SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

INFRA-2010-2.3.1 – First Implementation Phase of the European High Performance Computing (HPC) service PRACE

PRACE-1IP

PRACE First Implementation Project

Grant Agreement Number: RI-261557

D7.2.1
Interim Report on Collaboration with Communities
Final

Version:	1.0
Author(s):	Giovanni Erbacci, CINECA
Date:	25.06.2011

Project and Deliverable Information Sheet

PRACE Project	**Project Ref. №: RI-261557**	
	Project Title: PRACE First Implementation Project	
	Project Web Site: http://www.prace-project.eu	
	Deliverable ID: D7.2.1	
	Deliverable Nature: <DOC_TYPE: Report / Other>	
	Deliverable Level: PU / PP / RE / CO *	**Contractual Date of Delivery:** 30 / June / 2011
		Actual Date of Delivery: 30 / June / 2011
	EC Project Officer: Bernhard Fabianek	

* - The dissemination level are indicated as follows: **PU** – Public, **PP** – Restricted to other participants (including the Commission Services), **RE** – Restricted to a group specified by the consortium (including the Commission Services). **CO** – Confidential, only for members of the consortium (including the Commission Services).

Document Control Sheet

Document	**Title: Interim Report on Collaboration with Communities**	
	ID: D7.2.1	
	Version: 1.0	**Status:** Draft / Final
	Available at: http://www.prace-project.eu	
	Software Tool: Microsoft Word 2007	
	File(s): D7.2.1.docx	
Authorship	**Written by:**	Giovanni Erbacci, CINECA
	Contributors:	Riccardo Bunino, CINECA John Donners, SARA Andrew Emerson, CINECA Ivan Girotto, ICHEC Joerg Hertzer, HLRS Paul Sherwood, STFC Maria Fancesca Iozzi, UiO
	Reviewed by:	William Sawyer, CSCS Dietmar Erwin, JSC
	Approved by:	MB/TB

Document Status Sheet

Version	**Date**	**Status**	**Comments**
0.1	20/April/2011	Draft	First draft and preliminary content
0.2	27/May/2011	Revised Draft	Integration of progress report on application codes
0.9	9/June/2011	Draft ready for internal review	Revision from WP7
1.0	25/June/2011	Final version	

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, Scientific communities, Application codes, Petascaling,

Copyright notices

Table of Contents

List of Figures

List of Tables

References and Applicable Documents

[1] PRACE Preparatory Phase Deliverable D6.1, "*Final report on Application requirement*", 2008.

[2] PRACE 1IP Deliverable D 7.4.1, "*Applications and user requirements for Tier-0 systems*", February 2011.

[3] HET, The scientific case for European supercomputing Infrastructure, 2007.

[4] PRACE Preparatory Phase Deliverable D6.5, "*Report on porting and optimisation of applications*", 2009.

[5] Ivan Girotto, Yang Yang, ICHEC's GPU research: porting of scientific application on NVIDIA GPU, poster session GTC 2010, http://www.nvidia.com/content/GTC/posters/2010/I15-ICHECs-GPU-Research-Porting-of-Scientific-Application-on-NVIDIA-GPU.pdf

List of Acronyms and Abbreviations

ADF	Amsterdam Density Functional software
AISBL	Association internationale sans but lucratif
BCO	Benchmark Code Owner
BG/P	BlueGene/P computer
BLAS	Basic Linear Algebra Subprograms
BSC	Barcelona Supercomputing Center (Spain)
BSCW	Basic Support for Cooperative Work (software package for collaboration over the Web)
CC	Coupled cluster. A numerical technique for describing many-body systems, used in quantum chemical post-Hartree–Fock ab initio quantum chemistry methods
ccNUMA	cache coherent NUMA
CEA	Commissariat à l'Energie Atomique (represented in PRACE by GENCI, France)
CERFACS	Centre Européen de Recherche et de Formation Avancée en Calcul, France
CERN	Organisation Européenne pour la Recherche Nucléaire
CFD	Computational Fluid Dynamics
CI	Configuration Interaction method. A post-Hartree–Fock linear variational method
CIEMAT	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas
CINECA	Consorzio Interuniversitario, the largest Italian computing centre (Italy)

CMB	Cosmic Microwave Background
CPU	Central Processing Unit
CRPP	Centre de Recherches en Physique des Plasmas in Lausanne
CSC	Finnish IT Centre for Science (Finland)
CSCS	The Swiss National Supercomputing Centre (represented in PRACE by ETHZ, Switzerland)
CSR	Compressed Sparse Row (for a sparse matrix)
CUDA	Compute Unified Device Architecture (NVIDIA)
DARPA	Defense Advanced Research Projects Agency
DBCSR	Distributed Block Compressed Sparse Row library
DECI	DEISA Extreme Computing Initiative.
DEISA	Distributed European infrastructure for Supercomputing Applications. EU project by leading national HPC centres
DEMOCRITOS	DEmocritos MOdeling Centre for Research In aTOmistic Simulations, Trieste (Italy)
DFT	Density Functional Theory
DGEMM	Double precision General Matrix Multiply
DoE	Department of Energy
DP	Double Precision, usually 64-bit floating point numbers
EC	European Community
ECMWF	European Centre for Medium-Range Weather Forecasts
EDF	Électricité de France
EESI	European Exascale Software Initiative
EFDA	European Fusion Development Agreement
EMBL	European Molecular Biology Laboratory
EPCC	Edinburg Parallel Computing Centre (represented in PRACE by EPSRC, United Kingdom)
EPOS	European Plate Observing System. European research infrastructure in solid Earth Science
EPSRC	The Engineering and Physical Sciences Research Council (United Kingdom)
ESM	Earth System Model
ETHZ	Eidgenössische Technische Hochschule Züerich, ETH Zurich (Switzerland)
ESFRI	European Strategy Forum on Research Infrastructures; created roadmap for pan-European Research Infrastructure.
FFT	Fast Fourier Transform
FP	Floating-Point
FZJ	Forschungszentrum Jülich (Germany)
GB	Giga (= 2^{30} ~ 10^{9}) Bytes (= 8 bits), also GByte
Gb/s	Giga (= 10^{9}) bits per second, also Gbit/s
GB/s	Giga (= 10^{9}) Bytes (= 8 bits) per second, also GByte/s
GCS	Gauss Centre for Supercomputing (Germany)
GENCI	Grand Equipement National de Calcul Intensif (France)
GFlop/s	Giga (= 10^{9}) Floating point operations (usually in 64-bit, i.e. DP) per second, also GF/s
GHz	Giga (= 10^{9}) Hertz, frequency =10^{9} periods or clock cycles per second
GIPAW	Gauge-Including Projector-Augmented Wave Method. A first principles theory of solid state NMR
GNU	GNU’s not Unix, a free OS

GPAW	Grid-based Projector-Augmented Wave method. GPAW is a DFT Python code based on the PAW method.
GPGPU	General Purpose GPU
GPL	GNU General Public License
GPU	Graphic Processing Unit
GPW	Gaussian and Plane Wave approach
GRIB	GRIdded Binary, a concise data format commonly used in meteorology
GROMACS	GROningen MAchine for Chemical Simulations. A molecular dynamics simulation package originally developed in the University of Groningen
GWU	George Washington University, Washington, D.C. (USA)
HECToR	High-End Computing Terascale Resource (UK's national supercomputing service)
HET	High Performance Computing in Europe Taskforce. Taskforce by representatives from European HPC community to shape the European HPC Research Infrastructure. Produced the scientific case and valuable groundwork for the PRACE project
HF	Hartree-Fock method
HiPER	European High Power laser Energy Research facility
HLRS	High Performance Computing Center Stuttgart, Germany
HPC	High Performance Computing; Computing at a high performance level at any given time; often used synonym with Supercomputing
HPL	High Performance LINPACK
IB	InfiniBand
IBM	International Business Machines Corporation
ICE	Internal Combustion Engine Group (at Politecnico di Milano, Italy)
ICHEC	Irish Centre for High-End Computing
IC3	Catalonian institute for climate sciences, Spain
IDRIS	Institut du Développement et des Ressources en Informatique Scientifique (represented in PRACE by GENCI, France)
IFS	Integrated Forecast System. An operational global meteorological forecasting model developed by ECMWF
I/O	Input/Output
IOIPSL	I/O institute Pierre Simon Laplace library. The IPSL I/O library
IPB	Institute of Physics Belgrade
IPCC	Intergovernmental Panel on Climate Change
IPM	Integrated Performance Monitoring (a portable profiling infrastructure for parallel codes)
IPP	Max-Planck-Institut für Plasmaphysik
IPSL	Institut Pierre Simon Laplace, France.
IS-ENES	InfraStructure for the European Network for the Earth System Modelling; an FP7-Project
ISV	Independent Software Vendor
ITER	International Thermonuclear Experimental Reactor
JSC	Jülich Supercomputing Centre (FZJ, Germany)
KNMI	Royal Dutch Meteorological Institute, The Netherlands.
KTH	Kungliga Tekniska Högskolan (represented in PRACE by SNIC, Sweden)
LAMMPS	Large-scale Atomic/Molecular Massively Parallel Simulator. A molecular dynamics code developed at Sandia National Laboratories
LAPACK	Linear Algebra PACKage
LES	Large eddy simulation

LINPACK	Software library for Linear Algebra
LiU	Linkoping University, Sweden
LOBPCG	Local Optimal Block Preconditioned Conjugate Gradient algorithm
LQCD	Lattice QCD
LR-TDDFT	Linear Response Time Dependent DFT formalism
LRZ	Leibniz Supercomputing Centre (Garching, Germany)
MAPPER	Multiscale APPlications on European e-infRastructures. FP7 EU Projetc
MCSCF	Multi-Configurational Self-Consistent Field method
MD	Molecular Dynamics
MFlop/s	Mega (= 10^6) Floating point operations (usually in 64-bit, i.e. DP) per second, also MF/s
MHz	Mega (= 10^6) Hertz, frequency =10^6 periods or clock cycles per second
MKL	Math Kernel Library (Intel)
MPI	Message Passing Interface
NAMD	Not (just) Another Molecular Dynamics program. A parallel molecular dynamics code for large biomolecular systems
NCF	Netherlands Computing Facilities (Netherlands)
NEMO	Nucleus for European Modelling of the Ocean: A state-of-the-art modelling framework for oceanographic research,
NMR	Nuclear Magnetic Resonance
NTNU	Norwegian University of Science and Technology, Norway
NUI	National University of Ireland
NUMA	Non-Uniform Memory Access or Architecture
OASIS	Ocean Atmosphere Sea Ice Soil. A tool for coupling the atmosphere and ocean components of a meteo-climatology model
OpenCL	Open Computing Language
OpenFOAM	Open Field Operation And Manipulation. Open source CFD software
OpenMP	Open Multi-Processing
OS	Operating System
PABS	PRACE Application Benchmark Suite
PAW	Projector Augmented Wave method
PME	Particle-Mesh Ewald method
PNNL	Pacific Northwest National Laboratory in Richland Washington
PP	Particle Particle force calculation
PRACE	Partnership for Advanced Computing in Europe; Project Acronym
PW	plane waves
PWR	Pressurised water reactor
PWscf	Plane-Wave Self-Consistent Field. A code for electronic structure calculations.
QCD	Quantum Chromodynamics
QDR	Quad Data Rate
QE	Quantum Espresso
QM	Quantum Methods
QM/MM	Quantum Mechanics/Molecular Mechanics
QR	QR method or algorithm: a factorisation of a matrix into a unitary and upper triangular matrix
RMDIIS	Residual Minimization with Direct Inversion In Subspace algorithm
RT-TDDFT	Real Time - Time Dependent DFT formalism
SARA	Stichting Academisch Rekencentrum Amsterdam (Netherlands)
ScaLAPACK	Scalable LAPACK library

SEM	Spectral-Element Method
SGEMM	Single precision General Matrix Multiply, subroutine in the BLAS
SGI	Silicon Graphics, Inc.
SIESTA	Spanish Initiative for Electronic Simulations with Thousands of Atoms. A code for performing electronic structure calculations and ab initio molecular dynamics simulations
SISSA	International School for Advanced Studies, Trieste (Italy)
SMHI	Swedish Meteorological and Hydrological Institute, Sweden
SNIC	Swedish National Infrastructure for Computing (Sweden)
SP	Single Precision, usually 32-bit floating point numbers
SPC	Simple Point Charge model. A common simulation model for water
STFC	Science and Technology Facilities Council (represented in PRACE by EPSRC, United Kingdom)
SVN	Subversion system. A software versioning and a revision control system
TFlop/s	Tera (= 10^{12}) Floating-point operations (usually in 64-bit, i.e. DP) per second, also TF/s
Tier-0	Denotes the apex of a conceptual pyramid of HPC systems. In this context the Supercomputing Research Infrastructure would host the Tier-0 systems; national or topical HPC centres would constitute Tier-1
Tier-1	Denotes the national or topical HPC centres in the pyramid of HPC systems
UT	University of Technology
UKTC	UK Turbolence Consortium
VASP	Vienna Ab-Initio Simulation Package
VERCE	Virtual Earthquake and seismology Research Community in Europe e-science. FP7 EU project
VKI	Von Karman Institute for Fluid Dynamics
VMEC	Variational Moments Equilibrium Code
Wikki	Wikki Ltd (UK) / Wikki GmbH (Germany)
2D	2-Dimensional
3D	3-Dimensional

Executive Summary

Task 7.2 "*Application Enabling with communities* has the objective of developing relationships with important European applications communities and provide petascaling expertise to ensure that their key applications codes can effectively exploit Tier-0 systems. Deliverable D7.2.1 represents the interim report documenting the collaboration with scientific communities in the first year of activity of PRACE 1IP.

The document presents the methodology adopted to identify the scientific communities for cooperation and to select the application codes of interest to these communities. In addition, we include progress reports describing the activity which has been performed for each application.

So far, eleven key application codes which are of value to the main scientific communities at European level have been selected for petascaling in Task 7.2. The applications are: GROMACS and DL-POLY for Molecular Dynamics; Quantum Espresso and CP2K, for Material Science; GPAW for Nanoscience; DALTON for Computational Chemistry; EUTERPE for Plasma Physics; the EC-Earth 3 suite for Meteo-climatology, SPECFEM3D for Earth Sciences (earthquakes) and OpenFOAM and Code_Saturne for Engineering and CFD. For these applications, a long term collaboration is envisaged (in terms of 8-10 PM from Task 7.2, in some cases with the support of Task 7.5 and Task 7.6). For these eleven applications a total effort of around 100 PM from Task 7.2 has been allocated. It should be emphasized that the petascaling work involves also the owners of the application codes, or the scientific communities interested in the codes. They are collaborating actively with the PRACE staff, providing their own effort in terms of PM and expertise.

For five of these applications (GROMACS, Quantum Espresso, GPAW, EC-Earth 3 and OpenFOAM) the enabling work started in December 2010 and some preliminary results are presented. For the remaining six applications the work started in May 2011 after the selection process and the agreements with the scientific communities on the work plan to follow. The activity with these five application codes is scheduled to conclude around December 2011, so that the petascaled codes will allow the related scientific communities to apply to the subsequent regular access calls for effective use of the PRACE Tier-0 systems. For the other six application codes the collaboration will continue till spring 2012.

It is important to emphasise here the ongoing cooperation with the IS-ENES community (the European Network for Earth System Modelling). The long term objective with this community is to enable the six European climate models of interest for the community to run on the PRACE Tier-0 systems. The first step was to define a strong cooperation with the researchers of this community in December 2010 to petascale the EC-Earth 3 suite of codes which includes three main applications that are part of the six European climate models: Oasis, NEMO and IFS. The activity is making good progress and consolidates the strong cooperation between PRACE and IS-ENES.

Collaboration with other structured communities will be established and in the next few months agreement for petascaling some other application codes of key interest for these communities will be reached in order to exploit fully the HPC resources on the PRACE Tier-0 systems. For this future work more than 40 PM are still available from Task 7.2.

1 Introduction

Petascaling applications to enhance their scalability plays a key role in ensuring effective exploitation of the PRACE Tier-0 systems. In this context the collaboration with scientific communities is fundamental. Enhancing code scalability on Tier-0 systems is not trivial and generally requires not only a very careful analysis of the code and a knowledge of the Tier-0 architecture involved but also a thorough understanding of the scientific application and the numerical model and the parallelisation methods.

Task 7.2 "Applications enabling with Communities" has the ambitious goal of identifying opportunities to enable applications through engagement with selected scientific communities, industrial users and specific application projects. The work has two objectives:

1) Establish a long-term collaboration with scientific communities interested in computational sciences to facilitate their exploitation of Tier-0 systems;
2) Enabling new challenging supercomputing applications that can be put at the disposal of the whole scientific community. This allows not just a single researcher but a wider community to realise complex, demanding and innovative simulations on Tier-0 that would not be possible without this collaboration.

The report is organised in six sections. The next section provides a general characterisation of the scientific communities and describes the methodology adopted to identify and select them. Section 3 describes the methodology for cooperating with independent communities and that for selecting application codes to petascale, which are of interest for these independent communities, mainly groups of users involved in the same research field.

Section 4 discusses in detail the activity with the more structured communities and the plans to develop a long-term collaboration with them.

Section 5 presents the progress report for the selected applications, showing both the preliminary results and the collaboration with the relevant community and/or the owners of the code. Finally, some conclusions of the work and perspectives are discussed in Section 6. There are also two annexes, one, containing the form used to identify communities and codes of interest, and the second containing an abstract of the application codes selected.

Apart from WP7, the major audience for this report can be envisaged to be WP4, to contribute to further integrate the HPC-Ecosystem; WP5, to enhance cooperation with industrial communities; and WP6, to get feedback on the operability and service provision on Tier-0 systems. However, the authors believe that the document should also provide valuable information for the PRACE AISBL at large to provide a better infrastructure for computational research, providing applications optimised and petascaled and creating a more concrete cooperation with the European computational research community.

2 Scientific Communities

A scientific community can be defined as a group of scientists working in a particular scientific domain, with relationships and interactions among them, sharing opinions and using the same scientific methodologies to advance science in that specific domain. Scientific communities involved in computational sciences are generally interested in adopting similar computational approaches and methodologies, and using high end HPC resources to advance science through simulation methods. These communities, in general, adopt specific categories of application codes and similar mathematical models.

These communities can be broadly divided in two categories: *independent* and *structured* . The independent communities are organised as groups of scientists investigating specific research topics and characterised by using similar application codes and adopting a common investigation methodology. No specific governance rules and research programmes are defined between the groups of scientists involved in a community. Often the developers of the application codes and the related communities do not work closely together, nor do they necessarily coordinate their efforts.

Structured communities on the other hand gather groups of researchers carrying out research in the same scientific area at a European level, with a common governance, structure and specific research objectives. Sometimes these communities also have a coherent body which represents the whole community. In general these communities are structured in terms of organised infrastructures or projects at European level.

These communities often need considerable computational resources to do production runs for their research using applications which have been previously petascaled. This is different from the needs of independent research communities which benefit from cooperation with HPC experts in order to enable and petascale their own common set of codes and methodologies to enhance science in a specific discipline.

A survey carried out by WP6 in the PRACE-PP project [1] showed that while a considerable percentage of user groups and communities use their own application codes, there are programs that are widely used by specific communities. This trend was confirmed also by the survey on the Applications and user requirements for Tier-0 systems performed by Task 7.4 in PRACE 1IP [2].

To achieve an effective collaboration with the user communities we decided to organise the Task in five subtasks, each focusing on a specific group of scientific disciplines. The groups were selected in accordance with the subdivision already adopted in the HET Document "*The scientific case for European supercomputing Infrastructure*" [3] which contributed to the creation of PRACE. Unlike the HET document, we preferred to put Computational Chemistry in a single subtask, as we are convinced that this discipline needs a considerable effort to petascale codes that often have limited scalability.

The five subtasks under the coordination of Giovanni Erbacci, (CINECA) are:

- *Material Science, Nanoscience and Life Science*
  Led by Maria Francesca Iozzi, SIGMA, University of Oslo.
- *Computational Chemistry*
  Led by Paul Sherwood, STFC
- *Astrophysics, Cosmology, High Energy Physics and Plasma Physics*
  Led by Riccardo Brunino, CINECA
- *Weather, Climatology and Earth Sciences*
  Led by John Donners, SARA
- *Engineering and CFD*
  Led by Joerg Hertzer, HLRS.

The first objective was to indentify which communities of researchers needed support to petascale and optimise their key applications which are of value to the community itself.

The main criteria adopted were:

- effective use of the PRACE Tier-0 systems;
- long-term collaboration (on the order of 8 -10 months or more);
- applications of interest for a scientific community, not just for a single researcher;
- size and maturity of the community as HPC user;
- emphasis also on new and emerging scientific communities.

2.1 Material Science, Nanoscience and Life Science

Material Science, including Nanoscience, and Life Science form two large independent communities (see the definition above) that is *(i)* in both areas there exist many groups developing and/or using different codes and *(ii)* often developers and users communities do not work closely together. In order to meet the needs of such independent communities, it is important to identify macro-areas (large groups of users or developers) and understand the differences and similarities between the codes used or developed in those areas. Obviously the partitioning into macro-areas is somewhat arbitrary. In the following we analyse and interpret the needs of the two independent communities in terms of the computational methodologies each of them uses.

We observed that both life- and material-scientists make use of Quantum Mechanics (QM), Density Functional Theory (DFT) and Molecular Dynamics (MD) methods. The choice of the methods depends on the properties under study: for example, structures and physical-chemical properties are accurately determined by QM and DFT, whereas MD algorithms are used to investigate time-dependent properties and structures of very large systems. In general terms we noticed that QM and DFT methods are implemented in very lengthy codes and the parallelisation might become a difficult task if not introduced at the early stages of the code design. MD algorithms are more easily implemented and parallelised, but often require quite large memory and an efficient matrix distribution strategy.

Even if QM, DFT and MD methodologies are adopted in both Material Science and Life Science, their working equations and therefore the corresponding implementations depend on the area of usage. For example, a certain program designed to perform a QM calculation in Material Science cannot give the same accuracy when applied to a Life Science problem.

In the Material Science area, scientists are interested in determining structures and properties of solid materials and surfaces which are accurately described by means of QM and/or DFT calculations. The codes make use of plane-waves that are best suited to describe the periodicity and the band-like structure of a solid material. It is worth mentioning a few well-known codes based on these methodologies: VASP, QuantumEspresso, ADF-band, SIESTA, and GPAW. In Material Science time-dependent properties (such as diffusion, absorption, transition-phases) are also important and are more easily determined by using MD methods. Common codes in this area include for example LAMMPS and DL_POLY.

In the Life Science area the structures and properties can be determined by doing QM or DFT calculations. In this case the codes are based on the use of gaussian basis function and adopt a non-periodic approach. Programs often used in the area include GAUSSIAN, Turbomole, DALTON, Q-Chem, and many more. It is important to emphasise that the same codes can be used in both Life Science and Computational Chemistry, the border between the two being simply determined by the chemical nature (organic/inorganic molecule or bio-molecule) and size (less than or more than 100 atoms) of the system under study. Dynamics are also extensively used in Life Science to do structure recognition or determination at very large scale (i.e thousands of atoms). Codes widely used in the area are GROMACS and NAMD.

It is worth mentioning the CP2K code, and its module QUICKSTEP, respectively performing MD simulations and DFT calculations on very large scale. The important feature in these two programs is the use of a mixed plane-wave/gaussian approach which makes the calculations feasible on large systems for both Material Science and Life Science problems.

Many of the codes mentioned have a long history and were originally designed for a scalar architecture. Thus, the introduction of efficient levels of parallelisation (MPI or hybrid MPI/OpenMP schemes) might in some cases require large modifications of the innermost structure of the original code. For this reason, some widely used codes are not yet orienting their development towards petascale parallelisation; on the other hand some codes have evolved quickly towards high scalability and are now ready for petascale optimisation. The codes presently under optimisation for petascaling are: Quantum Espresso, GPAW, GROMACS, CP2K, DL_POLY and DALTON (the latter two will be described in the Computational Chemistry section).

2.2 Computational Chemistry

Computational Chemistry, in the context of this work package includes quantum chemistry (electronic structure of molecules) and the use of molecular dynamics and mechanics for the study of inorganic systems (solutions and materials). As such, there is a strong overlap with subtask "*Material Science, Nanoscience and Life Science*", and the two subtasks have been conducted in collaboration.

Computational Quantum Chemistry is a mature and established field, with widespread use in many areas of application. Probably the most common calculation of this type performed today is the application of Density Functional Theory (DFT) to small and medium sized molecules. Such studies can provide good estimates of molecular structure, spectroscopic properties, energy differences between reactants, products and the transition states for chemical reactions, thus providing valuable information on the mechanism of chemical reactions. A variety of other techniques exist, trading increased accuracy and reliability for increased execution time. Of these, important examples would be MP2 theory, a long-standing method of particular value in organic chemistry, and variants of Coupled Cluster Theory (e.g. CCSD(T)) which (for favourable cases) can be regarded as being of benchmark quality.

There is a wide variety of software available for computational quantum chemistry today, and the community (both of developers, and of users) lacks structure. Many of the quantum chemistry programs have a long heritage and many individual researchers have strong ties to one or more program packages. Developer groups, in particular, have strong allegiance to their chosen package, since implementation of a new method often builds (scientifically and technically) on the work that has gone before and each programme package offers a lot of infrastructure and utility functionality on which new development builds. There are not many examples of substantive strategic collaboration between the teams associated with a given code, although in some cases a piece of functionality may appear in more than one package as a result of a specific agreement.

Europe has a good track record in developing molecular quantum chemistry packages. An incomplete list might include Turbomole and Orca (Germany), MOLPRO (UK and Germany) Dalton (Scandinavia), GAMESS-UK (UK and the Netherlands) Columbus (Austria and US). Internationally the market is probably led by Gaussian (USA) and there are many other US-based packages, including a recent addition to the Open Source Software base, NWChem (developed using mainly DoE funding at PNNL). For the purposes of PRACE, it was felt better to choose a European package to maximise the benefit of (relatively modest) European funding to the wider EU research base.

Turning to Molecular Dynamics (MD) codes based on empirical force-fields, the predominant users are in the life sciences community, and GROMACS is the leading European example of a code designed for that community. More broadly, there are many applications of MD in the materials science and other areas of chemistry and one area in particular is becoming strategically important in Europe, specifically the simulation of large pieces of an inorganic material when subjected to radiation damage. These processes, which depend crucially on atomistic detail, require simulations of millions of particles and very large compute resources. The field is rapidly developing and new force field formulations are being developed to handle the interatomic interactions which often lie between the extremes of covalent and ionic bonding. DL_POLY in Europe, and LAMMPS in the USA are two examples of codes that are specialised for this type of work.

2.3 Astrophysics, Cosmology, High Energy Physics and Plasma Physics

Computational Astrophysics has traditionally been one of the scientific disciplines capable of gaining a considerable boost from the exploitation of high end computing systems. The development of innovative algorithms and their implementation have always been carried out with awareness of performance issues and platform optimisation, in order to get the maximum speed-up: these activities have traditionally been carried out by individual research groups and implemented in numerical codes used by the developers themselves and by a limited number of collaborators. In this sense, the Computational Astrophysics community is to be considered as "independent", even though a limited number of “suites” of numerical solvers have gained a certain popularity throughout the community itself. Such an example is the PLUTO code (http://plutocode.ph.unito.it/) that already has good scalability over tens of thousands of CPUs.

A particular role in Computational Astrophysics is held by Computational Cosmology: the peculiarity of the physical problem gives rise to many challenges requiring innovative algorithmic solutions and that have proven to be able to significantly stress current advanced supercomputer architectures. The very demanding effort of developing and maintaining such a numerical code has proven to be beyond the reach of many academic research groups'. A small set of numerical codes for cosmology has emerged in High Performance Computing: in Europe, Dr. Volker Springel and the Virgo Consortium have developed a code named Gadget-2 which belongs to the PRACE benchmark suite. The code version available to the public is not up-to-date, whereas the full-physics state-of-the-art versions of the code are essentially distributed among the Virgo Consortium collaboration which is the main organised body in the Computational Cosmology community. Contacts have been established in order to try to set up a fruitful collaboration. A wider spectrum of European numerical codes for Computational Cosmology can be retrieved at the following URL: http://www.itp.uzh.ch/~agertz/Wengen_2/Home.html, even though attempts to establish contacts with the authors (i.e. the code developers) have proven to be difficult.

Currently, much attention is focused on the possibility of producing energy in an environmentally friendly manner: promising candidates have been indentified in Plasma Fusion processes and two main experiments are in the process of being started, namely ITER (http://www.iter.org/) and HiPER (http://www.hiperlaser.org/). The momentum generated by these two initiatives has also influenced and stimulated the efforts for the development of computational models able to produce useful predictions that can efficiently drive the experimental design. In order to do so in an effective way, the best possible exploitation of current petascaling Tier-0 facilities offered by the PRACE partnership is a great opportunity: with such an aim, the EUTERPE code has been selected for petascaling activity within the present PRACE task.

Theoretical and Particle Physics are currently not yet engaged in a well established collaboration with the PRACE WP7: nevertheless, numerical modelling in these fields has shown to be already able to scale up to very large numbers of processing units and also to be able to exploit heterogeneous architectures. This appears to be partially due do the inherent characteristics of the adopted algorithmic approach. For example, some QCD applications have already shown to make very efficient use of petascale installations as demonstrated by the fact that very high fractions of the sustained peak performances are reached on PRACE Tier-0 systems (e.g. JUGENE).

2.4 Weather, Climatology and Earth Sciences

The community of Meteo-Climatology at European level is mainly related to the National Meteorological services and research groups spread across different universities and regional research centres. Most of those involved in this research discipline are linked via the IS-ENES infrastructure and cooperate together with a well recognised research activity. The IS-ENES community contacted PRACE and we agreed a specific cooperation plan to petascale applications for Tier-0 resources. For this reason in Task 7.2 it was decided to concentrate the contacts and the activity with the Meteo-climatology community specifically through the IS-ENES community. For a description of the activity with this community we refer to Section 4.

The Earth Sciences community and, specifically, the Geophysics community, are quite diverse and spreads across different disciplines and interests from earthquakes, to volcanology, surface dynamics and tectonics. Different research centres and national institutes are involved in specific research programs involving simulation activities using different models and applications. The earthquake and seismology research addresses both fundamental problems in understanding Earth's internal wave sources and structures, and augmenting applications for societal concerns about natural hazards, energy resources, environmental change, and national security. This community is central in the European Plate Observing System (EPOS), the ESFRI initiative in solid Earth, recently consolidated in a research infrastructure. Some contacts with the EPOS community as representatives of the Geophysics community indicated the SPECFEM3D code as a code of interest for Tier-0 systems.

2.5 Engineering and CFD

In engineering and CFD (Computational Fluid Dynamics) different users, depending on their individual problems, use quite different software. For this reason the application community of engineering and CFD is not organised as a single community.

In industry, codes from ISVs (Independent Software Vendors) dominate and usually the software vendors do not disclose the source code.

Such ISV codes are often used in research groups, while software developed especially for individual needs, is widely used in research groups. Nevertheless openly available codes are used both in industry and research. Typically, from this wide area, CFD codes have the highest requirements on computing power and consequently the highest demand for HPC systems.

Only codes not bound by the disclosure limitations of ISV, can be worked on within WP 7.2. During our discussions we found that OpenFOAM and Code_Saturne, both from the CFD area, are the most widely used codes being openly available and requiring top level HPC systems for current problems in research and industry.

2.5.1 *OpenFOAM*

The OpenFOAM® (Open Field Operation and Manipulation) CFD Toolbox is a free, open source CFD software package produced by a commercial company, OpenCFD Ltd. It has a large user base from most areas of engineering and science, from both commercial and academic organisations.

The user and developers community is wide and active throughout Europe, the US and India. The proponents are in touch with different research groups working on relevant scientific and engineering cases which will benefit from the enabling of OpenFOAM on a Tier-0 system to make advances on their research activity.

One of these research groups is the Internal Combustion Engine Group (ICE Group) at the Energy Department of the Politecnico di Milano, Italy, whose research activity is focused on internal combustion engines, with a particular emphasis on the development and application of advanced one-dimensional and multi-dimensional numerical models.

A major OpenFOAM user at the University of Stuttgart is the Institute of Fluid Mechanics and Hydraulic Machinery. Their focus is the analysis of incompressible flows in turbines and related technical equipment.

In collaboration with the Centre for Climate and Air Pollution studies at NUI Galway ICHEC prepared a test case for modelling the flow around a vessel. The motivation for this was the installation of a suite of sensors to directly measure the air-sea exchange of carbon dioxide. The overall objective is to improve the estimates of air-sea fluxes from ships.

2.5.2 *Code_Saturne*

Code_Saturne® is a multi-purpose Computational Fluid Dynamics (CFD) software, which features a co-located finite volume method for compressible and incompressible flows with or without heat transfer and turbulence for a wide range of cell types and grid structures. It has been developed by *Électricité de France Recherche et Développement* EDF-R&D since 1997 and was released as open source in 2007, distributed under a GPL license. The code was originally designed for industrial applications and research activities in several fields related to energy production; typical examples include nuclear power thermal-hydraulics, gas and coal combustion, turbo-machinery, heating, ventilation, and air conditioning. The main hub in the UK is the University of Manchester, who are part of the UK Turbulence Consortium (UKTC); they manage the *Code_Saturne*® Wiki, and have been using and co-developing the software since 2001.

It should be noted that *Code_Saturne*® is available for anonymous download from the official website and no registration is required. It is difficult to determine exactly how many *Code_Saturne*® users there are worldwide as the code is available via anonymous download. On average, the *Code_Saturne*® wiki has received over 4,000 hits per month since the code became open source. The usage of the code is truly world-wide as shown by the map of Wiki accesses presented in Figure 1.

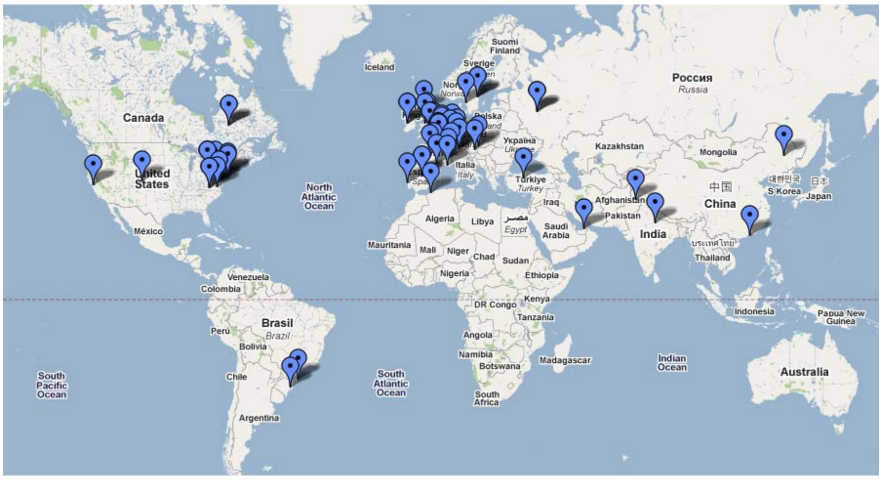

Figure 1: World map showing Wiki accesses to Code_Saturne

The main application areas of Code_Saturne are: nuclear power plant optimisation in terms of lifespan, productivity and safety, combustion (gas and coal), electric arc and joule effect, atmospheric flows and radiative heat transfer. For more information visit:

http://cfd.mace.manchester.ac.uk/twiki/pub/Main/MarcSakiz/CodeSaturneGeneral-2007-05-10.pdf

3 Cooperation with Independent Communities

As reported in Section 2, the independent communities are mainly characterised by groups of scientists interested in the same computational methodology, who use a specific group of application codes of value for the entire community, to advance their science. In this section we describe the policy defined to cooperate with these communities and the methodology adopted to select the applications considered for petascaling.

3.1 Policy to to Identify collaborations

The goal of Task 7.2 is to assist scientific communities with petascaling their applications, if possible through longstanding relations. We have selected those communities by applying four criteria:

1. Communities with a consolidated maturity in terms of computational methods used and proven access to the HPC facilities. These are typically the scientific disciplines which have already used HPC facilities for a long time. Here the objective was to cooperate with communities that really need to improve one or more of their application codes for efficient Tier-0 usage.
2. Importance and the representativity of the community. It was established that the communities of interest should be characterized from the use of specific codes that they develop, maintain and promote over a wide number of users spread in different countries, to advance science, In other words, it is important to work on application codes that once prepared for petascale performance can be made available on Tier-0 systems to a wide community and not just to a small group of users.
3. Long-term collaborations with the community. It is important to define a plan together with the owners of the code and some members of the community, to identify better the performance bottlenecks in the code and to cooperate to identify common methodologies and improvements. The PRACE members involved in Task 7.2 in general are experts in the HPC methodologies needed to petascale and optimize the codes, but not in the scientific discipline and the scientific model at the basis of the code. Thus, the cooperation with the owners of the codes plays an important role for providing improvements to the application, both with respect to the scientific methodology and the HPC performance.
4. Availability of the codes. It was decided to consider mainly application codes not subject to specific license schemes, in order to maximise the support for the users of these codes on different Tier-0 systems and towards the scientific community at large.

3.2 Methodology adopted

We now applied a two-pronged approach to start the work. We started by getting in touch with the structured communities on one side, while at the same time we started a process for selecting specific application codes which are widely used by important independent communities.

The scientific communities in general are envisaged not from their structure and organisation but mainly by the use of particular application codes in specific scientific domains, as reported in the survey conducted in PRACE PP WP 6 [1] and recently by a similar survey prepared by PRACE 1IP WP 7 Task 4 and documented in Deliverable D 7.4.1 "*Applications* *and user requirements for Tier-0 systems*" [2]. The same attitude was confirmed also from useful discussions we had with partners involved in PRACE 1IP WP 4 "*HPC Ecosystem Relations*" and WP5 "*Industrial User Relations*". The objective of these discussions was to identify, at the beginning of Task 7.2 activity, opportunities for enabling applications through engagement with selected scientific communities, industrial users involved in public research, and specific application projects.

In order to cope with independent communities, Task 7.2 has opened a call to identify important groups of scientific users, seen as a single scientific community, which need to petascale their applications to Tier-0 Systems.

The call was organised in cooperation with the partners in PRACE as they have a consolidated knowledge of the scientific communities and of the application codes they use. The scientific communities are spread among different countries and mainly use the computational resources provided by the different national HPC centres.

Apart from the actual selection of important application codes, another objective for the call was to get in touch with the community representatives, with them identify a number of applications which are relevant for the community, and then try to set up a collaboration and interaction with the application developers and the community as a whole.

3.3 The call for applications and related scientific communities

The call was defined in October 2010 and a form was prepared to gather the requirements in a uniform way.

The form contained a section to describe the main characteristics of the code with respect to scalability, in terms of algorithms, numerical methods, I/O methodology, parallelisation strategy, current performances and performance bottlenecks. A further section reported the motivations for petascaling, a description of the work to accomplish, possible risks and a contingency plan.

Furthermore, an estimation was requested of the involvement in terms of Person Months to complete the activity (PM's), both in terms of effort from PRACE WP 7 and involvement of the developers of the code and the scientific community using the code.

The form has been included in Annex I. A wide distribution of the form was issued through the main groups of users using the HPC Centres involved in PRACE. In addition, the subtask leaders appointed in Task 7.2 had the role of disseminating the information and contacting the scientific community of which they are a member.

The deadline for submission was December 10th, 2010 and 19 valid proposals were submitted. The proposals requested the optimization of the following applications codes:

- **Life Sciences**:
  Brain cores+ANSCore: (model of the global brain neocortical network and large-scale neural network simulator;

- **Material Sciences**:
  GROMACS**:** Classical particle Molecular Dynamics for biomolecular simulations;
  GPAW**:** Time Dependent DFT using the projector augmented wave method;
  QuantumEspresso**:** Atomistic simulations of solid state systems using ab-initio and DFT;
  CP2K**:** atomistic simulations of solid state, liquid, molecular and biological systems;

- **Computational Chemistry** :

Crystal**:** gaussian basis electronic structure on periodic materials;
Dalton**:** Quantum Chemical electronic structure calculatons;
DL_POLY**:** Mulecular Dynamics of materials or biological systems;

- **Astrophysics, Cosmology and Theoretical Physics**:
  PLUTO**:** Astrophysics Magnetic Hydrodynamics;
  PRMAT**:** Atomic, Molecular and Optical physics;

- **Plasma Physics**:
  OSIRIS**:** A 3D relativistic particle in cell code for modelling plasma based accelerators;
  EUTERPE**:** A code for solving the gyrokinetic equations via the particle-in-cell Monte-Carlo method;

- **Meteo-Climatology** (in cooperation with the IS-ENES Community):
  EC-EARTH 3 **suite** (***IFS*** model for atmosphere, ***NEMO*** model for ocean, ***OASIS 3-4*** the coupler);

- **Earth Sciences**:
  SPECFEM3D: Geophysics, earthquake simulations;

- **Engineering and CFD**:
  Code_Aster: Structural Mechanics;
  Code_Saturne: Computational Fluid Dynamics;
  OpenFOAM: Computational Fluid Dynamics;
  SIMSON & NEK5000: Computational Fluid Dynamics;
  TELEMAC: Computational Fluid Dynamics, especially hydrodynamics.

3.4 First selection of Applications

At the PRACE WP 7 Face-to-Face meeting on December 15-16, 2010 in Bologna, after an evaluation of all the applications submitted, five applications were selected. The selection was unanimous between all the partners involved in PRACE 1IP WP7.2 and approved by all the partners in the WP7 meeting.

In addition to the criteria defined in Section 2 for the community code selection, further criteria for choosing this group of applications were defined as follows:

1. The application is clearly an important application in the European field. This can be expressed by:
   - application that shows up in earlier PRACE surveys;
   - part of the PRACE Application Benchmark Suite (PABS);
2. The application is a clear representative of applications used by independent communities;
3. Quick start within Task 7.2 possible, including quantification of PRACE PM effort and committed scientific community effort;
4. Complete proposal, including information on and analysis of performance bottlenecks;
5. Potential industrial interest.

Additionally, during the selection process, Task 7.2 has ensured that:

- Only a limited number of PMs is allocated now, to ensure that adequate resources were left for further work;

- Selection of the applications so as to cover most scientific disciplines;

The applications selected at the Face-to-Face meeting are reported in Table 1. The table reports also the effort in terms of PM estimated in the proposals as a request to PRACE WP7, divided in the effort required directly to Task 7.2 for petascaling the applications, and for further cooperation for the activities in the realm of Task 7.5 and Task 7.6, such as hybrid programming with accelerators and support for enhancing I/O. In addition, the effort from the Community, or the owners of the codes, is also reported.

Three of the selected applications (GROMACS, Quantum Espresso and GPAW) are part of the PABS Benchmark Suite defined in PRACE PP. A performance analysis of these three codes was initially done in PRACE PP WP6 and some points were identified to better investigate to enhance the scalability, as reported in Deliverable D 6.5, "*Report on porting and optimisation of application* [4]. The indications provided then are an important starting point now to improve the scalability of these three application codes.

One of the codes selected (EC-Earth 3) was the result of an on-going collaboration between Task 7.2 and the IS-ENES Meteo-Climatology community (see Section 4). And another code (OpenFOAM) is of main interest also for the industrial communities and there was an expression of interest from PRACE 1IP WP5 "*Industrial User Relations*".

Application	**Community**	**Effort WP 7.2**	**Effort WP 7.5**	**Effort WP 7.6**	**Effort Community**
GROMACS	Biochemistry – Molecular Dynamics	6 PM	4 PM	4 PM	2 PM
GPAW	Nano Science	12 PM			1 PM
Quantum Espresso	Material Science	12 PM	5 PM		3 PM
EC-Earth 3	Meteo Climatology	10 PM		10 PM	5 PM
OpenFOAM	Engineering and CFD	15 PM		3 PM	1 PM

Table 1: First group of application codes selected, related communities and effort estimated

A total effort of about 50 PM from Task 7.2 has been allocated for petascaling this first group of applications. The petascaling activity started soon after the Face-to-Face meeting, after agreement on an updated work-plan of the activity, with a real distribution of work between the PRACE partners, the people external to PRACE, like the owners of the code and, where possible, participants from the scientific communities. The objective was to conclude the activity in an elapsed time of 12 months between January and December 2011.

The progress on these five application codes is documented in Section 5.

Not all the 14 remaining applications met all the requirements governing Task 7.2: some proposals did not document well a long-term effort required for petascaling, or the involvement with the owners of the codes and the scientific communities interested in using the applications. For example, proposals with a small effort in terms of PM should have been redirected directly to the preparatory access call in PRACE.

For these applications it was decided to give some more time to the proposers, in order to better clarify the real involvement in terms of PM and the real involvement of the owners of the codes and the scientific community. In the meantime it was decided to define the selection rules for this remaining group of proposals. The new deadline for resubmitting the revised applications was set for February 27th , 2011.

3.5 The selection process

At the deadline, 12 of the 14 original proposals had been resubmitted with a revised plan. Two proposals were withdrawn by the proposers: SIMSON & NEK5000, related to the CFD community, and TELEMAC, related especially to hydrodynamics within CFD.

It was agreed between the WP 7 task leaders to adopt a selection procedure based on a voting scheme, similar to the one already used in PRACE 1IP and approved for the internal call in Task 7.1. Furthermore it was agreed to allocate no more than 50 PM for Task 7.2 for this selection. This limit will allow WP7 to handle around 5–6 projects, depending on their size.

In this way, some more effort will remain available in Task 7.2 for further petascaling of applications or for cooperating with leading scientific communities to petascale new challenge applications in the last period of Task 7.2.

3.5.1 *Voting procedure*

The review process for the 12 proposals took place between March 9th and March 21st 2011 at 18:00 (CET). The 12 proposals to review were uploaded on the BSCW at: https://bscw.zam.kfa-juelich.de/bscw/bscw.cgi/523438 and circulated among the PRACE 1IP WP7 partners. Each partner had the opportunity to read and vote on the proposals. The voting rules were:

- Each partner had five votes
- Each partner had to use all five votes, or none at all
- Each vote had the same value
- Only one vote per proposal per partner
- Partners were allowed to vote for proposals which had participants from their own site.

A PRACE partner was here defined as the list of beneficiaries on page 4 in the PRACE-1IP DOW (only GCS for Germany). Only votes which had arrived before 18:00 CET March 21st 2011 were considered valid.

3.5.2 *Review criteria*

The following criteria were used to review the proposals:

1. The proposal must address petascaling/optimisation of a key application which is of value to a wide community of researchers, so that they can effectively use the PRACE Tier-0 systems, and not just to a single group of users or a single institution.
2. The scientific community and/or the owner of the application code should be involved in the petascaling activity with a clear written effort in terms of PM.
3. Size and maturity of the community as HPC users.
4. A long-term collaboration should be envisaged: at least 8-10 PM of total effort. This effort should be contributed from PRACE Task 7.2, but also from Task 7.5 and Task 7.6 and possibly from the scientific community or from the owners of the application code.
5. The underlying scientific problems and the datasets used for enabling have to be potentially relevant for Tier-0 systems. Minimum Tier-0 allocations are at the moment:

- 8000 cores on BG/P (JUGENE)
- 2000 cores on Nehalem QDR-IB machine (Curie).

All the proposals and a document describing the call, the selection process and the reviewing instructions were uploaded on the BSCW at https://bscw.zam.kfa-juelich.de/bscw/bscw.cgi/523438

3.5.3 *Voting results*

On March 23rd the results of the voting procedure were compiled: 18 countries voted and one abstained, out of 20 PRACE countries. The results of the voting action are presented in Table 2. The table reports also the effort required in terms of PM, requested from PRACE WP7 and the effort offered by the scientific communities or the owners of the application codes, as reported in the proposals.

The effort required by PRACE WP7 is mainly delivered from Task 7.2 but there are some applications requiring also some additional activities from Task 7.5 or Task 7.6, i.e. new algorithms, hybrid programming with accelerators, visualisation, parallel I/O optimisation, etc. Where appropriate, the efforts requested from Task 7.5 and Task 7.6 have been reported in the table. These efforts have been gathered directly by the proposals, when explicitly reported, or estimated, when only a qualitative description or interests were written in the proposal.

Name	**Total Votes**	**Total PM (WP7)**	**PM (7.2)**	**PM (7.5)**	**PM (7.6)**	**PM Non PRACE**
DL-POLY (Computational Chemistry)	**13**	**12**	**3**	**9**	**--**	**6+**
CP2K (Material Science and Life Science)	**12**	**10**	**10**			**2**
Code_Saturne (Engineering and CFD)	**12**	**14**	**12**		**2**	**2**
DALTON (Computational Chemistry)	**9**	**12**	**12**			**2**
SPECFEM3D (Meteo-Climatology and Earth Sciences)	**8**	**11**	**6**		**5**	**1**
EUTERPE (Astrophysics, Cosmology and Plasma Physics)	**7**	**6**	**2**	**2**	**2**	**3**
Code_Aster (Engineering and CFD)	**7**	**9**	**6**	**3**		**1**
PLUTO (Astrophysics, Cosmology and Plasma Physics)	**6**	**3**	**2**		**1**	**9**
CRYSTAL (Computational Chemistry)	**5**	**12**	**12**			**18**
Braincores&ANScores (Material Science and Life Science)	**4**	**6**	**5**		**1**	**6**
PRMAT (Astrophysics, Cosmology and Plasma Physics)	**4**	**8**	**3**	**2.5**	**2.5**	**--**
OSIRIS (Astrophysics, Cosmology and Plasma Physics)	**3**	**10**	**5**	**2**	**3**	**2**

Table 2: Voting results and PMs planned

3.5.4 *Accepted projects*

As reported at the beginning of Section 3.5, Task 7.2 had already allocated an effort of about 50 PM for petascaling the applications selected in December 2010. In the current selection the total effort to allocate in Task 7.2 was in the order of another 50 PM, so that sufficient PM remained for further cooperation with scientific communities.

From the results of the voting activity it was clear that the first three applications were the most supported, but after these, there was not a clear cut-off to indicate the groups of applications to select and the ones to reject. In fact the number of votes from DALTON to OSIRIS (9 applications) only ranges from 9 to 3.

WP7 therefore proposed to select the first three applications, DL-POLY, CP2K and Code_Saturne, for a total of 25 PM in Task 7.2, and discard directly the last 3 applications (which have less than 5 votes). Then it was decided to use the criterion which makes sure that Task T7.2 addresses those user communities that will need petaflop/s performances. In this way WP7 proposed to further select:

- DALTON: as it is the only real code in the area of computational chemistry (in fact DL-POLY is more in the area of Molecular Dynamics). Furthermore, DALTON is one of the codes used by the ScalaLife community, a community which recently was in contact with WP7 for petascaling applications (see Section 4);
- SPECFEM3D: this is the only code in the area of Earth Sciences (earthquakes) and is of great interest for the EPOS RI and the VERCE project recently approved by the EC in FP7;
- EUTERPE is the only code in the area of plasma physics, and used also by the ITER community.

In summary WP7 proposed to accept six applications in this round:

- **DL-POLY** (Molecular Dynamics)
- **CP2K** (Material Sciences)
- **Code_Saturne** (CFD)
- **DALTON** (Computational Chemistry)
- **SPECFEM3D** (Earth Sciences)
- **EUTERPE** (Plasma Physics)

The total effort requested for these 6 applications amounts to 45 PM for Task 7.2, a total effort of about 65 PM for the whole of WP7 (including Task 7.5 and Task 7.6) and a further direct involvement of more than 16 PM external to PRACE of the scientific communities or the developers of the codes.

In this way, considering also the 5 applications already selected in December 2010, the total effort allocated from Task 7.2 sums to about 95 PM leaving sufficient further PM effort in Task 7.2 to cooperate with other scientific communities in the near future, mainly with the communities which are not yet, or not sufficiently, represented in this list, like the Astrophysics and Cosmology, QCD and Theoretical Physics.

The results of the voting procedure and the considerations from WP7 were submitted to the Technical Board which decided to support the six projects as proposed by WP7 in its meeting of April 8, 2011, without any modifications.

The eleven applications selected in Task 7.2 (the five selected in December 2010 and the six selected in April 2011) are representative of the main scientific areas and involve the main scientific communities interested in huge Tier-0 resources, as reported in Table 3. The distribution of the scientific disciplines we support till now in Task 7.2 is in line with the distribution reported in D 7.4.1, "*Applications and user requirements for Tier-0 systems*" [2] in the scientific areas accessing current Tier-0 systems. Not covered yet by Task 7.2 are mainly theoretical physics and QCD which both demand a great amount of resources from the current Tier-0 systems, but have not yet requested support from Task 7.2. A reason for this could be the fact that application codes in this field (QCD in particular) already scale very well and are used in production runs on Tier-0 systems.

A brief abstract of each of the eleven application codes selected to petascale in Task 7.2 is reported in Annex II**.**

Scientifc discipline / Community	Application code
Molecular Dynamics	GROMACS DL-POLY
Material Sciences	Quantum Espresso CP2K
Nanosciences (DFT)	GPAW
Computational Chemistry	DALTON
Plasma Physics	EUTERPE
Meteo-Climatology	EC Earth 3
Earth Sciences	SPECFEM3D
Engineering and CFD	OpenFOAM Code_Saturne

Table 3: Application codes selected in Task 7.2 grouped by scientific disciplines

4 Cooperation with Structured Communities

Structured communities are not well represented and common in Europe, if compared to the different independent research groups developing and using their own common set of codes and methodologies to enhance science in a specific discipline.

Nevertheless some structured communities exist, like the ones organised around European laboratories, institutions and research infrastructures. Examples include the European Molecular Biology Laboratory (EMBL) for basic research in the molecular life sciences (http://www.embl.de); the VIRGO consortium for cosmological research (http://www.virgo.dur.ac.uk/); the community organised around CERN interested in the different aspects of theoretical particle physics; the European Fusion Development Agreement (EFDA) on Fusion research (http://www.efda.org/); the PLANCK project in charge for the first European space observatory whose main goal is the study of the Cosmic Microwave Background (CMB) i.e. the relic radiation from the Big Bang. (http://www.rssd.esa.int/index.php?project=Planck). Most of these communities have specific huge simulation activities programs but, in general, they have already well performing application codes and usually they only need access to computational resources for production activity. Furthermore, in some cases, they already have their own computational facilities.

We decided to get in touch with some of these communities in those scientific fields which need to access to heavy computational resources, to establish with them a long-term collaboration to enhance their application codes, in order to exploit the Tier-0 Infrastructure as efficiently as possible. Given the modest amount of person months available in Task 7.2, our policy was to cooperate with communities with a real need for petascaling and not just communities interested in production runs, or for which a petascaling activity was already in progress internally within the community itself. Currently, Task 7.2 has made progress with two communities: IS-ENES and MAPPER. The first one relates to the area of Meteo-Climatology and the second one gathers scientists (via an EU FP7 project) interested in multi-scale activities in different areas. In this Section we present the ongoing activities with these communities.

4.1 IS-ENES

Global climate models are at the basis of climate change science and of the provision of information to decision-makers and a large range of users. In Europe, the European Network for Earth System Modelling (ENES, http://www.enes.org) gathers together the European climate/earth system modelling community, which is working on understanding and prediction of future climate change. IS-ENES the InfraStructure for the European Network for the Earth System Modelling is an FP7-Project (March 2009-February 2013) funded by the European Commission under the Capacities Programme, Integrating Activities.

ENES, through IS-ENES, promotes the development of a common distributed modeling research infrastructure in Europe in order to facilitate the development and exploitation of climate models and better fulfill the societal needs with regards to climate change issues. IS-ENES gathers 18 partners from 10 European countries and includes the 6 main European global climate models. IS-ENES combines expertise in climate Earth System Modelling (ESM), in computational science, and in studies of climate change impacts.

Climate earth system models are key tools for understanding climate change and its effects on society and are at the basis of the IPCC (http://www.ipcc.ch).

Several European earth system models are in use at the moment: EC-EARTH, C-ESM, CNRM-CM5, COSMOS, HadGEM2-ES and IPSLESM. See https://verc.enes.org/models/earthsystem-models. This heterogeneity of models is needed to get an idea of the uncertainties that are inherent in any model. An earth system model consists of several components: an atmosphere model, ocean model, land-surface model, and possibly models for the carbon cycle, oceanic biogeochemistry, atmospheric chemistry and ice sheets.

One direction of research is to simulate the climate at a higher resolution at long enough timescales and with more additional components than is possible nowadays. Small-scale causes and effects are highly important to better answer policy-relevant questions: to understand the effects of climate change on hurricanes, or the link between coastal upwelling zones, fisheries and local climate and how it will affect local communities. New developments are the integration of a dynamical model of the Greenland ice sheet, which could be a major contributor to global sea-level rise. These single, high-resolution models are now run at an O(1000-10000) concurrency, which needs to be further increased to take advantage of the next generation of supercomputers. The inclusion of more earth system components requires a lot more computing power, even at the same resolution as is used nowadays.

Another approach is the use of ensembles to investigate uncertainties. A large number of simulations, each with an O(10-100) concurrency, is run independently and analysed afterwards. To reduce the time needed to complete all computations, this is usually done in parallel. At this moment, petascale computing is seen as a necessary condition to provide the answers in a reasonable time. It is still under investigation how petascale computing can have added benefits beyond a faster time-to-solution.

Although there are worldwide several large data repositories specialised for handling the IPCC (Intergovernmental Panel on Climate Change)-related modelling activities, there is a need for solutions to handling the large amounts of data that will result from petascale climate simulations. The EUDAT Data Infrastructure is looking into some of these issues, and PRACE needs to collaborate to provide adequate and integrated solutions that help scientists from their initial hypotheses to the final publications that answer them.

4.1.1 *Collaboration with IS-ENES*

In order to define the collaboration, a meeting with the IS-ENES community was organised in Paris on December 1st 2010. In this meeting 20 people from IS-ENES and PRACE 1IP Task 7.2 participated. During the meeting the different applications of interest for the community were analysed: three European coupled models needed to be run at very high resolution. These high end simulations currently reach to the order of 2000 processors (1/4° for atmosphere and ocean, through the integration with Oasis 3 which is pseudo parallel). At the meeting it was agreed that the scalability of the coupled system (atmosphere + ocean + coupler) is still poor and needs to be enhanced by a factor of 2 to 10 in order to fully exploit the Tier-0 PRACE facilities. Furthermore there was interest from the community to perform ensemble runs altogether which would allow accounting for statistical variability or multi-model needs.

The long-term objective is to be able to run the six European climate models on PRACE Tier-0 systems, however, as a first step, ENES proposes to start with one coupled model and benefit from this experience for all the six models.

At the end of the meeting it was agreed to start a collaboration with PRACE-1IP. Focus was given to the EC-Earth 3 application, which includes OASIS (used by all the models), NEMO (used by 5 out of 6 models) and IFS (spectral model in 4 of the 6 atmospheric models). This experience should benefit all the modelling groups, e.g. sharing results on scalability, optimisation, etc.

After the meeting a proposal was prepared and submitted for approval to Task 7.2 in December. The proposal is called EC-Earth 3 and became one of the five proposals selected to work on starting in December 2010.

4.2 MAPPER

Science today can benefit from advanced computational methods, thanks to the availability of huge HPC resources and advanced instruments (sensors, satellites, micro arrays, synchrotrons, etc) to collect and analyse large quantities of data. The models are increasing in complexity and scientists are commonly faced with the challenge of modelling, predicting and controlling multiscale systems which cross scientific disciplines and where several processes acting at different scales coexist and interact.

MAPPER is a project funded by the EC in FP7 to develop computational strategies for multiscale simulations across disciplines, exploiting existing European e-Infrastructures (http://www.mapper-project.eu/). MAPPER has the objective to deploy a computational science environment for multiscale computing, cooperating with other research infrastructures and EU Projects. The MAPPER Project contacted PRACE WP6 (in charge for the evolution of the distributed infrastructure) and WP7 (in charge for petascaling applications) to evaluate possible forms of collaborations between the two projects.

A meeting was organised on May 13th 2011 between MAPPER and PRACE WP6 and WP7. A total of about 20 people attended the meeting and for WP7 the WP7 Leaders and 3 representatives of Task 7.2 attended the meeting. The meeting offered the occasion to better understand the application in which the different members of MAPPER are involved and agree on a cooperation plan. The multiscale scientific applications of interest for the MAPPER community span over five challenging scientific domains: fusion, clinical decision making, systems biology, nanoscience and engineering.

4.2.1 *Planned collaboration with MAPPER*

At the meeting it was decided that the initial cooperation with PRACE Task 7.2 was for Task 7.2 to analyse a document prepared by MAPPER describing the scientific applications and the codes in which MAPPER is interested. The objective is to better understand and investigate the application requirements, the needs for petascaling, if petascaling for the selected applications is required, and how this could be established. A document describing the application of interest for MAPPER will be sent to WP7 after June 2011. Once this document has been analysed, a collaboration plan to petascale one or more of these applications on Tier-0 Systems will be designed. The objective will be to release a multiscale application that can exploit the Tier-0 PRACE Infrastructure to produce science in the challenge domains of interest for MAPPER.

4.3 Collaboration with other Communities

The activity in Task 7.2 continues to further improve contacts with other communities. In this direction, contacts are in progress with the Astrophysics and Cosmology communities to identify their special requirements better. As we have already emphasised with respect to the VIRGO consortium, many scientists from the astrophysics and cosmology community are interested in using computational resources both at Tier-0 and Tier-1 level. The impression is that they already have petascaled applications and only need resources for production runs. In any case, some more attempts will be made to get in contact with them. A major occasion to establish new contacts with the community at large will be the summer school "*Advances in Computational Astrophysics: methods, tools and outcomes"* organised in June 13-17 in Cefalù, (Sicily). by Prof. Roberto Capuzzo Dolcetta, member of the PRACE Access Committee. At the summer school the main computational experts in the astrophysics community around Europe will participate and a meeting with a representative of Task 7.2 will be organised during the school.

In addition, contacts with other scientific communities will be established in the next months in order to investigate their applications for exploiting efficiently the PRACE Tier-0 infrastructure. Some contacts with the ScalaLife Project, a EU project for Life Science software, are on going and are trying to identify specific cooperations for petascaling software of interest for the life science community.

5 Progress Report on Selected Applications

As reported in Section 3.4, the activity on the five application codes selected in the first round started early January 2011. Preliminarily, the partners in PRACE-1IP who had a stake in Task 7.2 were identified to perform the work requested in each application project. Then a more pragmatic work plan was defined to carry out the activity in a time frame of around 12 months, in order to complete the activity for these application codes in time to allow the communities to apply for resources in the next regular calls before the completion of PRACE-1IP. Then the activity started, in cooperation with the owners of the codes and, where possible, with representatives from the scientific community involved in each of the projects.

In general the work is proceeding well for all the applications and it is in time with the defined work plan. While accounts on the JUGENE and Curies Tier-0 systems were being prepared, the profiling activity and the initial performance tests for the application codes were done on local and Tier-1 HPC systems at the premises of the PRACE Partners involved in Task 7.2. In April 2011, access procedures to JUGENE and Curie were finalised, allowing the work to be transferred to Tier-0.

Subsequently we present the progress so far on the activities related to the five application codes selected in the first round.

For the other six application codes, selected in April 2011, the work is starting now. The partners from Task 7.2 to work on these new applications have been identified, and they have agreed to the effective work plan with the owners of the codes and the representatives of the communities involved. The last WP7 Face-to-Face meeting held on May 30 and 31, 2011 in Oslo has been instrumental in discussing the activity and for agreeing on specific cooperation with the other tasks involved, i.e. Task 7.5 and 7.6.

5.1 GROMACS

GROMACS is a scientific application in the field of Material Science and biomolecular simulations.

5.1.1 *Short description of the application and the planned work*

The activity has been structured in two branches:

1) Increasing the efficiency of reduction of arrays over threads (this activity is done with the support of Task 7.5).
2) Improving the domain decomposition scheme by implementing a topology aware communication

NCSA from Task 7.5 are involved in the first activity, while CINECA and KTH are working on the second part of the project.

In Molecular Dynamics simulations more than 90% of the time is spent in calculating the forces acting on the particles. In biomolecular or electrolyte simulations long-range electrostatic interactions play an important role. In most atomistic simulation packages such interactions are usually calculated using the Particle-Mesh Ewald (PME) method. PME splits the electrostatic interactions in a rapidly decaying short-range part and a slowly fluctuating long range part. For the long-range part the charges are mapped on to a grid and the problem is solved in k-space, which requires a 3D Fast Fourier Transform (FFT). This 3D FFT requires communication over the whole system. Since the particle-particle and the mesh force calculations are completely decoupled, the two calculations can be done on different processes. This allows the 3D FFT communication to be overlapped with particle-particle (PP) force calculations. The manner in which the PP and the PME tasks are allocated could strongly affect the performance of simulation and finding a way to distribute efficiently such tasks could be a way to reach a better scalability on massively parallel machines.

The aim of this project is to find the best way to partition the PP and PME nodes on the available Tier-0 machines, exploiting as much as possible their technical features.

The total plan for petascaling consists of 6 PM distributed between CINECA (4PM) and KTH (2PM). Furthermore, Erik Lindahl and Berk Hess from Sweden, two of the developers of the GROMACS application, are cooperating fully in the activity.

The activity actually started in April, during a Face-to-Face meeting held in Stockholm where all the partners involved in this activity were represented. During this meeting we identified the tasks and agreed on the distribution of work between the collaborators.

Based on the existing expertise of the teams and on the availability of a small prototype of BlueGene/P installed in CINECA, we decided to focus the activity of the CINECA team on the optimization of GROMACS on the JUGENE BlueGene/P architecture, the KTH staff being assigned the task of the optimization on the Curie and Cray machines.

The coordination of the work between the two teams is assured by regular conference calls and by a constant interaction with Berk Hess, as the contact person from the code developers.

In addition, we are making use of a Wiki, where all the documents and minutes of the conference calls are, and of a SVN repository where modifications of the code can be shared.

This activity should end in October 2011 and the first preliminary results should be available around July 2011.

5.1.2 *Status and preliminary results*

A preliminary part of the work consisted of the installation of version v.4.3.1 of the program on the JUGENE computer. Then a minor modification of the code was introduced in order to provide more detailed information on the partitioning scheme of the particle-particle (PP) and PME MPI tasks. Three standard systems were provided by the GROMACS developers for performing the tests, but we decided to start with a simple water box containing a total of 521564 SPC water molecules (SPC is a common simulation model for water). This system was chosen because it is fairly simple, but also because it contains a large number of atoms and therefore should show reasonable scaling on a BlueGene/P architecture. Other system parameters used for runs were fairly standard (2fs time step, Langevin thermostat, etc.) and each simulation was run for 2000 steps, i.e. 4 ps of simulation time. For the moment we decided to fix the number of cores used for the PME calculation to 1/3 of those used for the PP calculation, as recommended by the GROMACS developers. For launching batch jobs the **`mpirun`** command was used with –mode=VN (i.e. 4 MPI tasks/node) and with the exception of the number of tasks n=1024, the –bgconnection mode was set to TORUS (for n=1024 only –bg_connection=MESH is possible).

For these first simulations we decided to concentrate on two parameters controlling the partitioning scheme of the PP and PME interactions, one an option of GROMACS **(`-ddorder`**) and the other a flag available to the **`mpirun`** launch command present on the BG/P system **(`-mapfile`**). These options can be found from the documentation of GROMACS and the BG/P respectively, but in brief:

- **`-ddorder`** (possible values cartesian, interleave or pp_pme). This option of the mdrun command of GROMACS controls how the nodes for PP and PME calculations are allocated. Thus, -interleave will alternately allocate PP and PME nodes, while –cartesian will separate them (in this work we haven't studied the influence of –pp_pme).
- **`-mapfile`** (XYZT, TXYZ, or any other permutation or a file). The **`–mapfile`** parameter of the BG/P mpirun command can be used to specify the order in which the MPI tasks are distributed across the nodes and cores of the partition reserved. The default mapping on JUGENE is to place the tasks in XYZT order, where X,Y, and Z are the torus coordinates of the nodes in the partition and T is the number of the cores within each node (T=0,1,2,3). In our runs we have investigated the performance with –**`mapfile`** XYZT and **`–mapfile`** TXYZ.

The scaling runs we performed for our water box obtained by varying these two parameters are reported in Table 4, where the performance is quoted in terms of nanoseconds of simulation time per day, as indicated in the output of GROMACS.

	Performance (ns/day)			
	mapfile XYZT		mapfile TXYZ	
cores	-ddorder cartesian	-ddorder interleave	-ddorder cartesian	-ddorder interleave
1024	1.739			
2048	2.355	2.356	2.365	2.552
4096	2.425	2.251	2.429	2.485
8192	1.332		1.382	

Table 4: GROMACS scaling tests on JUGENE as a function of the -mapfile and -ddorder options

The table is incomplete as runs are still being performed but even from these limited data we can draw some tentative conclusions at least for the SPC water box simulated here:

- Despite the relatively large system size, the system does not scale well beyond n=4096 cores.
- There is a small but significant performance increase in using –mapfile TXYZ, compared to –mapfile XYZT for all cases.
- The –ddorder option also has an effect, particularly in the case of n=2048 cores.

Once this set and the benchmark runs have been completed we will have a better picture of the scaling behaviour but it is already clear that on the BG/P architecture how the nodes of the hardware are allocated to the different interactions (PP or PME) is important.

5.1.3 *Cooperation with the owners of the code, the Scientific Community or users*

The reference persons for the developer's community are Erik Lindahl and Berk Hess. Both of them are located in the KTH in Stockholm where also the KTH team involved in this task is based. CINECA is in regular contact with Berk Hess and KTH via phone or video conferences. The developer's team provided also the benchmark cases to work on and some documentation to better understand the directions to follow to fulfil the aims of this project.

5.2 Quantum Espresso

Quantum Espresso is a scientific application in the field of Material Science.

5.2.1 *Short description of the application and the planned work*

The project aims to enable the Quantum Espresso package to perform the linear response NMR calculation through PWscf and GIPAW codes at large scale. While the PWscf code has shown good scalability up to thousands of processors as reported by PRACE benchmarking activity and some DECI projects, GIPAW and the general linear response codes show poor scalability beyond a certain scale. The main goal within Task 7.2 is to overcome this limit and eventually enable the response code to run at petascale. The original idea in the project proposal was to implement a hybrid parallelisation OpenMP/MPI model for GIPAW that mirrors the scheme already implemented in PWscf. We believed that this would have improved significantly the scalability. However, as a result of a deeper performance analysis and an intensive collaboration with the developers community we opted for a different approach. We eventually decided to develop two additional levels of parallelisation, consisting of splitting compute intensive loops across processes (low level) and in replicating process images across multiple numbers of processors (high level). The validation of this development and parallelisation activity will be tested on two different scientific cases, briefly (see the application project proposal for more detailed information):

1. The determination of the cholesterol crystal structures in human gallstones from first principles;
2. The characterisation of the aluminium and silicon distribution in zeolites by computing the solid-state NMR spectra.

In addition, an activity within Task 7.5 has been carried out to port the NVIDIA GPU architecture into the PWscf code. Our main goal here is to enhance the performance behaviour of PWscf on NVIDIA GPU. Positive results have been found when running the serial 1GPU/1CPU version developed at ICHEC [5]. Based on this progress we want to improve the CUDA version so that the user community can make use of a distributed version for parallel hybrid CPU/GPU systems when using OpenMP, CUDA and MPI.

The PRACE partners ICHEC and UNINETT-Sigma, together with the collaboration of Quantum Espresso users and developers, were principally involved in the activity in Task 7.2. The scientific community is mostly represented by the MML group at Oxford University, the University of Udine and the DEMOCRITOS group (and SISSA).

In line with our original project plan, we will complete the first parallelisation phase of the GIPAW code at the end of June, and start testing in the following few weeks. Afterwards we aim at improving the exact-exchange term of the DFT functional and optimising the cut-offs. In the overall project ICHEC has the main role of development while UNINETT-Sigma will be more involved in the testing phase. Further, both PRACE partners are involved in leadership and management of the whole project. The non-PRACE partner has the important role for proposing and discussing new solutions, collaborating in the analysis of the code and validating the results obtained. Due to a lack of man-power at ICHEC we will likely extend the pre-estimated project termination by two months, i.e. the project will proceed until the end of the year but not beyond.

5.2.2 *Status and preliminary results*

During the first phase of this project we improved the parallelism of the GIPAW package which performs the calculation of NMR spectra on physical systems. In order to understand the GIPAW parallelism it is worth describing the hierarchy of MPI communicators available in QE, where 5 levels of communicators have been defined. Three in particular are of interest for this project and are described below:

- **world**: native MPI group, all the processes (MPI_COMM_WORLD).
- **images**: processes are divided into different “images”, corresponding to a point in configuration space (i.e. to a different set of atomic positions) for NEB calculations; to one (or more than one) “irrep” or wave-vector in phonon calculations.
- **pools**: when k-point sampling is used, each image group can be subpartitioned into “pools”, and k-points can distributed to pools. Within each pool, reciprocal space basis set (PWs) and real-space grids are distributed across processors. This is usually referred to as “PW parallelisation”. All linear-algebra operations on array of PW / real-space grids are automatically and effectively parallelised.

For testing the scalability of the code on the PRACE machines the scientific community has provided us with three test inputs from the first scientific case mentioned above, i.e. the structures of cholesterol crystals. The three tests have different sizes, namely 592, 616 and 1184 atoms with band structures formed by 640, 672 and 1280 bands respectively. A smaller test case with 160 atoms and 180 band was provided and used for the phase of development.

A careful analysis of the profiler trace revealed that in the pristine version of the GIPAW code nearly 40% of the CPU time is spent diagonalising the DFT Hamiltonian, and 60% in the linear response routine (see the first entry in Table 5). This is characterised by several independent and computing intensive loops over the bands. The basic idea we used for addressing large-scale parallelism is to introduce a new level of parallelisation in order to divide these loops across different processes. For this purpose we make use of a new communicator named **bgrp** (band group). All processes (n_world) are hierarchy bunched in smaller sub-sets across communication levels starting from *world*. When the number of *images* (n_images) and *pools* (n_pools) are selected the number of processes is firstly distributed in a set of *images*. Each set is given by n_world/n_images. Then, splitting each new group of processes by n_pools we obtain the number of processes for each group of communication identified as a *pool*. The new *bgrp* adds a deeper level of communication as it allows us to further group the number of processes of each *pool*. We parallelise loops over the number of bands across this new set of processes, n_bgrp as follows:

```
! serial
do i = 1, nbnd
  a(:,i) = … Computation over i
enddo

! nbnd represents the total number
! of bands in the system
```

```
! parallel
do i = ibnd_start, ibnd_end
  a(:,i) = … Computation over i
enddo

! ibnd_start and ibnd_end are
! starting and ending index for
! each group of processor bands
```

The cost of the loop is now reduced by a factor of n_bgrp. The strategy is logically trivial. Since the array a(:,:) is initialised to 0 before every loop it is possible to perform each section independently and rebuild the global value afterwards. For this action we use *MPI_Allreduce.*

In Table 5 we show that this approach gives good scaling for the linear response with an acceptable efficiency on the global time up to using twice the initial number of processes.

Number of cores	*n_bgrp* value	Linear Response Functions Wall Time (sec.)				Total Time
		green_f	*cgsolve*	*ch_psi*	*h_psiq*	
48	1	1014.05	1004.62	996.87	848.75	25m 49.49s
96	2	567.56	558.83	554.74	457.92	16m 45.48s
192	4	339.85	331.41	329.24	261.87	12m 39.20s
384	8	203.27	194.30	192.85	143.50	10m 23.02s

Table 5: Timing report for linear response calculation performed with GIPAW code

A further level of parallelisation is currently under development. With this work we aim to introduce a higher-level of process distribution between the communication levels *images* and *world* distributing the embarrassing parallel computation executed for the 7 different *q* points. Due to the complex structure of the code, this is a tougher step to implement but it will allow scaling at a higher level. When this work is completed we expect users to be able to scale around 2 x 7 = 14 times more than the highest efficient level of scalability reachable with the older version of the code.

We now describe the results obtained by the porting of PWscf code on CUDA performed on Task 7.5. This version allows mixing the use of OpenMP, CUDA and MPI to fully exploit hybrid configurations of GPU/CPU systems. In Figure 2 we show that the CPU+GPU combination outperforms the CPU-only OpenMP version by a factor of 3 and delivers a factor of 7.8 increased performance over a single-core (CPU threads and GPU kernels on one NVIDIA C2050 card and one Intel six-core Westmere X5650). Further improvement can be obtained by exploiting all the available processing power within the workstation using both CUDA and OpenMP. In Figure 3 we show tests performed on distributed GPU/CPU nodes. The chart shows the best results possible for different configurations when running on CINECA's PLX GPU cluster. The results demonstrate that the best performance is achieved when using MPI across nodes and OpenMP plus CUDA within a node. Both the tables below report performance results obtained running the benchmark AUSURF112. Due to the relative small size of the 112 atom AUSURF112 benchmark, performance drops off after 16 nodes.

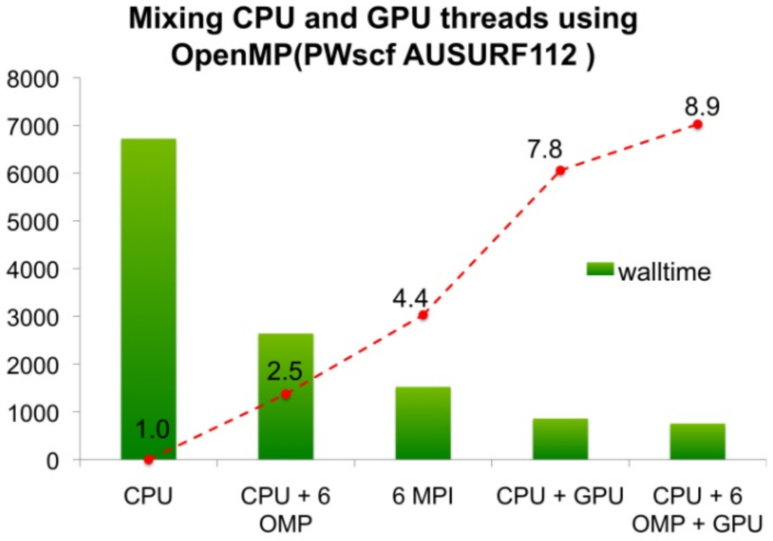

Figure 2: Timing for hybrid parallelisation

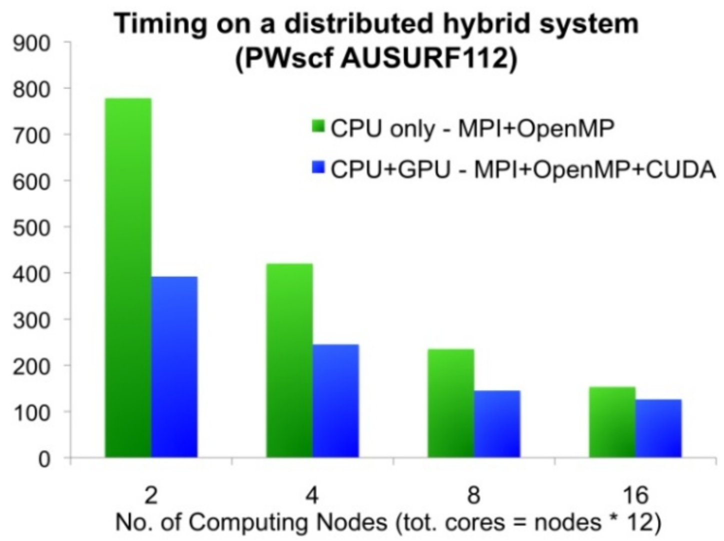

Figure 3: Benchmark on the GPU cluster

We obtained a speed-up of 1.9 when using two GPUs (where a single GPU is shared between 2 MPI processes), over the best possible CPU-only parallel configuration, running four MPI processes containing three OpenMP threads per MPI process (e.g. 12 cores in total). The CUDA kernels demonstrated the following speedups over the best OpenMP/MPI performance: about 7.8 times (**addusdens**) and about 3 times (**newd**). The parallel "**vloc_psi**", which contains most of the parallel FFTs performed during the calculation, is not yet fully developed and runs only on the CPU. We have also checked that the speed-up for parallel CPU+GPU GEMM operations is approximately 4.8 times.

Utilising a four GPU system, we were able to use one GPU per MPI process (rather than a single-GPU for two MPI processes). This increased the performance by about 2.3 times. The performance of the CUDA kernels did scale when using more GPUs: approximately 14.4 times for "**addusdens**", 5.9 times for "**newd**" and 8.3 times for the linear algebra calculations. This behaviour needs further investigation but it indicates that with this version of the code, it is better to add more GPU computing nodes rather than GPUs per node.

5.2.3 *Cooperation with the owners of the code, the Scientific Community or users*

The intrinsic characteristic of this project is the intensive collaboration between the scientific community and PRACE partners. This is primarily underlined by our aim to validate the results and the scalability of the new implementation basing the analysis on scientific projects proposed by members of the Quantum Espresso user community. Members of the user and developer community have been personally involved in the deliverable of this project as well as during the development phase currently ongoing.

The code analysis phase was done mainly during the first stage of the project, working at Oxford University and then at SISSA (Italy) and DEMOCRITOS (http://www.democritos.it) with members of the community and the developers group. At Oxford University the collaboration was mainly geared towards performance analysis of the GIPAW application. At SISSA the collaboration was to analyse strategies of parallelisation. The cooperation with the developers aimed to ensure that the newly developed code was perfectly aligned and coherent with the existing complex architecture of the Quantum Espresso package. This is essential for the following two reasons:

- To get the final product and the effort spent through PRACE to be recognised, validated and accepted by the scientific community;

- To be able to upload the new code into the official repository making it available for the whole community.

This cooperation has given outstanding results with the activity performed in Task 7.5. On May 5th, 2011, in agreement with the developers and maintenance team of the Quantum Espresso project, ICHEC released the first beta release of the GPU version of the PWscf code. The GPU version is currently hosted in a SVN branch of the official repository. The upload of the new version has been followed by an announcement on the PWscf forum and news was published on the official Quantum Espresso web site. In addition, a dedicated official forum was created in qe-forge.org.

As a follow-up of this task which aims at improving the scaling of the linear response calculation, we are planning to improve also the scaling of the diagonalisation present in the PW code. Note that GIPAW is linked against PW and uses PW's computational kernels for the diagonalisation. We have discussed possible parallelisation strategies together with Dr. Ceresoli and the QE developers at SISSA. The first strategy is to avoid diagonalising the DFT Hamiltonian for very large systems (which are the focus of this project), by implementing a “commensurate" linear response routine. This is currently being done by Dr Ceresoli at Oxford and requires very few changes to the code. Moreover, the commensurate approach boils down to multiplying the wavefunction by a simple phase factor and this operation can be easily parallelised over the star of q-points and over bands, with linear scaling (as opposed to the cubic scaling of diagonalisation).

The second strategy, in the long run, consists in modifying the PW code by introducing an extra level of parallelisation over bands, similar to what has been done in GIPAW. This will allow the exploitation of distributed and block-diagonalisation algorithms such as the residual minimisation with direct inversion in subspace (RMDIIS) or the local optimal block preconditioned conjugate gradient (LOBPCG).

5.3 GPAW

GPAW is a scientific application in the field of Materials Science and Nanoscience.

5.3.1 *Short description of the application and the planned work*

GPAW implements density-functional theory (DFT) using the projector augmented wave (PAW) method. The discretisation of the equations can be done with two complementary approaches, either with a finite-difference method using uniform real-space grids or with localised atomic orbital basis. The real-space approach is more relevant for the petascale applications. The PAW approximation offers good description over the whole periodic table of elements, and especially for first row and the transition metal elements it provides typically a better description than the more conventional norm-conserving or ultrasoft pseudopotential approaches.

In addition to standard density-functional theory, GPAW implements also the time-dependent density functional theory which can be used for studies of excited state properties such as optical spectra, which are beyond the realm of standard DFT. Two different forms of time-dependent density functional are implemented, the real-time formalism (RT-TDDFT) and the linear response formalism (LR-TDDFT).

The main objectives are the following:

- Reducing the initialisation overhead of Python in Petascale calculations
- Parallel distribution of Ω-matrix in LR-TDDFT calculations.
- Hybrid MPI/OpenMP implementation

- Better scaling alternatives to ScaLAPACK

The work plan and distribution of work between the partners is as follows:

• Python Initialisation:	Jan-June 2011, CSC
• Matrix distribution:	Feb-June 2011, CSC
• Hybridisation:	June-December 2011, CSC, IPB, Argonne
• ScaLAPACK alternatives:	March-October 2011, UmU

The first two objectives, Python initialisation and Ω-matrix distribution have been largely implemented on local systems. The activity is now underway on the Tier-0 systems to test the implementations at the large scale. The second priority is to get the hybridisation activity started. CSC will provide instructions and the initial code version so that IPB and Argonne will be able to start their work.

5.3.2 *Status and preliminary results*

We have implemented a patch for modifying the CPython interpreter in such a way that only single process performs the I/O related to initialisation, and uses MPI for distributing the data to other process. The patch has been tested in CSC's Cray XT5 system, and it seems to reduce the initialisation overhead significantly, as shown in Figure 4.

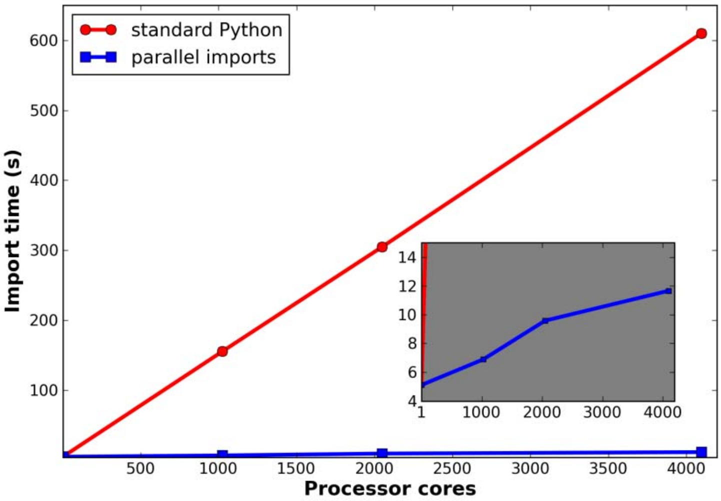

Figure 4: Initialization time of standard and modified Python interpreters

The patch is currently being tested at a larger scale on the Tier-0 systems.

The major part of the parallel distribution of Ω-matrix has been implemented. The remaining implementation work is related mainly to smoothing out the user interface for this parallelisation option. The calculation of matrix elements has already previously been implemented in parallel with good scalability. The matrix distribution is needed for solving the remaining memory bottleneck in very large calculations. The hybridisation activity has not yet started, but we are aiming to begin during June 2011.

Instructions about types and sizes of matrices have been communicated to UmU for the investigation of ScaLAPACK alternatives, however, only preliminary testing has been done within this action.

Both of these actions are being performed in collaboration with Task 7.5.

5.3.3 *Cooperation with the owners of the code, the Scientific Community or users*

The coordinator of the project, Jussi Enkovaara, is one of the key developers of GPAW. Thus, the project has good contacts with the whole GPAW development team, and cooperation has been done via developer-mailing lists, private emails, and teleconferences.

In early May, a GPAW course was given at CSC with several participants from Finnish user groups. In early June a presentation discussing also some key issues related to the project will given in the International Conference on Computational Science in Singapore. In August, a two week course " Density Functional Theory for Nanostructures" will be given in the 21st Jyväskylä Summer School. The hands-on exercises of the course will be done with GPAW, the course participants come from various European countries.

We are planning to participate in suitable physics/material science conference, and hopefully presenting also some results related to this PRACE 7.2 project.

5.4 EC-Earth 3

EC-Earth 3 is a scientific application in the field of Meteo-Climatology.

5.4.1 *Short description of the application and the planned work*

EC-Earth is a project, a consortium and a model system. The EC-Earth model is a state-of-the-art numerical earth system model (ESM) based on ECMWF's Seasonal Forecasting System. Currently, the EC-Earth consortium consists of 24 academic institutions and meteorological services from 11 countries in Europe. EC-EARTH Version 3, which is the most recent version, is a coupled ESM comprising ECMWF's atmospheric model IFS, including H-TESSEL land model, the general ocean circulation model NEMO, the LIM sea ice model, and the OASIS coupler.

The tasks to be performed are the following (partners involved are given in parantheses):

1. Make a performance analysis of the high-resolution configuration of the coupled EC-EARTH 3 on different platforms. (LiU/SARA)
2. Validate the OASIS4 coupler for the EC-Earth coupling configuration. (CERFACS)
3. assess the performance of and improve the coupling implementation, including the potential benefits and costs of upgrading to OASIS4. If beneficial, upgrade EC-EARTH 3 to OASIS 4. (CERFACS)
4. Perform the mapping of MPMD hybrid MPI+OpenMP threads and MPI-only tasks efficiently onto cores on different architectures. (SARA)
5. Investigate if CUDA-enabled routines can improve the performance and/or scalability of the coupled or atmosphere-only model. (ICHEC)
6. Investigate and improve scalability of I/O within IFS and NEMO. (NTNU)
7. Design of one application to run ensemble simulations. Monitoring and fault tolerance of such a mega-model (which, in itself, is desirable)must be addressed in the design. The final implementation of such an ensemble system is of secondary importance and is only addressed if time permits. (SMHI/IC3)
8. Optimise the mapping of MPI tasks of the sub-model NEMO (possibly also IFS) to cores on different architectures. (SARA)
9. Implement dynamical memory allocation and load-balanced domain decomposition. (STFC, not supported by PRACE)
10. look at scalability of reading of forcing file for ocean-only run. (SARA)

11. improve the efficiency of CDO for post processing of GRIB data, which is very I/O intensive. This is partly due to archiving, but also do to the “heavy” I/O (CDO).

The partners involved are:

SARA: National Academic Computing Centre, The Netherlands
ICHEC: High-performance Computing Centre, Ireland.
LiU : Linkoping University, Sweden.
NTNU: Norwegian University of Science and Technology, Norway.
SMHI: Swedish Meteorological and Hydrological Institute, Sweden.
CERFACS: European Centre for applied mathematical research, France
IPSL: Institut Pierre Simon Laplace, France.
KNMI: Royal Dutch Meteorological Institute, The Netherlands.
IC3: Catalonian institute for climate sciences, Spain.
STFC: Science and Technology Facilities Council

5.4.2 *Status and preliminary results*

EC-Earth has been ported and benchmarked on the Ekman cluster at PDC. The difficulty with benchmarking EC-Earth is that the right balance between the number of processors for the atmosphere (IFS model), the ocean (NEMO model) and the coupling (OASIS) needs to be found. In an ideal case, neither the atmosphere nor the ocean should be waiting for input from the other component. As it can be seen in Figure 5 the high resolution EC-Earth scales well up to 1500 MPI processes.
For this benchmark, the number of NEMO processes was kept constant at 320, while the number of IFS processes was increased. In the Figure, the ideal scaling is relative to the atmospheric component only.

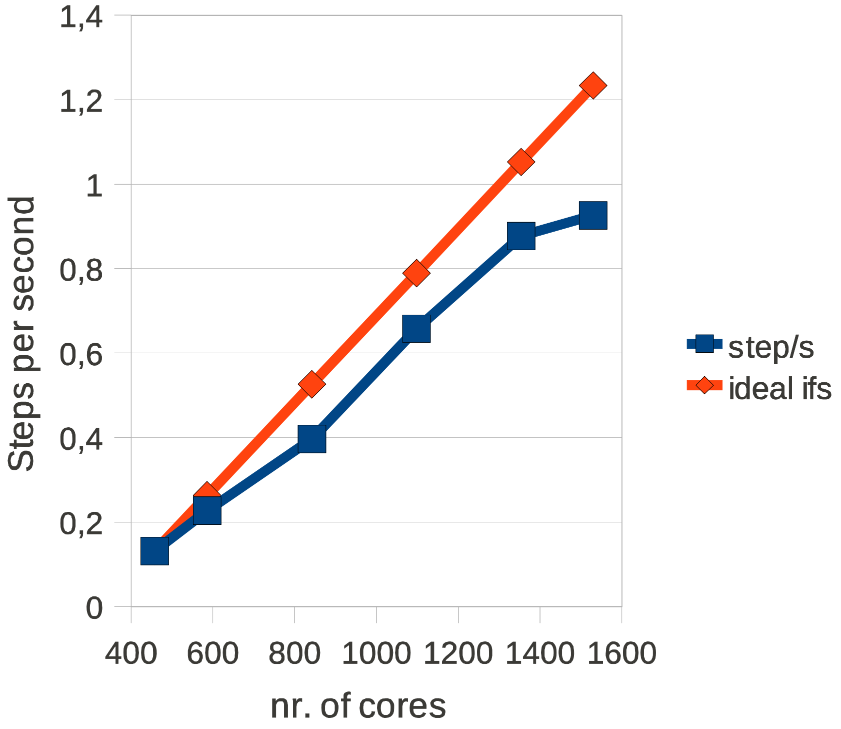

Figure 5: Benchmark of EC-EARTH, including full I/O

The causes for the drop-off at more than 1400 cores has not been well investigated yet. As has been shown in the PRACE PP project, NEMO can scale up to at least 1000 cores so a more accurate performance analysis of the EC-Earth model is required to investigate if it is possible to address further scalability.

The atmospheric I/O steps in EC-Earth are 2–3 times more expensive and scale less due to their serial nature than the computational steps, which clearly shows the need for I/O modifications. The actual impact of faster or slower file systems, as well as local disks for intermediate storage, needs to be investigated. This becomes especially urgent if the I/O frequency would be increased.

Porting of EC-Earth to other platforms, at the moment the IBM Power6 platform running both Linux and AIX, is still ongoing. Porting issues on these platforms have already been translated into modifications of the EC-Earth development branch.

The very high-resolution NEMO ocean-only configuration has been ported to the IBM Power platform and modifications identified in a study done in the PRACE1PP project will be implemented. This work is also relevant for Tasks 7.4, as NEMO is part of the benchmark suite, and 7.1, where the developers have applied for preparatory access to JUGENE.

Discussions have been ongoing about the way to implement ensemble simulations that are suitable for petascale machines. Two possible approaches have been compared:

Autosubmit is a suite of Python scripts that manages ensemble runs as a sequence of (inter-dependent) jobs. It handles multi-stage workflow (e.g. pre/post processing) and includes live monitoring. The disadvantages are that the target platform needs to support python and that it handles ensemble members as individual jobs for the queuing system. This may not be acceptable to the Scientific Steering Committee, that decides about PRACE allocations. A discussion at the NEMO User's meeting in Toulouse, end of June 2011, should clarify this point.

OASIS Ensemble Mode, OASIS handles separate ensemble members in a similar way to the pseudo-parallel mode. The advantage would be that this is a true one-application-approach, where all members are running in one MPI universe. Furthermore, it allows for the (partial) calculation of online ensembles, which would decrease the disk I/O and space requirements. Since this would be a radically different approach from what has been done before in the climate community, it means a higher risk of failure. Since no code has been written yet, it would also require more effort to implement.

At the moment, OASIS3 is being used to couple the atmospheric and oceanic component together. OASIS3 is only quasi-parallel, where an exchange variable cannot be split up and processed by multiple tasks. OASIS4 is fully parallel and is already coupling the oceanic component. Work is ongoing to couple OASIS4 to the atmospheric component, which should be finished before the end of 2011.

5.4.3 *Cooperation with the owners of the code, the Scientific Community or users*

PRACE has an active collaboration with the scientific community and the owners of the code. PRACE developers have access to the development portal that is used for EC-Earth 3 and can directly discuss with the other developers. Monthly teleconferences are used to keep track of the progress.

Preparatory access calls were submitted by the IS-ENES community to prepare ARPEGE/NEMO (a French climate model that is very close to EC-Earth3) and NEMO-only for petascale architectures.

A session at the NEMO User's Meeting at the end of June in Toulouse will be dedicated to EC-Earth developments.

5.5 Open FOAM

OpenFOAM is a scientific application in the field of Engineering and CFD.

5.5.1 *Short description of the application and the planned work*

The OpenFOAM (Open Field Operation and Manipulation) CFD Toolbox is a free, open source CFD software package. The core technology of OpenFOAM is a flexible set of efficient C++ modules. These are used to build a wealth of solvers, to simulate specific problems in engineering mechanics, utilities to perform pre- and post-processing tasks ranging from simple data manipulations to visualisation and mesh processing, and libraries to create toolboxes that are accessible to the solvers/utilities, such as libraries of physical models.

The scientific cases require running with a finer mesh to increase the accuracy of the simulations, therefore to achieve a better understanding of the phenomena and thus to make validation studies of new and more accurate models possible. To do this in a reasonable amount of computation time a significantly higher number of processing cores is needed. This requires more efficient parallelisation, leading to better scaling behaviour.

Enabling of OpenFOAM to a Tier-0 system advances both on the scientific problem and on the OpenFOAM software development itself. Previous work shows very good scalability of the simpleFOAM solver up to 2000 cores, further work is necessary to explore higher scalability of this solver and to analyse the performance and scalability of the most relevant solvers and to identify possible bottlenecks. Generally optimisation of MPI parallelisation to achieve better scaling will be required and issues related to communication in collect operations will be analysed.

The partners involved are CINECA, EPSRC (EPCC), GCS (LRZ and HLRS) and NUI Galway (ICHEC).

The first activity was the selection of test cases and currently two scientific test cases have been selected and analysed:

- At CINECA a model from the Internal Combustion Engine Group (ICE Group) at the Energy Department of Politecnico di Milano,
- At ICHEC a model of the flow around a vessel to improve the estimates of air-sea fluxes from ships, taken from the Centre for Climate and Air Pollution studies at NUI Galway.

In addition, a 3D version of the OpenFOAM's 2D cavity flow tutorial and a 3D version of OpenFOAM's 2D Dam Break tutorial have been prepared for tests by EPCC.

The next steps will be to identify the bottlenecks and to select modules to be optimised.

The directory structure for input/output files for OpenFOAM as described in the OpenFOAM User's Guide, may limit the scalability in event of several thousand processes accessing such a structure. A profiling of the test cases with the I/O-profiling tool Darshan will show how I/O is performed and can give indication on whether the I/O structure can be a limiting factor. The activity must also assess other I/O strategies for simulations with very large CPU-core counts. Implementing the use of SIONlib or ADIOS are paths that could also be considered.

5.5.2 *Status and preliminary results*

LRZ has contacted the community at LRZ for collaboration in petascaling tasks. It was given the opportunity to use one of the user cases as a starting point. They tried to generate a very large mesh to be used at high core counts. Unfortunately, they have been so far unable to achieve the required mesh quality and have dropped these tasks for the time being.

At the same time, at LRZ, a complete scalability test was performed on one of the OpenFOAM solvers for a current CFD problem, Large Eddy Simulations. The purpose of the tests was first to determinate the scalability limits of OpenFOAM and secondly to collect enough data to investigate parallelism issues and MPI implementations.

A very effective tool for profiling the parallel applications, IPM (http://ipm-hpc.sourceforge.net), was tested, because tools such as Scalasca, Vampir and PTP failed to collect useful data. As a result, it can be shown that for OpenFOAM there are many aspects of the code which could be optimised. Some examples include the multi-grid family of solvers for partial differential equations, the I/O footprint or the MPI strategies for parallelism.

At CINECA OpenFOAM 1.7.1 was ported on the new CINECA IBM PLX cluster (3312 core Xeon E5645 with 528 GPU nVIDIA Tesla M2050). Some efforts were initially made to better understand the OpenFOAM structure and then the activity continued with running the test cases, profiling to identify bottlenecks and make a performance comparison. The scientific test case was taken from the Politecnico di Milano and currently tests are being performed on it.

EPCC has produced a report describing the installation of OpenFOAM on the Cray XT4 part of HECToR, which includes profiling OpenFOAM for three large test cases. Two of these test cases, a 3D version of the OpenFOAM's 2D Cavity Flow tutorial and a 3D version of OpenFOAM's 2D Dam Break tutorial, have been made available to the group. The third test case belongs to an industrial partner.EPCC has ensured that the existing connections within OpenCFD Ltd, the authors of OpenFOAM, remain strong. OpenFOAM 1.7.1 and 1.6 will be made available on the Cray XE6 part of HECToR in the near future.

ICHEC compiled and built OpenFOAM 1.7.1 using GNU compilers on the SGI Altix ICE 8200EX machine. Profiling the code with Intel Trace and additional profilers for single process CPU (e.g. Vtune, IPM, gprof) are ongoing. Initial benchmarks have been completed for a test case, modelling turbulent airflow around a vessel with LES, using varying mesh sizes and time steps. The test case is taken from the Centre for Climate and Air Pollution studies at NUI Galway. Preliminary results show that I/O is a bottleneck and a full performance analysis is underway.

At HLRS the coordination for the work on OpenFOAM and the administrative work for planning and reporting has been done. This includes regular phone conferences to discuss the status of work, individual phone talks and e-mail exchange.

Intensive discussions with the OpenFOAM community at University of Stuttgart were performed, this included the organisation of a local OpenFOAM user workshop.

OpenFOAM versions 1.6-ext and 1.7.1 were compiled and tested at the test system for the upcoming Tier-0 system (Cray XE6) at HLRS. A test case, modelling a complete Francis turbine, from the Institute of Fluid Mechanics and Hydraulic Machinery of the University Stuttgart was analysed on the Cray XE6. As this test case requires OpenFOAM extensions it has not been selected for further analysis within WP 7.2 for the time being.

5.5.3 *Cooperation with the owners of the code, the Scientific Community or users*

Hrvoje Jasack, OpenFOAM Chief developer at WIKKI (http://www.wikki.co.uk/), was invited as a speaker at the Third PRACE Industrial Seminar. It was an important occasion to get in touch with him who is very interested in the work we are doing on OpenFOAM. He is interested in discussing our ideas and we are attempting to improve our collaboration. A further occasion could be the "hands on" school held in Zagreb every September, in which a member of our group could participate.

We have established an initial contact with the VKI (http://www.vki.ac.be/), the biggest European institute research in the field of Fluid Dynamics. They have spent quite some efforts in investigating the OpenFOAM capabilities in a number of fields, Aero acoustics being one of those. They are interested in the migration of OpenFOAM to a GPU/CPU cluster.

The scientific test cases were selected by proposals of users of OpenFOAM, each in strong collaboration with one of the PRACE partners involved in OpenFOAM. Discussions on experiences and requirements were done with these users of OpenFOAM.

6 Conclusions

During the first year of activity, Task 7.2 worked to fulfil its objectives, firstly contacting the scientific communities and selecting the application codes of interest, then starting to work with these communities, in a long-term cooperation, to petascale the codes selected.

So far, eleven application codes, in the main domains of computational sciences have been selected. For five of these applications the enabling work started in December 2010 and some preliminary results have been already achieved. For the remaining six applications the work started in May 2011, after the second selection process and the agreements with the scientific communities on the work plan to implement. The activity on the first five application codes is scheduled to conclude around December 2011. In this way the petascaled codes will allow the related scientific communities to apply to the next regular calls for access to Tier-0 resources. The work on the other six codes will run until the end of PRACE-1IP.

With respect to structured communities, cooperation has been established with the IS-ENES community, with the long-term objective to run the six European climate models, of interest for this community, on PRACE Tier-0 systems. Currently the enabling activity is proceeding with the suite EC-Earth 3 which includes three complex application codes (IFS, NEMO and Oasis) that are part of the six European models. The enabling activity is making good progress and is expected to consolidate a strong cooperation between PRACE and IS-ENES.

Further collaborations with other structured communities will be addressed in the next few months, and other application codes of key interest for these communities will be investigated to petascale. Contacts will be established with the ScalaLife Project in the field of Life Science, with the Astrophysics community and with the High Energy Physics community. In addition, the collaboration with the MAPPER project will improve to petascale one or more multiscale applications of interest for this community.

In summary, the activity done so far has demonstrated the utility of working closely with the scientific communities and the code developers. We have been quite impressed with the progress which has been made in the initial stages of the optimisation and petascaling of many applications, progress which we believe would not have been possible without the close collaboration of the communities and developers involved. This activity will be continued in the remaining period of PRACE 1IP to fully complete the petascaling process and to make the applications available to enhance science by means of Tier-0 systems. Furthermore, the activity will be extended to new communities and applications in order to enlarge the spectrum of scientific codes on PRACE systems. This approach is clearly successful and could be used as a model within PRACE for exploiting the Tier-0 facilities and the Tier-1 resources which will be available in late 2011.

7 Annex I: Selection Form

Form used to select applications and the related scientific communities

PRACE 1IP WP7 – Task 7.2a

Please provide the following information for each of the communities/applications you wish to propose…

Community/Application

Scientific field	
Macro area	(e.g. physics)
Specific area	(e.g. cosmology, galaxy clusters, galaxy dynamics)
Application name – version - License	(e.g. Enzo)
Application web site	http://

Short description of the involved communities:
In this part we expect to collect information about:

- The group that has developed the proposed application
- The main research groups worldwide that use the proposed application
- The relationship between the proponents and the developers team
- The relationship between the proponents and one or more of the research group adopting the application
- The degree of involvement of the developers team in the present project
- The degree of involvement of one or more scientific users in the present project

Main publication related to the proposed application
In this section you have to provide a list of the 5/10 most important publications related to the proposed application

Confidentiality
Is any part of the project covered by confidentiality? If YES, please give the reasons for confidentiality:

Short description of the application (algorithms, I/O, parallelisation strategy, current performances, performances bottlenecks)

The case for Petascale
Describe

- Why the code should be enabled to petascale (scientific cases, size of the community who would exploit the petascale version…)
- The work to be accomplished on the code to be petascaled
- Possible risks of failure and contingency plan
- Expected effort (PM) and who will do what

Project contacts (Max. 3):

Name1	Role/Institution1	e-mail1
Name2	Role/Institution2	e-mail2
Name3	Role/Institution3	e-mail3

Any other relevant comment:

8 Annex II Application Codes Selected in Task 7.2

A quick abstract of the eleven application codes of interest for the scientific communities selected to petascale in Task 7.2 is presented in this Annex. This section has been included, from one side, to synthesise in a common vision all the different scientific disciplines and methodologies covered by the applications, and, from the other, to have a more concrete idea of the importance of the codes and the complexity of the work to accomplish.

8.1 GROMACS

GROMACS is one of the world's most widely used packages for classical particle Molecular Dynamics, mainly used for biomolecular simulation and modelling. The code was originally developed at Groningen University. Presently the developer community consists of a large international team effort led by Prof. Erik Lindahl at the KTH - Royal Institute of Technology in Stockholm. Additional main development labs are located at Uppsala University (Sweden); Max-Planck-Institut in Göttingen (Germany); Stanford University, University of Virginia, and Oak Ridge National Labs (US). SNIC-KTH is supporting the development activities within GROMACS.

GROMACS implements highly-optimised and dynamically load balanced 3D domain decomposition algorithms with a multiple-program multiple-data approach for the long-range interactions. The computational kernels are written in assembly for maximum performance. On multi-core nodes GROMACS uses thread-based parallelization to automatically spawn the optimum number of threads, while MPI is only required for parallelisation over the network. GROMACS will soon have support for collective I/O when running on large clusters.

8.1.1 *Case for Petascale*

Historically, the primary aim for GROMACS has been efficiency and performance rather than relative scaling, but over the last couple of years there have been major efforts at combining this with better scaling. We believe this makes the package an exceptionally important strategic target for PRACE, since the combination of said efficiency with petascaling will truly enable calculations that previously have not been possible for scientists to accomplish. It is also a truly European code (with all three main developers in the EU), and the free software aspect means work can be reused in other codes.

Presently, GROMACS has already been shown to scale to 150,000 cores on Jaguar at Oak Ridge in the US. However, this required (a) systems with 100M particles (~650 atoms/core), and (b) the use of reaction-field interactions rather than the PME (particle-mesh Ewald) alternative typically preferred for biomolecular systems.

The initial goal for the PRACE project would be to scale individual simulations to over 8,000 cores even for relatively small systems in the 1-2M atom range. This will require an improvement of the strong scaling down to ~250atoms/node, which should be relatively straightforward by combining a better load balancing with and an efficient hybrid communication scheme MPI with threads/OpenMP on nodes, rather than using only MPI. The GROMACS developers are also working actively on improving the PME lattice summation algorithms to reach the same scaling level.

The ultimate goal within PRACE will be to develop a separate new parallelisation layer that is designed from the scratch to execute on 10 million cores, and then combine it with the current acceleration present in GROMACS. For larger systems (~10M atoms), this should enable the codebase to reach much higher scaling than the minimum PRACE threshold.

Smaller systems (10-100k atoms) will benefit equally from these scaling improvements, and these to will be able to surpass the PRACE threshold through the use of replica-exchange and similar sampling enhancement techniques.

8.2 Quantum Espresso

Quantum ESPRESSO is an integrated suite of codes for electronic-structure calculations and materials modelling at the nanoscale. It is based on density-functional theory, plane waves, and pseudopotentials (both norm-conserving and ultrasoft).

Quantum Espresso is an initiative of the DEMOCRITOS National Simulation Centre (Trieste) and SISSA (Trieste), in collaboration with the CINECA National Supercomputing Centre in Bologna, Ecole Polytechnique Fédérale de Lausanne, Université Pierre et Marie Curie, Princeton University, Massachusetts Institute of Technology, and Oxford University.

8.2.1 *Case for Petascale*

ICHEC and University of Oslo aim within this project to enable the Quantum Espresso package to perform the linear response along with NMR calculation through PWscf and GIPAW codes at large scale. This functionality has been requested by the community and they have provided two cases to study where the current limit of the application scalability is an issue. The PWscf code has shown good scalability within thousands of processors as reported by PRACE benchmarking activity and some DECI projects, however GIPAW and the general linear response codes still show poor scalability at that scale. Our proposal aims to address this issue by improving the data distribution and introducing hybrid parallelism for GIPAW following the schema currently implemented in PWscf. The scalability of the GIPAW application needs to be enhanced to allow the NMR calculation to meet the minimum scaling requirements to run on a PRACE system (4K/8K cores depending on the infrastructure).

Regarding the PWscf code we propose to focus on both the parallelisation of the hybrid functionals and Van der Waals approach to improve the accuracy of the SCF calculation. The Van der Waals approach is currently implemented using a parallel FFT method which is well known to cause a bottleneck at large scale so we propose to assess the implementation of this approach in real space.

One of the most important results from NMR experiments is to provide a very detailed inspection of the atomic scale structure and dynamics in the vicinity of the observed nuclei in ordered or disordered materials.

The methodology presented above will be carried out over two real scientific cases briefly here described.

1. Determining the cholesterol crystal structures in human gallstones from first principles. For studying this system we need to sample the phase space of cholesterol crystals using an experimental NMR spectra as a guideline, calculate the ab initio NMR chemical shifts of the candidate structures and compare them to the experimental results. The first step will be achieved by performing metadynamics simulations within LAMMPS in order to generate trial configurations and then relax the structure using ab initio methods. For the second stage we will use the Gauge-Invariant Projector Augmented Waves (GIPAW) method. Both simulations will be carried out in the framework of plane-wave density functional theory (DFT) and using the pseudopotential approach. Since the NMR spectrum is very sensitive to electronic structure, we will need to achieve a realistic description of the atomic and electronic

structure. This will be done by incorporating the Van der Waals interaction at the ab-initio level.

2. The most important applications of Zeolites are in acid catalysis. The acidity arises as some of the silicon atoms are substituted by an aluminium-proton pair. While the opportunities for tailoring the chemical and physical properties of zeolites are enormous, zeolite synthesis is still a highly empirical undertaking. Characterisation of the aluminium distribution cannot be done with XRD, since Al and Si have so similar nuclear charges. Solid-state NMR shows promise, but assigning the spectra has proven difficult. Access to computed spectra would allow great progress in the characterisation of these materials. Useful results would require the reliable computation of Al and Si NMR spectra for periodic systems with P1 symmetry, cell dimensions of several tens of Ångströms, and thousands of atoms in the unit cell.

8.3 GPAW

GPAW implements density-functional theory (DFT) using the projector augmented wave (PAW) method. The discretisation of the equations can be done with two complementary approaches, either with a finite-difference method using uniform real-space grids or with localised atomic orbital basis functions. The real-space approach is more relevant for the petascale applications. The PAW approximation offers good description over the whole periodic table of elements, and especially for first row and transition metal elements it provides typically better description than the more conventional norm-conserving or ultrasoft pseudopotential approaches.

Also, the real-space discretisation offers good parallelisation possibilities. To our knowledge, GPAW is currently the only publicly available code implementing the PAW method with real-space grids. In addition to standard density-functional theory, GPAW implements also the time-dependent density functional theory which can be used for studies of excited state properties such as optical spectra, which are beyond the realm of standard DFT. Two different forms of time-dependent density functional are implemented, the realtime formalism (RT-TDDFT) and the linear response formalism (LR-TDDFT).

8.3.1 *Case for Petascale*

Petascaling of GPAW enables standard DFT calculations for larger system sizes. Especially studies of biomolecules, new electronic energy storage and catalysis would benefit from possibility of larger scale calculations. For time-dependent DFT calculations the excellent scalability prospects should allow relatively easy utilisation of tens of thousands of processes. TDDFT calculations, either in the real-time or in the linear response mode are much more time consuming than the standard ground state calculations, and currently a single run takes easily several days. Petascale computers would bring the simulation times down from days to few hours.

The petascaling activity will address the initialisation task, the distribution of Ω-matrix for LR-TDDFT calculations, Hybrid MPI/OpenMP implementation and matrix diagonalisation methods alternative to ScaLAPACK. Neglecting linear scaling DFT calculations (which cannot be applied for example to metallic systems), GPAW is in our opinion one of the most promising DFT code packages for petascale.

Some specific scientific problems where petascale calculations would be useful are: Optical properties of Au clusters, optical properties of Si nanocrystals, surface catalysis, etc.

8.4 EC Earth 3

The activity in EC-Earth 3 has been agreed upon with the e-INES community and involves three main codes in EC-EARTH, one for atmosphere (IFS/ Harmonie), one for the ocean (NEMO 3.3) and a coupler (OASIS3/4).

The ECMWF's IFS code is a parallel spectral weather model that is also used for seasonal climate prediction. Its structure is similar to climate codes from NCAR, including CCM. It uses domain decomposition in two-dimensions and performs both spectral and Legendre transformations on the grid data. Furthermore, it includes many state-of-the-art physical parametrisations that are adjusted to scale with the resolution. The file format for I/O is GRIB, which is an international standard defined by the WMO. There are developments of a parallel I/O library for GRIB outside

ECMWF, but as of yet the I/O of IFS/Harmonie is serial. This is a serious bottleneck for scaling up this application. Since GRIB is a used worldwide, (preferrably open-source) solutions would benefit the whole meteorological community.

The NEMO model consists of several components, all based around the OPA physical ocean model. It uses MPI with a regular domain decomposition and finite differences. It needs to calculate the free surface height, which is now done using a less-than-ideal conjugate gradient or overrelaxation method, which limits the scalability. A new method is implemented that calculates the free surface using an explicit method, which does not have the scalability issues of the elliptic solvers. I/O is done using the IOIPSL library, which now uses parallel I/O using NetCDF4. Detailed experiments have been conducted to achieve a good output performance using e.g. chunking. Boundary conditions are read from file at the moment that they are needed, but this causes huge delays at scale on systems like BlueGene. This reading of input can and should be done beforehand, as long as the memory permits it. It can run both in a realistic topography and forcing, or an idealised setup. The domain decomposition strategy should be adapted to the depth and coastline in case of a realistic topography, which is work done by STFC.

8.4.1 *Case for Petascale*

The main actions planned are:

- Improve the free surface solver, decrease the number of iterations, and the communication aspects.
- Investigate the new I/O server in NEMO and possibly improve the implementation.
- Optimise the mapping of MPI tasks to cores on different architectures.
- Implement dynamical memory allocation and load-balanced domain decomposition.
- Improve the scalability of the forcing files for the ocean-only run.

The difficulty of scaling EC-Earth is in the coupling of both models, which is done using three separate applications (IFS, NEMO, OASIS3/4) that interact through MPI.

Other actions are related to the following points:

- Make a performance analysis of EC-Earth T799+0.25 on different platforms. It has been run only on a local cluster so far and better understand the bottlenecks (I/O, communications, load-balancing) in the coupler.
- Get OASIS4 working with EC-Earth.
- Mapping of MPMD hybrid MPI+OpenMP threads and MPI-only tasks efficiently onto cores on different architectures.

- Investigate if CUDA-enabled routines can improve the scalability of the coupled or atmosphere-only model.
- Investigate and improve scalability of I/O.

8.5 OpenFOM

The OpenFOAM® (Open Field Operation and Manipulation) CFD Toolbox is a free, open source CFD software package produced by a commercial company, OpenCFD Ltd. It has a large user base across most areas of engineering and science, from both commercial and academic organisations. OpenFOAM has an extensive range of features to solve anything from complex fluid flows involving chemical reactions, turbulence and heat transfer, to solid dynamics and electromagnetics. However, considerable community effort has contributed to further development of OpenFOAM concerning specific areas such as Turbo machinery, e.g. GGI interfaces etc. The core technology of OpenFOAM is a flexible set of efficient C++ modules.

The relation of the proponents with the developer community is based on the fact that there is a significant number of OpenFOAM users at several sites involved in PRACE. Moreover, the proponents are in touch both with OpenCFD Ltd. and Wikki GmbH consulting companies and have already started to discuss collaboration with them. Collaboration with these consulting companies would be through support.

8.5.1 *Case for Petascale*

OpenFOAM has a large user community and a wide range of applications, relevant scientific cases like Large Eddy Simulation (LES), combustion and Direct Numerical Simulation (DNS) rely on OpenFOAM. Its enabling to a Tier-0 system advances both the scientific problem and the OpenFOAM software development itself. Previous work shows very good scalability of the simpleFOAM solver up to 2000 cores, further work is necessary to explore higher scalability of this solver and to analyse the performance and scalability of the most relevant solvers and to identify possible bottlenecks. Generally optimisation of MPI parallelisation to achieve better scaling will be required and issues related to communication in collect operations will be analysed.

8.6 DL-POLY

DL_POLY is a general purpose classical Molecular Dynamics simulation software developed at Daresbury Laboratory by I.T. Todorov and W. Smith. The DL_POLY project was originally conceived in 1993 by William Smith at the Molecular Simulation Group at Daresbury Laboratory under the auspices of the Engineering and Physical Sciences Research Council (EPSRC) for the EPSRC's Collaborative Computational Project for the Computer Simulation of Condensed Phases (CCP5). DL_POLY is a community code and CCP5 is the body that recommends new feature developments.

8.6.1 *Case for Petascale*

As the availability of higher core counts increases, the opportunity to look at micro-scale problems as evolutions of fracture and defect dynamics on an atomistic level becomes real. The research benefitting most from this is modelling materials behaviour under irradiation. It is at utmost importance for predicting time to failure of components of PWR nuclear reactors as well as for understanding, predicting and designing metal alloys for use in fusion reactors and ceramic matrices for encapsulation of nuclear waste.

DL_POLY is routinely used as a modelling tool in a variety of scientific domains. We expect that the proposed work will have significant outcomes, not only to this community of modellers, but also for the whole scientific HPC community. The proposed work is to optimise the performance of the currently available hybrid version of the DL_POLY_4 code (MPI + OpenMP + CUDA). Moreover, not only we do intend to improve the multithreading CPUs capacity of the code but also introduce OpenCL alternative along the CUDA code and thus extend the test range for the performance analysis which will indicate whether this is a beneficial future direction.

The work to be accomplished on the code to be petascaled mainly focus on the following points:

1. Introduction of OpenMP into the MPI domain decomposition framework.
2. Testing and evaluation of libraries for:
   - non-equi-spatial DD load balancing
   - long-range electrostatic evaluations from the Simulation Lab "Molecular Systems" at FZ-Julich.
3. Testing, optimisation and further porting of algorithms on GPU enabled PRACE petascale systems.

8.7 CP2K

CP2K has been developed over 10 years by a group of around 20 developers, including the research groups of Prof. Jürg Hutter and Dr. Joost VandeVondele of the Physical Chemistry Institute at the University of Zurich.

CP2K is a freely available (GPL) package, written in Fortran 95, to perform atomistic and molecular simulations of solid state, liquid, molecular and biological systems. It provides a general framework for different methods such as DFT using a mixed Gaussian and plane waves approach (GPW), and classical pair and many-body potentials.

From an algorithmic perspective, the code makes use of a number of distributed data structures, including regular grids (which are transformed by 3D FFT in the process of shifting from one basis representation to the other) and distributed block-structured sparse matrices (handled by the highly scalable Distributed Block Compressed Sparse Row – DBCSR – library). Small systems of up to 1000 atoms are typically limited by the scalability of the FFT and current large systems are fully dominated by sparse matrix operations, and development of the DBCSR library is key to continued improving scalability.

8.7.1 *Case for Petascale*

Scaling CP2K calculations to the tens or hundreds of thousands of cores on Petascale systems requires a mixed-mode MPI/OpenMP approach, where MPI is used for inter-node communication and OpenMP divides work between the available cores within the node.

Scalability depends on the problem size and level of theory used, however, typical systems of ~1000 atoms studied using GGA or similar functionals have been shown to scale up to 10,000 cores on the HECToR2, and a small system of 216 atoms using Hartree-Fock theory was shown to scale to 65,536 cores on JaguarPF3.

Nevertheless, usage of the mixed-mode code is still low compared with the MPI-only version in the user community. In order to enable CP2K users to take full advantage of mixed-mode parallelism, we propose the following workplan:

1. Extending OpenMP to support the full range of exchange-correlation functionals in CP2K – currently less than half of the ~20 functionals are OpenMP-enabled.
2. Improve the existing OpenMP parallelism of the realspace grid functionality. At present there are known inefficiencies in the collocation of Gaussian basis functions and distribution of the density matrix, which cause OpenMP to be effectively limited to within a NUMA region, rather than across an entire node
3. Parallelise the calculation of the core Hamiltonian integral matrix using OpenMP. This part of the code does not make use of threads currently and can therefore become a bottleneck in using the highly accurate MOLOPT basis sets or basis sets with large angular momentum. This parallelisation builds on the work done in WP 7.1 to implement the QS neighbour lists as arrays instead of linked lists.
4. Porting and verification: at present the OpenMP in CP2K is known to work well with recent gfortran compilers (4.4+). However, in order to increase the utility of the code to the wider user community, we will verify the correctness of mixed-mode code under a range of popular and important compilers. We will also ensure that CP2K works correctly and efficiently on the PRACE Curie system using OpenMP across the 32-way nodes.
5. Benchmarking and dissemination: In order to raise the profile of the mixed-mode implementation in CP2K with the wider user community, we will gather several representative test cases from known CP2K user groups, and demonstrate the performance gains achievable using OpenMP in addition to pure MPI.

8.8 Code_Saturne

Code_Saturne is a multi-purpose Computational Fluid Dynamics (CFD) software, which has been developed by Électricité de France Recherche et Développement EDF-R&D since 1997. The code was originally designed for industrial applications and research activities in several fields related to energy production; typical examples include nuclear power thermal-hydraulics, gas and coal combustion, turbo machinery, heating, ventilation, and air conditioning. Code_Saturne has been released as open source in 2007 and is distributed under a GPL licence.

8.8.1 *Case for Petascale*

Many scientific cases of current interest calculated using Code_Saturne require larger, finer grids in order to simulate large-scale complex industrial flows, often involving areas of turbulence. Current examples of interest are flows around submarines, aircraft and water cooled nuclear reactors. In order to model these flows accurately meshes of the order of a few millions to even billion of cells have to be generated. It is expected that all the research organisations listed above would take advantage of the petascale version if access to petascale machines were made available to them.

Mesh generation of billion of cell meshes has to be parallel, but no open-source parallel mesh generators are yet available, as far as the proponents are aware. Therefore other routes must be followed and the one considered here deals with parallel global mesh refinement (or mesh multiplication), i.e. an initial mesh of about 100 million cells would be read by Code_Saturne, then each of its cells would be split (into 4 if initial tetrahedral cells, into more cells if initial polyhedral ones). This process could be repeated several times in order for the Navier-Stokes solver to run on a several billion cell mesh, while post-processing would be carried out on the initial 100 million cell mesh.

The work is thus split into three steps:

- Generation of a demonstration mesh of about 2 billion (2B) cells to be used for partitioning testing (see second step). This mesh is generated by joining several meshes of about 100M cells each,
- Some partitioners already exist in *Code_Saturne*®, but have not yet been tested for 2B cell meshes, where the 64-bit barrier is broken (connectivity index might be larger than 2B). The first part of the work will be to test the existing partitioner at such a scale. Following this a new partitioner will be implemented. Tests will be carried out firstly on Jugene to investigate the behaviour of the code on more than 60,000 cores, and then on Curie to assess *Code_Saturne*® performance depending on the partitioner used on a smaller amount of cores, but more powerful ones than the powerpc's.
- Parallel global mesh refinement (or mesh multiplication) will be implemented in *Code_Saturne*® to circumvent the mesh generation issue.

8.9 DALTON

The DALTON program package is a long-term Scandinavian collaboration in development of versatile Quantum Chemistry methods for the evaluation of various molecular properties. During the years DALTON developer communities have expanded and currently involve research groups from Denmark, Norway, Sweden, Italy, Germany and Spain.

The DALTON program package from version 3.0 (preliminary name DALTON 2011, scheduled release date end of January 2011) consists of two main modules: DALTON and LS-DALTON. The first module is designed to perform electronic structure calculations using HF, DFT, MCSCF, CI and CC methods for small and medium size molecular systems and the second module is designed to carry out HF and DFT calculations for large molecular systems within a linear scaling regime. For extremely large molecular systems (beyond several hundreds of atoms), where pure quantum mechanical description of the system is too expensive, the DALTON program package adopts so-called hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) approaches in order to carry out electronic structure and molecular property calculations. Currently, the QM/MM methods (principal developer Dr. J. Kongsted, SDU) are integrated into the DALTON module and work of porting these methods to LS-DALTON is currently ongoing.

8.9.1 *Case for Petascale*

DFT/MM methods in DALTON currently scale relatively well up to more than 500 CPUs. To increase scalability for petascale systems we aim to parallelise three main steps in hybrid DFT/MM calculations of optical and magnetic properties:

1. on-fly creation of polarisable force fields for inhomogeneous systems, like cellular membranes or proteins;
2. “domain aware” construction of MM region contribution to the Kohn-Sham matrix;
3. “domain aware” evaluation of MM region contribution to response vectors of various orders. All the above outlined algorithms will be parallelised using the domain decomposition approach to the whole system in which it will be

decomposed into smallest possible units, like amino acids residues. Suitable computational techniques will be applied to obtain desirable contributions to quantities relevant for the different steps, for example in the second and third step we will use a modified version of a fast-multipole method to compute the electrostatic Coulomb contribution in a linear scaling fashion.

In order to achieve the outlined performance targets the algorithms used to evaluate the MM contributions to the Kohn-Sham and response matrices must be made “domain aware” and new strategy for evaluation of these contributions must be implemented, which is based on the modified fast-multipole method as well as sparse data storage must be integrated in all algorithms.

Other points to be addressed for petascaling are:

- Hybrid implementation (Open MP and MPI) for LS-Dalton.
- Storage and I/O: Distributed in-core storage of matrices and recomputation of integrals will allow minimal to no disk usage.
- Algorithms: Matrix-multiplication based wave-function and linear-response optimisation and on-the-fly integral evaluation.

8.10 SPECFEM3D

The software package SPECFEM3D_GLOBE simulates seismic wave propagation at the scale of the whole earth-based upon the spectral-element method (SEM). The SEM is a continuous Galerkin technique, which can easily be made discontinous. SPECFEM3D is written in Fortran90 with full portability in mind, and uses MPI with domain decomposition for parallel processing.

The SPECFEM3D_GLOBE package was first developed by Dimitri Komatitsch and Jeroen Tromp at Harvard University and Caltech, USA, starting in 1998, based in part on earlier work by Dimitri Komatitsch and Jean-Pierre Vilotte at Institut de Physique du Globe (IPGP) in Paris, France from 1995 to 1997. Since then, widely distributed development team has developed it. SPECFEM3D is used throughout the world and is also used by a broad European community. Several European projects (EPOS, Verce) have shown their interest to use SPECFEM3D. Also the oil and gas industry uses these simulations for their explorations.

8.10.1 *Case for Petascale*

SPECFEM3D has two modes of operation - it can perform either global (Earth-scale) or local (continental-scale) simulations. The petascaling activity in this project would focus mainly on the global earth version, which is called SPECFEM3D_GLOBE.

The activity will be done in two main directions:

1. Investigate and design a mixed MPI + OpenMP implementation to create a hybrid application that petascales efficiently: for instance to take advantage of the “fat node” structure of some of the nodes of the Curie / PRACE machine at TGCC in France.
2. Implement a 3D domain decomposition in the case of more complex meshes and meshing issues.

8.11 EUTERPE

EUTERPE is a particle-in-cell (PIC) gyrokinetic code for global linear and non-linear simulations of fusion plasma instabilities in three-dimensional geometries, in particular in tokamaks and stellarators.

EUTERPE solves the non-linear gyrokinetic equations by discretising the distribution function using markers. Each marker contributes a part of the distribution function whose evolution is given by the gyrokinetic Vlasov equation. Perturbation theory is employed in the calculation of the distribution function, thus reducing the discretisation noise and the computational resources needed. Both the magnetic and the electric fields are taken into account in the equations of motion of the markers. The magnetic field structure is assumed to be fixed throughout a simulation and is computed using the Variational Moments Equilibrium Code (VMEC), which is the standard code for calculating three-dimensional equilibria.

EUTERPE, like other gyrokinetic codes, is at the forefront of plasma simulations and requires a huge amount of computational resources. EUTERPE was created at Centre de Recherches en Physique des Plasmas (CRPP) in Lausanne as a global linear particle-in-cell code. It has subsequently been further developed at Max-Planck-Institut für Plasmaphysik (IPP) and has been adapted to different computing platforms.

The Fusion Theory unit from Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) collaborates with IPP and Barcelona Supercomputing Center (BSC) on the development and exploitation of this code.

8.11.1 *Case for Petascale*

Simulations of plasma micro-instabilities are a necessary complement for stellarator and Tokamak experiments such as Wendelstein 7-X and ITER. Especially important are full torus simulations for three-dimensional stellarator configurations. Gyrokinetics as a first principle based theory is well suited to describe the relevant physics. An established and flexible method for solving the gyrokinetic system of equations is the simulation via the particle-in-cell (PIC) Monte-Carlo method.

For this purpose the EUTERPE code has been developed which originally solved the electrostatic gyrokinetic equation globally in arbitrary three-dimensional geometry. The full kinetic treatment of the electrons and the inclusion of magnetic field perturbations extends the scope of applicability to the magnetohydrodynamic (MHD) regime and will make EUTERPE the first gyrokinetic code worldwide that is able to simulate global electromagnetic instabilities in three dimensions.

In all the simulations, especially for the non-linear ones, the amount of computational resources that a global three-dimensional PIC code requires for typical simulations is huge. A single simulation of turbulence transport of several hundreds of microseconds in a cylindrical domain could take 100.000 CPU hours. An extension of the simulation time and domain could easily become two orders of magnitude higher than this. The petaflops systems are a critical and necessary infrastructure in this scenario.

The work that will be accomplished in this proposal can be divided into the following tasks:

1. The parallelisation of the code has been extended by means of introducing OpenMP in the most time-consuming routines in the code and developing a new hybrid solver (mixing MPI and OpenMPI) to solve the quasi-neutrality equation. What remains is to perform a study to improve the scalability of this new solver.

2. Porting EUTERPE to GPU architecture. The information collected in the previous development of the hybrid version will be used to port the code to this heterogeneous architecture.
3. Analysis of the parameters which influence the performances (e.g. memory accesses), and the quality of the solution (e.g. improving the noise filtering), when the execution uses hundreds of thousands of processing elements.
4. Identification of new bottlenecks due to the execution of the code in a heterogeneous architecture. EUTERPE reads/produces a great amount of data from data storage. Therefore, an estimation of the I/O performance will be carried out using ZOID one of the most relevant and scalable parallel I/O libraries.