

CONTRACT NUMBER 508830

**DEISA**  
**DISTRIBUTED EUROPEAN INFRASTRUCTURE FOR  
SUPERCOMPUTING APPLICATIONS**

**European Community Sixth Framework Programme**  
**RESEARCH INFRASTRUCTURES**  
Integrated Infrastructure Initiative

**JRA1: Grid-enabling of selected Materials Science  
Simulation Codes**

Deliverable ID: DEISA-D-JRA1-1

**Due date : Oct, 31, 2004**

**Actual delivery date: November 18, 2005**

**Lead contractor for this deliverable: RZG, Germany**

**Project start date : May 1<sup>st</sup>, 2004**

**Duration: 5 years**

<b>Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)</b>		
<b>Dissemination Level</b>		
<b>PU</b>	Public	
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	<b>X</b>
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

---

## Table of Contents

Table of Contents .....	1
1. Introduction .....	2
1.1 Executive Summary.....	2
1.2 References and Applicable Documents .....	2
1.3 List of Acronyms and Abbreviations .....	3
2. General remarks .....	5
3. Grid-enabling of selected materials science codes.....	5
3.1 Selected Materials Science Codes.....	5
3.1.1 CPMD.....	5
3.1.1.1 Introduction .....	5
3.1.1.2 CPMD Code version and licensing for DEISA .....	6
3.1.2 WIEN2k .....	6
3.1.2.1 Introduction .....	6
3.1.2.2 WIEN2k Code Version and Licensing for DEISA.....	7
3.2 Code preparation for usage on different DEISA core sites .....	7
3.2.1 Activities for CPMD .....	7
3.2.2 Activities for WIEN2k .....	7
3.2.3 Modules for CPMD and WIEN2k .....	8
3.3 A prototype for a portal for CPMD .....	8
3.3.1 UNICORE configuration.....	9
3.3.1.1 CPMD plug-in.....	9
3.3.1.2 Testing CPMD configuration and execution via plug-in .....	10
4. Scientific Community .....	12

## 1. Introduction

### 1.1 Executive Summary

This document “JRA1: Grid-enabling of selected Materials Science Codes” is the 6-month deliverable DEISA-JRA1-1 for Joint Research Activities in Materials Science. It describes porting activities for the selected codes CPMD and WIEN2K to the different environments encountered in DEISA core sites CINECA, IDRIS and RZG, and the provision of a prototype portal for the CPMD code, thus hiding the distributed DEISA infrastructure for the user.

### 1.2 References and Applicable Documents

- [1] <http://www.cpmid.org>
- [2] M. E. Tuckerman, D. Marx, M. L. Klein, and M. Parrinello, Efficient and general algorithms for path integral Car-Parrinello molecular dynamics, *J. Chem. Phys.* **104**: 5579-5588, 1996
- [3] Manual for CPMD version 3.9.1, May 18, 2004, <http://www.cpmid.org>
- [4] <http://www.wien2k.at>
- [5] R. Dohmen, J. Pichlmeier, M. Petersen, F. Wagner, and M. Scheffler: Parallel FP-LAPW for Distributed-memory Machines. *Computing in Science & Engineering* **3**: 18, 2001
- [6] <http://www.UNICORE.org/documents/UNICOREPlus-Final-Report.pdf>
- [7] <http://www.fz-juelich.de/unicore-test/cpmid>
- [8] V. Huber, “UNICORE: A Grid Computing Environment for Distributed and Parallel Computing”, in *Parallel Computing Technologies*, Proceedings of the 6th PaCT Conference 2001, pp. 258, Springer-Verlag LNCS 2127, 2001.
- [9] <http://sourceforge.net/projects/unicore>
- [10] <http://modules.sourceforge.net>
- [11] <http://www.csc.fi/gopenmol>
- [12] Marcella Iannuzzi, Alessandro Laio, and Michele Parrinello, “Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics”, *Phys. Rev. Lett.* **90**, 238302 (2003); <http://dx.doi.org/10.1103/PhysRevLett.90.238302>
- [13] DEISA-D-JRA3-1

### 1.3 List of Acronyms and Abbreviations

AFS	Andrew File System (a distributed file system)
AJO	representation of A user JOB, managed and submitted to target system by the NJS
BSS	Batch Sub-System on the target machine (IBM LoadLeveler)
CINECA	Consorzio Interuniversitario, Bologna, <a href="http://www.cineca.it">http://www.cineca.it</a>
CPMD	Car-Parrinello Molecular Dynamics (MD) simulations code: calculates an <i>ab initio</i> electronic structure using a plane wave/pseudopotential implementation of the density functional theory (see <a href="http://www.cpmc.org">http://www.cpmc.org</a> )
cpmd2cube	application that converts calculated data into the Gaussian cube file format (see <a href="http://www.cpmc.org">http://www.cpmc.org</a> ). The cube file format originates from the Gaussian software package and describes volumetric data as well as atom positions.
DEISA UNICORE gateway	a server on a dedicated machine that provides a consistent access to the NJS servers at every DEISA site
DEISA site	partner site of the DEISA consortium
FZJ	Forschungszentrum Jülich, <a href="http://www.fz-juelich.de">http://www.fz-juelich.de</a>
g94cube2pl	application of the gOpenMol package that converts volumetric data (e.g. electron densities), represented by the Gaussian cube file format, into a gOpenMol readable format.
gOpenMol	application for the visualisation and analysis of molecular structure data, chemical properties, total electron densities, molecule orbitals (see <a href="http://www.csc.fi/gopenmol">http://www.csc.fi/gopenmol</a> )
GUI	Graphical User Interface
IDB	Incarnation Data Base: the IDB provides site specific descriptions of available computing resources such as CPMD
IDRIS	Institut du Développement et des Ressources en Informatique Scientifique ( <a href="http://www.idris.fr">http://www.idris.fr</a> )
IPP	Max-Planck-Institut für Plasmaphysik, Garching, <a href="http://www.ipp.mpg.de">http://www.ipp.mpg.de</a>
JRA	Joint Research Activity
Module	Modules provide for the dynamic modification of a user's environment via Tcl ( <a href="http://tcl.sourceforge.net">http://tcl.sourceforge.net</a> ) based modulefiles ( <a href="http://modules.sourceforge.net">http://modules.sourceforge.net</a> ).
NJS	Network Job Scheduler: a UNICORE server component that instantiates AJOs or actions (e.g. directory listing on a target machine) by means of a corresponding incarnation data base, schedules jobs and manages job status information and data transfer between the TSI and a gateway or other NJS servers.
TSI	Target System Interface: UNICORE server component which represents the interface to the batch sub-scheduler.
UID	a user's Unique Identifier
UNICORE	UNiform Interface to COmputing Resources,

<http://www.unicore.org>

UNICORE configuration a UNICORE configuration established within DEISA or between some DEISA sites

UUDB UNICORE User Data Base, maps user certificates to uids

WIEN2k program package that performs electronic structure calculations of solids using density functional theory (<http://www.wien2k.at>)

## 2. General remarks

In the first year of the DEISA project JRAs in general face the fact that the infrastructure to be built up and provided by the different Service Activities has not yet been fully established. On the other hand, JRAs cannot wait for their complete availability in order to be able to start and proceed with their tasks. This requires extra work to be done by the JRAs in order to prevent severe delays for the achievement of the JRA deliverables which ensure the availability and usability of key applications as soon as the DEISA production environment will be ready. The JRAs have had to implement workaround environments and have provided the SAs with feedback. In particular, JRA1 and JRA3 gave significant support for the installation and DEISA-optimized configuration of a UNICORE test environment at CINECA and RZG.

The objective of this JRA is to grid-enable key applications from materials science for usage in DEISA, and to support large projects of the European computational materials science community. For this task, the CPMD and WIEN2K codes have been selected as first applications.

- It must be assured that the codes can be executed at the different core sites in an automated way and that it will be able to later use DEISA wide global file systems, so that DEISA-wide job scheduling and rerouting will become possible.
- For new scientists and to generally facilitate code usage and job preparation by new users, a portal should be provided for ease of use.

Future activities will focus more and more on the direct involvement of the European scientific community under the guidance of Prof. M. Parrinello.

## 3. Grid-enabling of selected materials science codes

### 3.1 Selected Materials Science Codes

#### 3.1.1 CPMD

##### 3.1.1.1 Introduction

The CPMD code is a parallelized plane wave/pseudopotential implementation of Density Functional Theory, particularly designed for ab-initio molecular dynamics [1].

CPMD main characteristics:

- ? works with norm conserving or ultrasoft pseudopotentials
- ? LDA, LSD and the most popular gradient correction schemes; free energy density functional implementation
- ? isolated systems and systems with periodic boundary conditions; k-points
- ? molecular and crystal symmetry
- ? wavefunction optimization: direct minimization and diagonalization
- ? geometry optimization: local optimization and simulated annealing
- ? molecular dynamics: constant energy, constant temperature and constant pressure

- ? path integral MD [2]
- ? response functions
- ? excited states
- ? many electronic properties
- ? time-dependent DFT (excitations, molecular dynamics in excited states)
- ? coarse-grained non-Markovian metadynamics

CPMD is copyrighted jointly by IBM Corp and by Max Planck Institute, Stuttgart, and is distributed free of charge to non-profit organizations. CPMD runs on many different computer architectures and it is well parallelized (MPI and Mixed MPI/OpenMP).

### **3.1.1.2 CPMD Code version and licensing for DEISA**

The most recent version 3.9.1 has been taken as the basis for the activities described here because, compared to earlier versions, advanced features have been included [3]:

- ? the methodology for reactive Car-Parrinello metadynamics [12] has been implemented
- ? new functionality including
  - ? Hockney-type Poisson Solver for slabs with influence function in G-Space
  - ? code for trajectory analysis
  - ? calculation of dipole moments using the Berry phase and in real space
  - ? transition matrix elements between orbitals
  - ? growth function for constraints and restraints
  - ? new code for applying static electrical fields
- ? the TDDFT implementation is improved, several bugs have been removed; new exchange-correlation functionals have been implemented to study the effect of different kind of approximations on the excitation properties
- ? the OpenMP parallelisation has been further optimized
- ? keeping wavefunctions in real space is again possible, leading to a reduction of necessary FFT's and thus to a faster execution times and parallel efficiency

RZG has obtained a license valid for the Max Planck Society; virtual organisations or projects as DEISA are, however, not globally covered. Scientists that would like to use CPMD via DEISA are requested to obtain their own license from [www.cpmc.org](http://www.cpmc.org) beforehand (free for non-commercial research institutes). Any publication relating to this code must include the following reference: CPMD V3.9, copyright INTERNATIONAL BUSINESS MACHINES CORPORATION (1990-2004) and MAX PLANCK INSTITUTE FUER FESTKOERPERFORSCHUNG STUTTGART (1995-2001)

## **3.1.2 WIEN2k**

### **3.1.2.1 Introduction**

The program package WIEN2k from the Vienna Technical University [4] performs electronic structure calculations of solids using density functional theory (DFT). It is based on the full-potential (linearized) augmented plane-wave ((L)APW) + local orbitals (lo) method, one of the most accurate schemes for band structure calculations. In DFT

the local (spin) density approximation (LDA) or the improved version of the generalized gradient approximation (GGA) can be used. WIEN2k is an all-electron scheme including relativistic effects and has many features. WIEN2k runs on many different computer architectures and it is well parallelized [5].

The calculated properties are:

- ? energy bands and density of states
- ? electron densities and spin densities, x-ray structure factors
- ? Baders's "atoms-in-molecule" concept,
- ? total energy, forces, equilibrium geometries, structure optimization, molecular dynamics
- ? phonons, with an interface to K.Parlinski's PHONON program
- ? electric field gradients, isomer shifts, hyperfine fields
- ? spin-polarization (ferro- or antiferromagnetic structures), spin-orbit coupling
- ? x-ray emission and absorption spectra, electron energy loss spectra
- ? optical properties
- ? fermi surfaces
- ? LDA, GGA, meta-GGA, LDA+U, orbital polarization
- ? centro- or non-centrosymmetric cells, all 230 spacegroups built in

### **3.1.2.2 WIEN2k Code Version and Licensing for DEISA**

The actual version has been taken as the basis for the described activities. RZG has obtained a license valid for the Max Planck Society; virtual organisations or projects as DEISA are, however, not globally covered. Scientists that would like to use WIEN2k via DEISA are requested to obtain their own license from [www.WIEN2k.at](http://www.WIEN2k.at) beforehand. In any publication in the scientific literature the program has to be referenced as follows:

P. Blaha, K. Schwarz, G. K. H. Madsen, D. Kvasnicka and J. Luitz, WIEN2k, An Augmented Plane Wave + Local Orbitals Program for Calculating Crystal Properties (Karlheinz Schwarz, Techn. Universität Wien, Austria), 2001. ISBN 3-9501031-1-2

## **3.2 Code preparation for usage on different DEISA core sites**

### **3.2.1 Activities for CPMD**

The most recent CPMD version (version 3.9.1) has been installed, configured according to the environments and tested locally at RZG, CINECA and IDRIS for DEISA usage. Applications for data post-processing, cpmd2cube [3] and gOpenMol [11], have been installed and tested at RZG and CINECA.

### **3.2.2 Activities for WIEN2k**

The WIEN2k code consists of about 20 executables, which have to be executed in a certain order for a complete run. Three of the programs are parallel programs. The course of the program is controlled with scripts, in which among others the commands

for starting the parallel programs are constructed from a kind of template which has to be specified in an input file. The code was initially installed and compiled on the IBM of the RZG.

The code package was adapted to other DEISA core sites. The adaptation to the different sites (RZG, CINECA, IDRIS) consisted of configuring the environment required by WIEN2k by specifying the location of necessary resources.

The template for the command script to start parallel code execution (the *poe* command script for IBM systems) had to be modified according to the demands of the respective site. The batch system submit file (here for IBM LoadLeveler batch system) also had to be adapted accordingly.

### 3.2.3 Modules for CPMD and WIEN2k

Under the leadership of the SA4 team the concept of a Central Production Environment has been developed and introduced. It has been agreed to employ the module package [10] at the DEISA sites. Specific modules for CPMD and WIEN2k have been defined and implemented for the RZG environment. The modules specify the paths to resources and the invocation commands for the respective executables. These modules are now ready to be exported to remote DEISA sites.

## 3.3 A prototype for a portal for CPMD

To facilitate usage of CPMD and WIEN2K applications by novel scientists of the European scientific community dealing with simulations in materials science, that do not (yet) have a deep knowledge about code specific parameter setups, provision of a portal would be very helpful. Such a portal has to facilitate the application handling, i.e. job preparation, job submission and output handling. This is a way to hide details of the infrastructure from the user.

Such functionality is principally provided by special features of the UNICORE software package [6], so-called plug-ins for the UNICORE client. In order to be able to test the suitability of such features for CPMD, a complete UNICORE test environment had to be built up at two DEISA sites at least. This was done in a joint effort of SA3, JRA1 and JRA3 for CINECA and RZG. SA3 focuses more on middleware functionality aspects in general. JRA1 (and JRA3) were especially interested in the plug-ins for the UNICORE client. This plug-in technology can be developed further in the scope of the JRAs.

The UNICORE client GUI is of particular interest because it potentially provides a consistent access point for applications such as CPMD and WIEN2k. The modular plug-in technology of the client makes UNICORE a serious candidate for providing the necessary portal functionality. Therefore, UNICORE has been employed to test and demonstrate the grid-enabling of the selected applications.

A UNICORE test environment had to be installed and configured in order to be able to investigate the suitability of the UNICORE client to be employed as a portal for material science applications such as CPMD. Work had to be done for each of the three UNICORE layers:

1. An already available CPMD plug-in [7,8] has been configured and an appropriate visualisation tool [11] has been installed
2. the incarnation data bases (IDB) of the NJS had to be configured appropriately (see also 2.3.1.2)
3. the environments on the target machines had to be customised, particularly a specific adaptation of the TSI was necessary. In addition the modules packages [10] have been installed and configured.

### *3.3.1 UNICORE configuration*

In close cooperation with the SA3 team (focusing on DEISA middleware) and JRA3, the UNICORE configuration has been established [13]. This configuration allows the submission of CPMD jobs to either RZG or CINECA HPC resources for investigations. As a first step the test installation has been done at RZG and CINECA with the option to extend this scalable infrastructure and integrate other DEISA sites (e.g. IDRIS and FZJ). UNICORE client 5.1.2, NJS server 4.0.3, gateway server version 4.0.1 and TSI server 4.0.4 from <http://unicore.sourceforge.net> [9] have been deployed. The configuration has been tested successfully by submitting complex CPMD job chains, prepared with the CPMD plug-in, between CINECA and RZG.

Applications that are made available in the UNICORE framework must first be registered in the UNICORE IDB at each DEISA site. Each registration implies a reconfiguration of the NJS server. The module package turned out to be a powerful tool to avoid frequent requests for changes of the IDB. Applications can be registered in general by declaring the appropriate module load command in the applications section of the IDB, whereas the design and implementation of the appropriate modules can be delegated to the user. Thus, changes in the design of the application's environment requirements, that may be frequent during the application developing phase, do not imply changes in the UNICORE IDB.

The UNICORE client, the CPMD plug-in, the CPMD executable and the executables for data post-processing (cpmd2cube, g94cube2pl) and visualisation (gOpenMol) can be distributed via AFS. The new DEISA AFS cell (/afs/deisa.org) is basically accessible by RZG, CINECA and IDRIS, but currently only used at RZG. Provided the user node runs an AFS client, the UNICORE client can be fetched from AFS and the CPMD plug-in as well as the gOpenMol visualisation toolkit are loaded from there. Additionally at CINECA the software resources are located on the target machine.

#### *3.3.1.1 CPMD plug-in*

The CPMD executable is either located in a separate directory on the target machine at each site or in a global directory provided by a distributed file system. It is invoked by the TSI and receives the input data, controlled by the UNICORE client, via the NJS. The UNICORE client loads a dedicated plug-in that allows configuration of the CPMD job chains. We employed the CPMD plug-in version 4.0.8 that has been developed at FZJ [7,8]. The plug-in provides a complex wizard, i.e. a GUI for input parameter selection and configuration. In addition, the visualisation tool gOpenMol [11] has been chosen for the investigation and presentation of the results of the CPMD calculations (see figure 1).

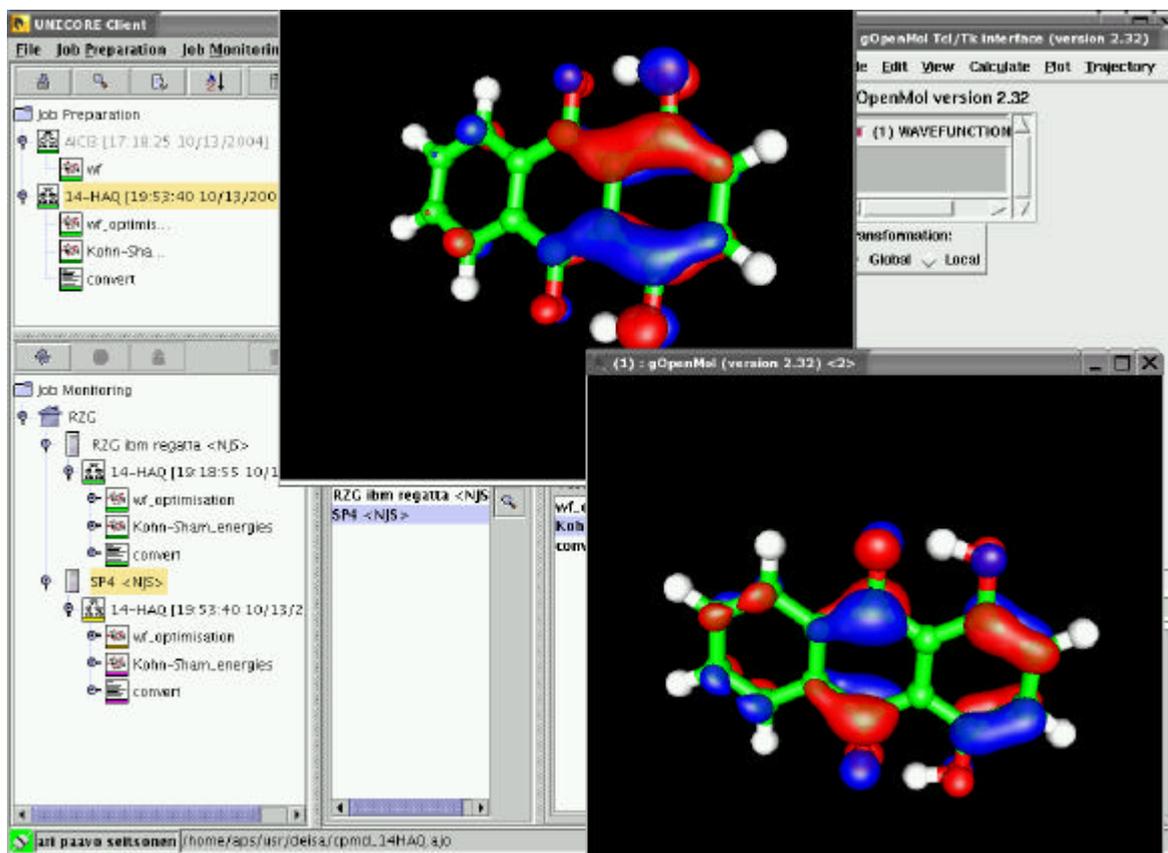
### 3.3.1.2 Testing CPMD configuration and execution via plug-in

The CPMD plug-in was tested for its principal suitability for usage with CPMD. The wizard turned out to be helpful, although a few keywords and parameters of the most recent CPMD version 3.9.1 are currently not fully supported (e.g., the input for the TDDFT method cannot be prepared).

A CPMD simulation is typically configured as a job chain with at least two jobs. The first job performs an optimisation of the wave functions. Starting from these a subsequent job is performed where the ionic coordinates and electronic wave functions are propagated according to the Verlet algorithm. This is the key part of the Car-Parrinello method, which enables one to perform stable molecular dynamics.

Finally, in order to visualise the results with gOpenMol, a post-processing job must be submitted which includes two conversion steps. First the volumetric density or wave function profile, representing the electronic structure of the simulated molecule, is derived from the calculated wave functions. This calculation is performed by `cpmd2cube` [3] and its output is a Gaussian cube file, which is converted by `g94cube2pl` [11] into a format readable by gOpenMol. Both the post-processing tools `cpmd2cube` and `g94cube2pl` have been compiled under IBM AIX 5.2 (the operating system on the target machines).

We generated the CPMD submit scripts for the two chained CPMD jobs with the wizard of the UNICORE CPMD plug-in, based on the parameters that have been entered. Required additional parameters and control keywords (e.g., the section `&TDDFT` for the calculation of electronic excitation energies) have been inserted manually using the script editor provided by the CPMD plug-in.



**Figure 1:** Demonstration of a grid-enabled CPMD execution via the CPMD plug-in showing the job submission to the DEISA sites RZG (successfully finished) and CINECA (still executing), and, visualisation of the result with gOpenMol after the completion of the job chain (at RZG). Depicted are the wavefunction isosurfaces of the molecular orbitals of 1.4-DHAQ (1.4-dihydroxy-9.10-anthraquinone, or chrysazin) that are most significant for the evaluation of its chemical reactivity. Negative parts of the wavefunctions are colored in blue, positive parts in red.

Our CPMD calculations and the post-processing have been performed on both DEISA sites, CINECA and RZG. The UNICORE client with the CPMD plug-in has been successfully verified under Linux and Windows: authentication by certificates, job preparation, submission, execution and result retrieval have all been tested. The visualization tool gOpenMol has been tested under Linux and Windows XP.

---

The functional components have been verified in a number of test cases by the CPMD developer and user Ari Seitsonen, working in the group of Jürg Hutter at the University of Zurich. Figure 1 presents one of the typical results.

The CPMD plug-in and the UNICORE client, in combination with the UNICORE test infrastructure and the environment modules, represent a first basic prototype of a CPMD application portal that can already be used by other DEISA users. The added value provided to the CPMD user includes:

- ? Convenient design of the CPMD output using the CPMD plug-in wizard.
- ? Efficient setup of several CPMD tasks, ranging from different computational steps like initialization of wave functions, main body of the calculations (e.g. Car-Parrinello molecular dynamics), post-processing of internal results into standard file formats and analysis of the results, e.g., the projected density of states. Also the different control structures provided by the UNICORE client such as loops turned out to be very useful for CPMD usage.
- ? The possibility to add shell commands to the generated submit script extends the applicability of the client, allowing the user to execute small tasks without having to rely on the transfer of temporary files to the local machine etc.
- ? Transparent and general control of computational parameters, for example the configuration of the queuing system, relieving the normal user from the need of studying all the technical parameters, thus leaving him/her more time to concentrate on the scientific issues.

As already stated earlier, however, not all new features of the most recent CPMD version 3.9.1 are supported by the CPMD plug-in. Since a few keywords and parameters are not fully supported, the input for the TDDFT method cannot be prepared. This highlights the need for further extension of the plug-in.

All these aspects are most relevant in order to gain acceptance for the CPMD portal by a European user community. Production runs, however, will require further achievements from the other DEISA Service Activities, and, of course, the transfer of the functionality of the test infrastructure to the production systems, a goal to be achieved within 12 months of project start.

After this positive experience with the CPMD plug-in, development of a plug-in for WIEN2k has already started.

#### **4. Scientific Community**

With respect to key users from the European materials science community, Prof. Parrinello from ETHZ (CPMD user) and Prof. Scheffler from FHI (WIEN2K user) have been informed about the progress with the setup of the DEISA infrastructure, and they have been asked about their specific needs for large runs on DEISA. Additionally Prof. Coveney from the UK Reality Grid project has been contacted about his needs for large DEISA runs. Future JRA1 activities will try to more and more directly include the scientific community under the guidance of Prof. Michele Parrinello.