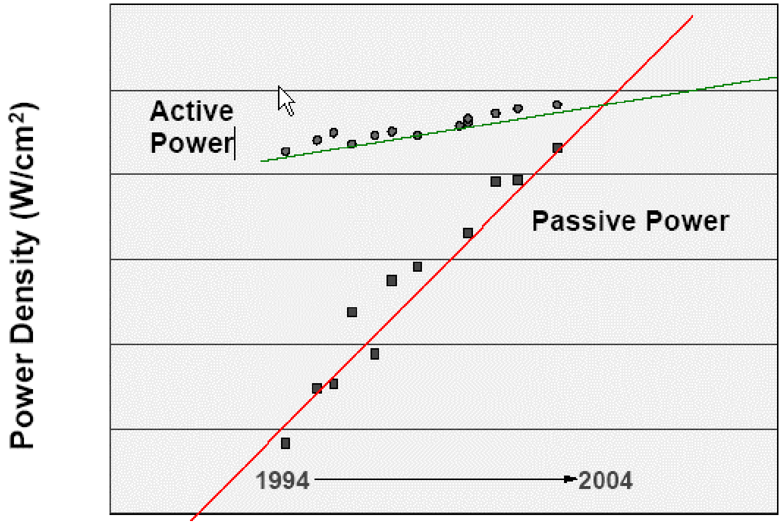
# Prace partners for WP6 Code_Saturne benchmarking, petascaling & optimization)

| Participant Role | Participant organisation name | Part. short name | Country |
|---|---|---|---|
| WP 6.3 Task Leader | Netherlands Computing Facilities Foundation | NCF | Netherlands |
| WP 6.4 Task Leader | Barcelona Supercomputing Centre | BSC | Spain |
| WP 6.5 Task Leader | CSC Scientific Computing Ltd | CSC | Finland |
| Code_Saturne Benchmark Code Owner (BCO) | Engineering and Physical Sciences Research Council (EPCC, STFC DL) | EPSRC | United Kingdom |
| Code_Saturne Contributor (Vector) | Universität Stuttgart - HLRS | USTUTT-HLRS | Germany |
| Code_Saturne Contributor (Cell) | Barcelona Supercomputing Center | BSC | Spain |

# IBM Blue Gene & Green Top 500

# The Power Wall

- Power requirement comprises
  - Active power
  - Passive power (gate leakage)
- Power densities can not increase indefinitely
- Active cooling is usually required
  - Cooling itself consumes power




| | | | |
|---|---|---|---|
| HPCx (2002) | 10 Tflop/s peak | 1 MW | |
| Red Storm | 20 | 2 MW | |
| Earth Simulator | 40 | 7 MW | €6M/year |
| HECToR (2006) | ~50 | ~ 10 MW | €10M/year |

Science & Technology Facilities Council

Feb 2008

| Green500 Rank | MFLOPS/W | Site* | Computer* | Total Power (kW) | TOP500 Rank* |
|---|---|---|---|---|---|
| 1 | 357.23 | Science and Technology Facilities Council - Daresbury Laboratory | Blue Gene/P Solution | 31.10 | 121 |
| 2 | 352.25 | Max-Planck-Gesellschaft MPI/IPP | Blue Gene/P Solution | 62.20 | 40 |
| 3 | 346.95 | IBM - Rochester | Blue Gene/P Solution | 124.40 | 24 |
| 4 | 336.21 | Forschungszentrum Juelich (FZJ) | Blue Gene/P Solution | 497.60 | 2 |
| 5 | 310.93 | Oak Ridge National Laboratory | Blue Gene/P Solution | 70.47 | 41 |
| 6 | 210.56 | Harvard University | eServer Blue Gene Solution | 44.80 | 170 |
| 7 | 210.56 | High Energy Accelerator Research Organization /KEK | eServer Blue Gene Solution | 44.80 | 171 |
| 8 | 210.56 | IBM - Almaden Research Center | eServer Blue Gene Solution | 44.80 | 172 |
| 9 | 210.56 | IBM Research | eServer Blue Gene Solution | 44.80 | 173 |
| 10 | 210.56 | IBM Thomas J. Watson Research Center | eServer Blue Gene Solution | 44.80 | 174 |
| 11 | 210.56 | Renaissance Computing Institute (RENCI) | eServer Blue Gene Solution | 44.80 | 175 |
| 12 | 210.56 | University of Canterbury | eServer Blue Gene Solution | 44.80 | 176 |
| 13 | 208.31 | Forschungszentrum Juelich (FZJ) | eServer Blue Gene Solution | 179.20 | 28 |
| 14 | 208.31 | Computational Biology Research Center, AIST | eServer Blue Gene Solution | 89.60 | 52 |
| 15 | 208.31 | EDF R&D | eServer Blue Gene Solution | 89.60 | 53 |
| 16 | 208.31 | Ecole Polytechnique Federale de Lausanne | eServer Blue Gene Solution | 89.60 | 54 |

http://www.green500.org

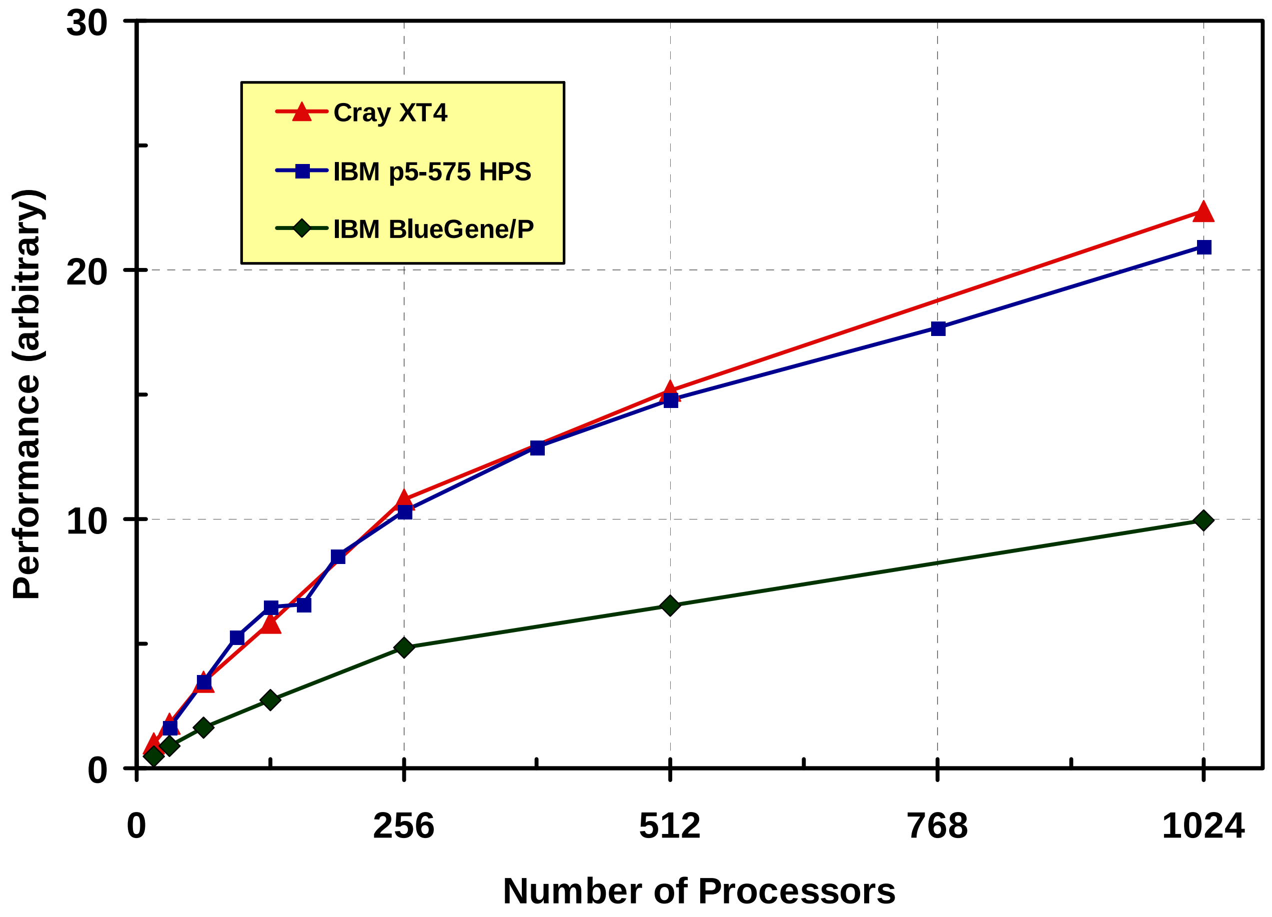
Science & Technology Facilities Council

# Green Top 500 Supercomputer List (Feb 2008)

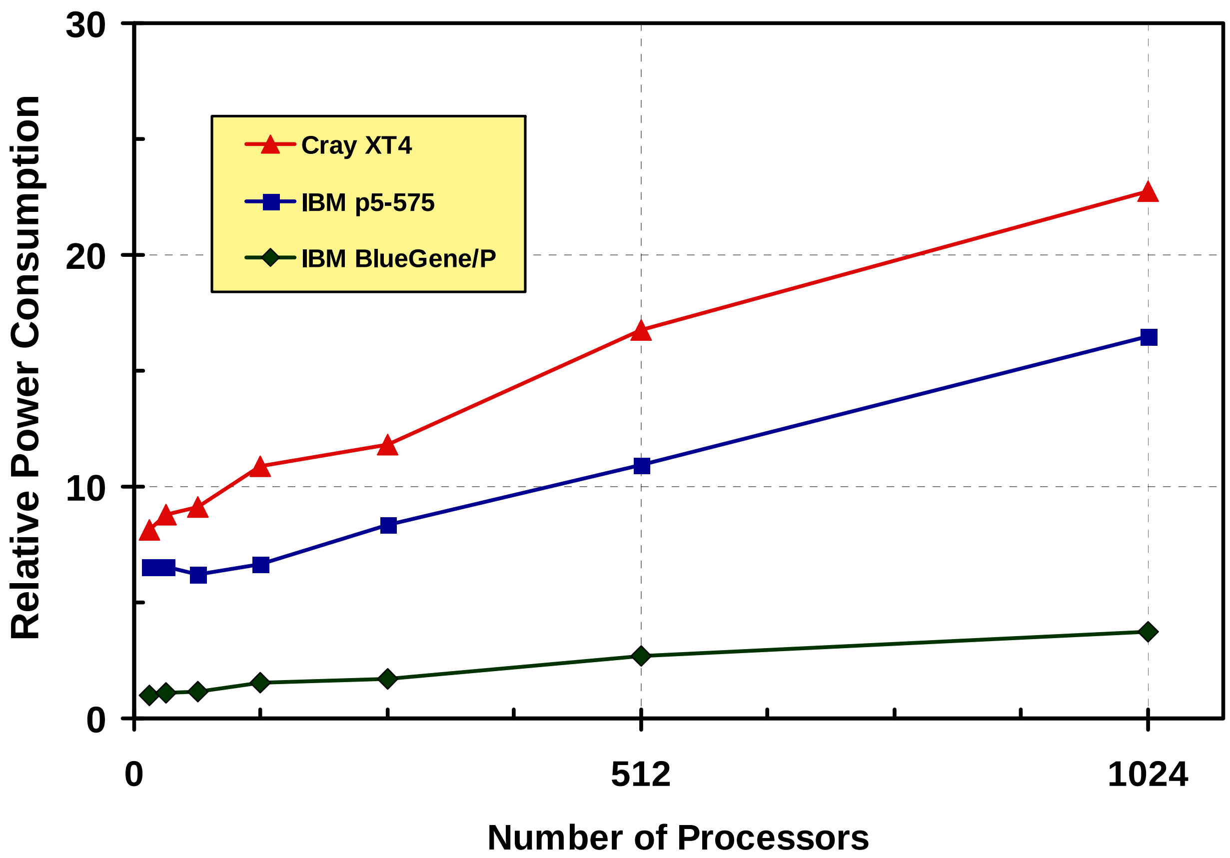
| Machine | Mflops/ watt | Green Top 500 Ranking | Top 500 Ranking |
|---|---|---|---|
| IBM BG/P (STFC) | 357.23 | **1** | 121 |
| IBM BG/L (EDF) | 208 | **15** | 53 |
| IBM p5-575 (HPCx) | 30.96 | **409** | 101 |
| Cray XT4 (HECToR) | 21.01 | **460** | 17 |

http://www.green500.org

## Relative Eigensolver Performance on HPC platforms



# Relative Eigensolver Power Requirements on HPC platforms




**Flops / Second or Flops / Watt ?**

# Porting

## Completed

- Blue Gene / L, P (DL, UK)
- IBM AIX, Power5 (DL, UK)
- Cray XT4 (EPCC, UK)
- IBM AIX, Power6 (SARA, NL)
- Cluster x86 (SARA, NL)
- NEC-SX8 (HLRS, DE)

## Not Completed

- Cell (BSC, SP)

## Optimisation in Progress

Science & Technology Facilities Council

# Machines

## BG/L

*2,048 PowerPC 440 700MHz 5.7 TeraFlops at Peak*

## BG/P

*4,096 PowerPC 450 850MHz 13.92 TeraFlops atPeak*

## IBM AIX (HPCx)

HPCx CAPABILITY COMPUTING

*2,560 POWER5 1.5GHz*

*15.36 TeraFlops at Peak*

## Cray XT4 (HECToR)

HECToR

*11,328 cores (DC opteron) 2.8GHz*

*59 TeraFlops at Peak*

# Current Performance (1.3.2)

**Conditions for the simulation:**

- **Partitioning on a SGI (100Gb RAM)**
- **Copy of input files to the HPC Machine**
- **Runs (if necessary Restart Files)**

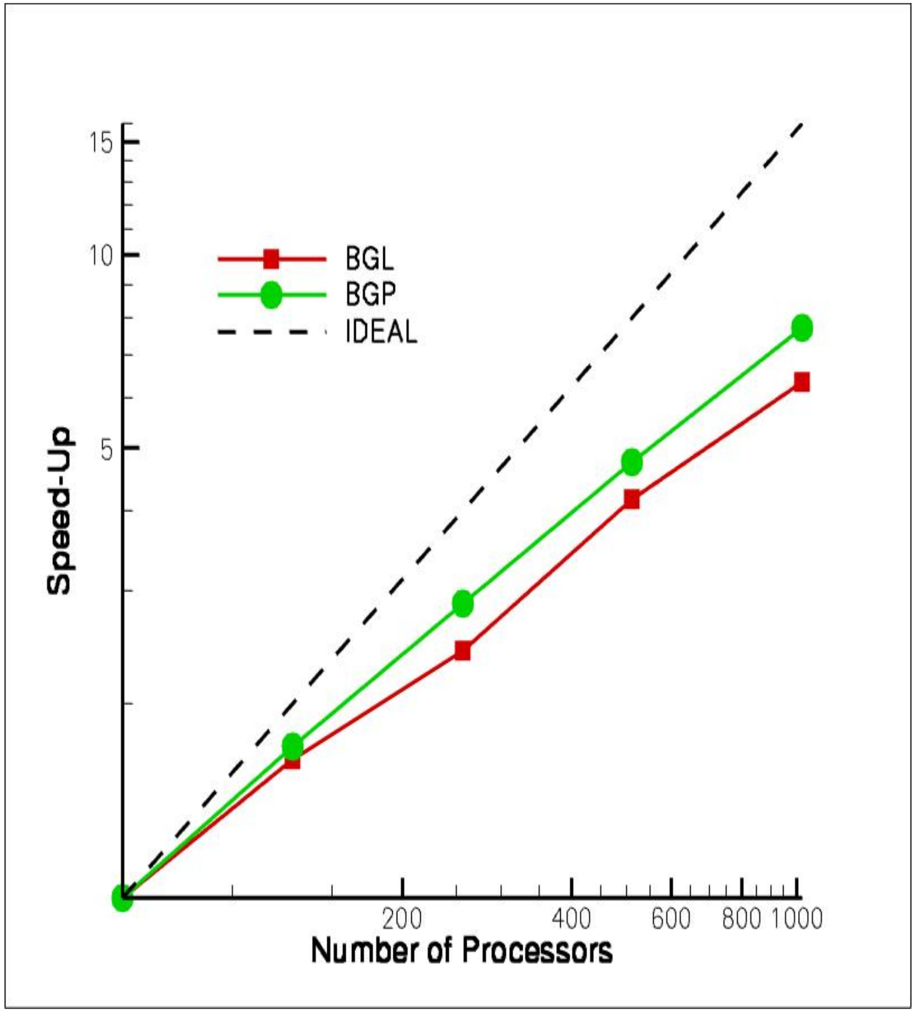
Science & Technology Facilities Council

# Current Performance (1.3.2)

**16M cells on BGL/BGP**

**(LES in a Channel Flow)**

**Calculation Part**

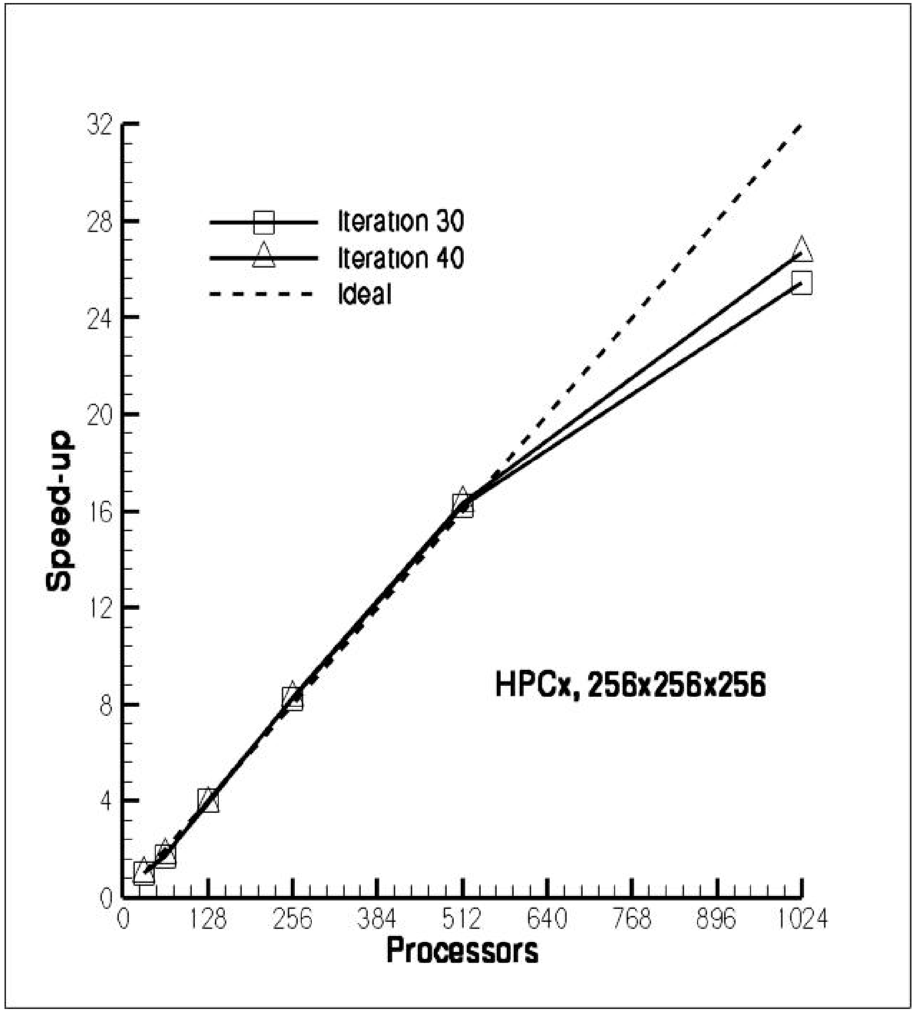


# Current Performance (1.3.2)

***16M cells on HPCx***

***(LES in a Channel Flow)***

***Calculation Part***

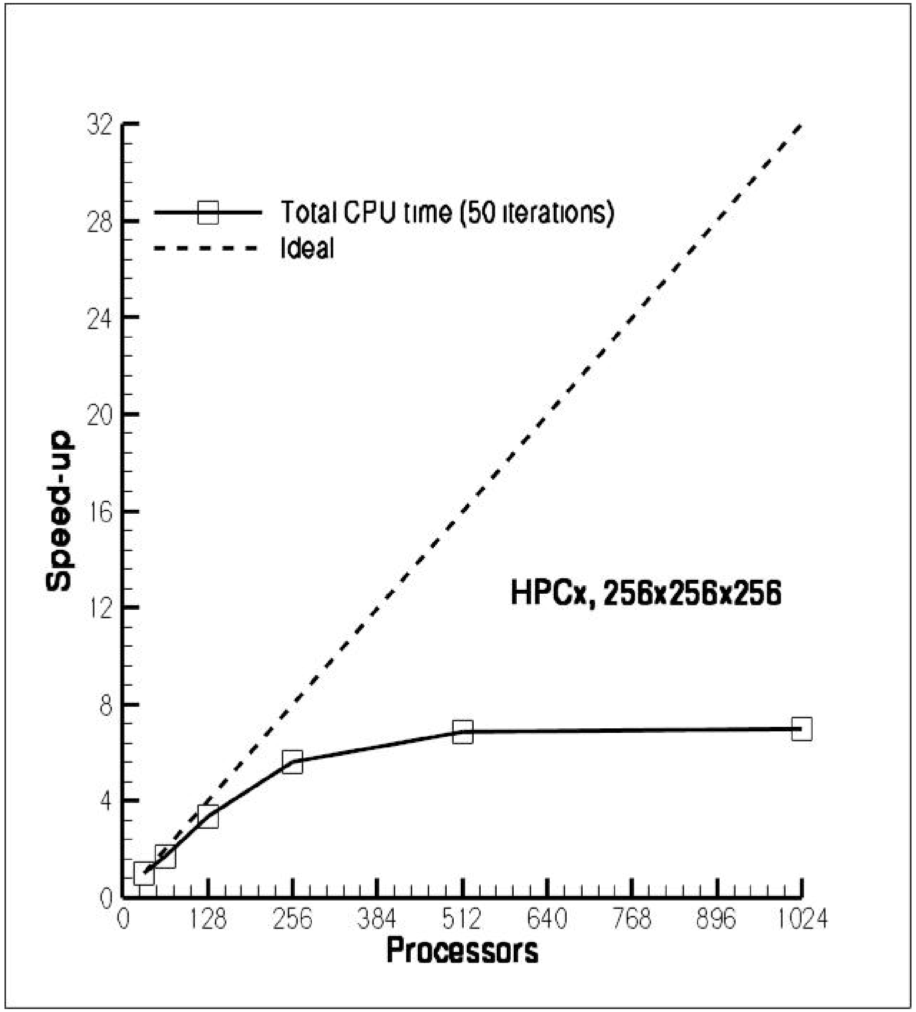


# Current Performance (1.3.2)

***16M cells on HPCx***

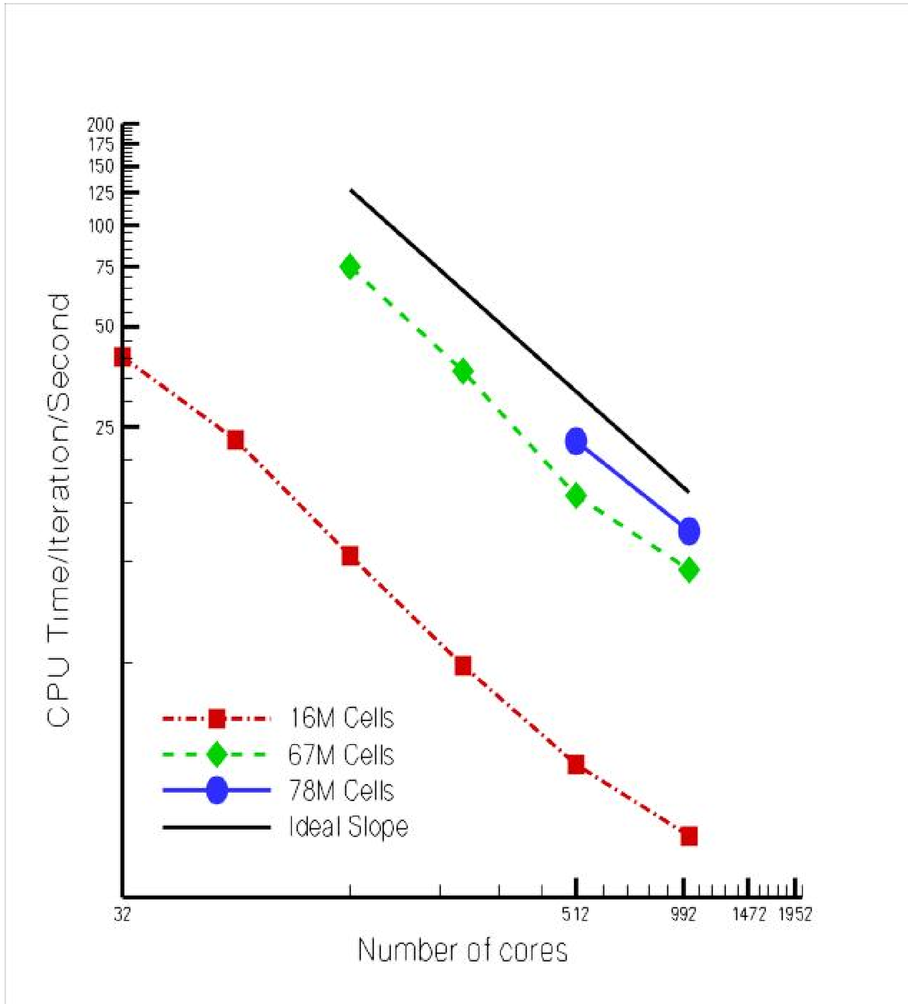
***(LES in a Channel Flow)***

***Overall Simulation***



# Current Performance (1.3.2)

## *78M cells on HPCx (LES in a Channel Flow – Gold Award)*

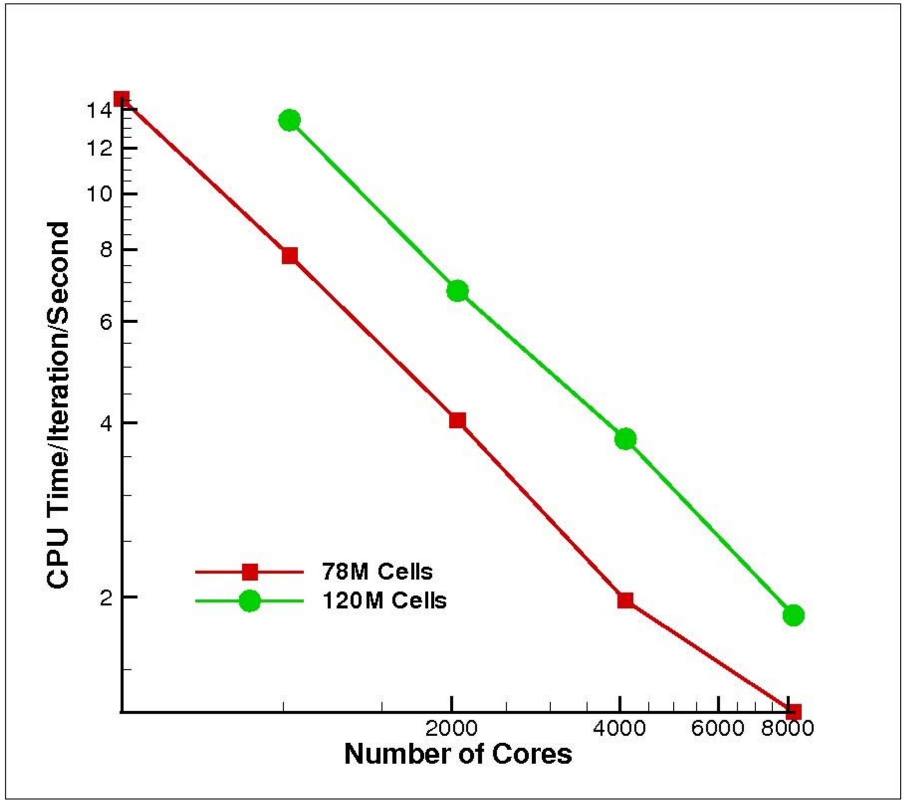



**9.3% of the Peak (96 cores)**

Science & Technology Facilities Council

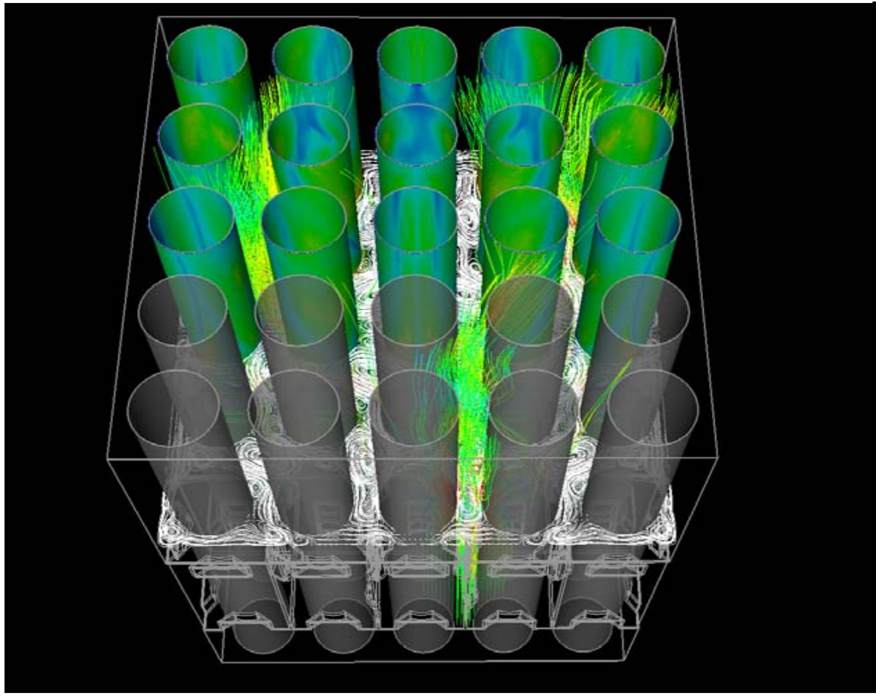
# Current Performance (1.3.2)

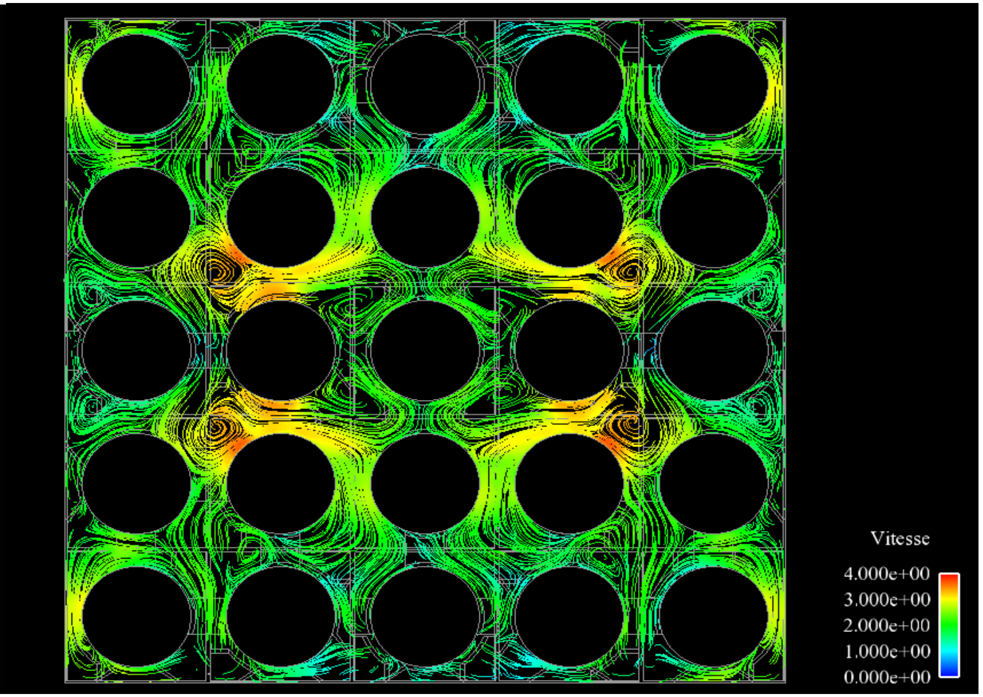
## 78M-120M cells on HECToR (LES in Channel Flow)



# Current Performance (1.3.2)

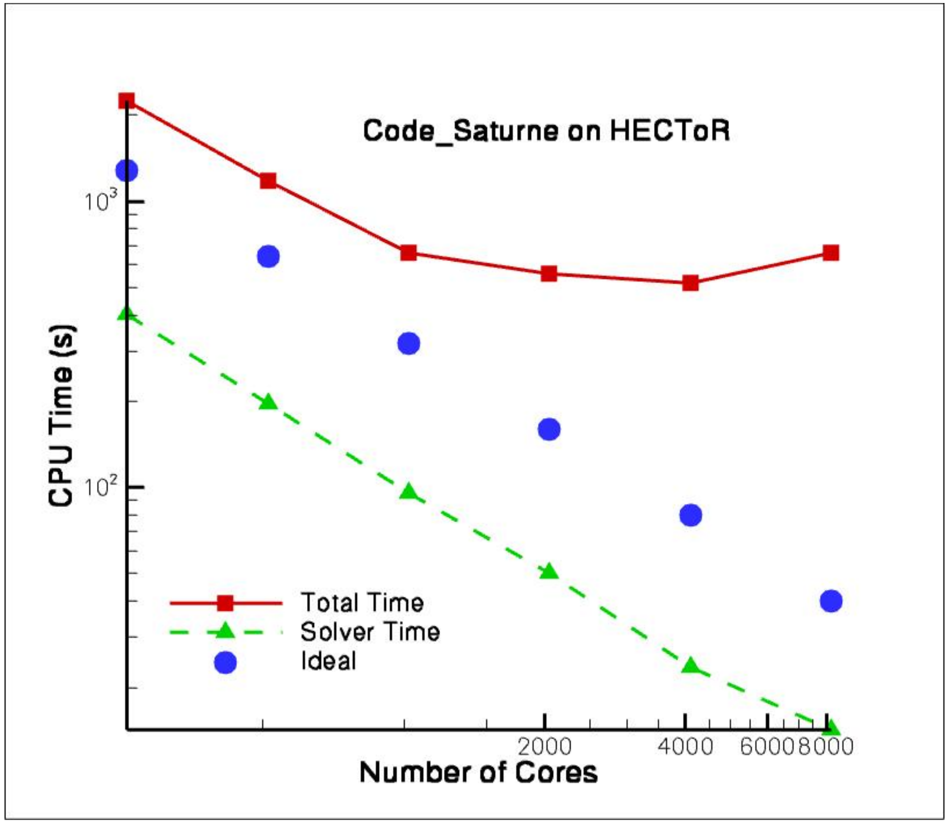
## 100M cells on HECToR (Mixing Grid)





# Current Performance (1.3.2)

## 100M cells on HECToR (Mixing Grid)



# Dissemination

• **HPCx Newsletter**

http://www.hpcx.ac.uk/about/newsletter/HPCxNews11.pdf

•**Frontiers 2008** (STFC yearly letter)

•**Gold Award Medal on HPCx** (Speed-Up of 1.7 from 512 to 1,024 procs and 30% discount on all simulations)

•**Conference**: A.G. Sunderland, M. Ashworth, N. Li, C. Moulinec, Y. Fournier and J.C. Uribe. "Towards Petascale Computing with Parallel CFD Codes". Parallel CFD 2008, Lyon.

# Conclusions – Future Work

## Conclusions

• Code_Saturne in PRACE

•Very good performance for the calculation part

•Overall perfomance not as good, due to (in 1.3.2) I/O's, Serial Partitioning

## Future Work

• Optimisation Cray XT4 (HECToR)

•Vector Machine (NEC and Cray X2)

•NAG project (18 months) Partitioning, I/O

•Porting and Tests on a Cray XT4 (Jaguar, 25,000+ processors)

•Porting and Tests on BG/P Argonne (4 racks)

# Future Issues

- **Solver part: Good performance**

# Future Issues

- **Solver part: Good performance**
- **Input/Output: expected good performance within a few months**

# Future Issues

- **Solver part: Good performance**
- **Input/Output: expected good performance within a few months**
- **Parallel partitioning: expected good performance within a few months**

Science & Technology Facilities Council

# *Future Issues*

- **Solver part: Good performance**
- **Input/Output: expected good performance within a few months**
- **Parallel partitioning: expected good performance within a few months**
- **What about parallel meshing?**