SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

INFRA-2011-2.3.5 – Second Implementation Phase of the European High Performance Computing (HPC) service PRACE

PRACE-2IP

PRACE Second Implementation Project

Grant Agreement Number: RI-283493

D8.1.1
Community Codes Development Proposal

Final

Version:	1.0
Author(s):	Claudio Gheller, Constantia Alexandrou, Miguel Avillez, Ivan Girotto, Sylvie Joussaume, Leonidas Linardakis, Ben Moore , William Sawyer, Thomas Schulthess
Date:	30.09.2011

3.3 Material Science

3.3.1 Introduction and Scientific Challenges

In contrast to traditional areas of physics such as cosmology or high-energy physics, where a few problems (e.g. dark matter, validation of the standard model) consume large parts of the respective communities, materials science consists of many big problems (e.g. photovoltaics, superconductivity, magnetism, nanoscience in various flavours, microelectronics, etc.) that have direct impact on technology and thus society, but the field may appear defocused to the uninitiated.

In particular, electronic structure simulations have emerged as seminal tools for scientific explorations in materials science. This is due to a powerful combination of theoretical developments (in which Europe has been leading the way), new algorithms, and the stunning increases in computer performance as all scales (from workstations to supercomputers). The impact of this development is measurable with a simple metric: 16 of the 20 most cited papers in the history of the American Physical Society (all Physical Review Journals of the APS starting in 1893 [1]) are concerned with quantum mechanical simulations of condensed matter, atoms, and molecules.

Electronic structure simulations of materials are based on first principles, i.e. they do not depend on empirical parameters and are thus predictive. Modern methods based on density functional theory (DFT) [2,3] are very reliable in describing ground state properties of most classes of materials, such as structure, magnetic phases, and as well as basic electron transport properties within linear response theory. In fact, many of the advances in semiconductor electronics are due to the success of band structure theory and corresponding electronic structure simulations of semiconductor heterostructures.

On the other hand, there are classes of materials, such as strongly correlated transition metal oxides, and problems, such as electron transport far from equilibrium or the proper description of electronic excitations in spectroscopy and during chemical reactions, where traditional DFT based electronic structure methods fail systematically. During the past decade, new methods have emerged that can tackle these problems, however there come at a tremendous computational expense. Fortunately, these new methods map well on to petascale systems, both in terms of scalability and efficiency. In fact, the most efficient codes on petascale systems today all fall in this category, as is demonstrated by a series of Gordon Bell Prize finalists and winners in recent years [4,5,6,7,8].

The European electronic structure community is leading the way in terms of theory and community code development [9,10]. However, the ideas that made quantum simulations so successful on petascale computing systems have not yet made it into the European community codes. The main goal of the service activities of this work package will be to bring this high-end technology into community codes and to prepare these codes for the coming age of multi-core and hybrid multi-core/GPU systems.

3.3.2 Numerical Approaches and Codes

Modern electronics structure codes in materials science build on Kohn-Sham Density Functional Theory (KS-DFT) [2] and some approximation to the exchange and correlation energy, such as the Local Spin Density Approximation (LSDA) or the Generalised Gradient Approximation (GGA) [11]. Recent developments in orbital dependent and time dependent density functional theory (TD-DFT) have also found their way into community codes and are routinely used to study electronic excitation spectra of materials, nanoscale systems, and molecules [12]. The codes are used by theorists, simulation experts, as well as experimentalists, and thus have to be robust, easy to use, and portable.

The numerical methods used in these codes are mostly determined by the desired accuracy and the material type on is investigating. For example, plane wave pseudo potential methods (PWPP) [11,13] are used when moderate accuracy is required, and localised core and semi-core states can be absorbed into a pseudo potential, thereby smoothening the potential and drastically reducing the number of plane waves in the expansion of the wave functions. When the accuracy needs are low, one can typically use numerical basis sets that a localised in space (but are often not systematic in the way plane waves are), which leads to sparse problems that can be solved in order *N* time (most electronic structure methods scale with the cube of the number of electrons). On the other hand, when the desired precision is high, even the use of pseudo potentials can be problematic and one has to resort to all-electron methods, where localised electrons feature in the calculations and expansions in plane waves are no longer feasible. Spherical harmonic expansions around the nuclei that are augmented with plane waves in the intermediate region between atoms are used instead – the linearised augmented plane wave LAPW method [13] has emerged as the gold standard in electronic structure simulations.

All these methods rely on the effective use of (in most cases dense) linear solvers. The optimal algorithms may vary, as the basis sets of the various methods differ, i.e. the fractions of eigenvalues and eigenvectors needed to represent the electronic structure depends on the method.

In plane wave codes, the basis set is prohibitively large and the entire Hamiltonian matrix cannot be stored at any given time. Instead, one uses (distributed) Fourier transforms to switch between real space and Fourier space representations, in order to compute different contributions to the Hamiltonian in the most efficient way. One extracts a sub-space with the electron orbitals of interest, in which a sub-space diagonalisation performed, where typically more than 50% of the eigenvalues/eigenvectors have to be computed. Subsequent iterative solves, the distributed Fourier transforms, and the sub-space diagonalisation, constitute the numerical bottlenecks of the method. With an ingenious reorganisation of the way how orbital and plane wave expansion have been distributed over a parallel supercomputer, Kent [14] was able to run large simulations using the standard electronic structure package VASP with 5000 MPI ranks on productive simulations already in 2007. It turned out the sub-space diagonalisation becomes a major computational/scalability bottleneck when scaling the calculations to thousands of nodes. This calls for more efficient multithreaded eigenvalue solvers.

In the LAPW method, the basis set is much more efficient and the entire Hamiltonian of the system can be computed and stored in a uniform way, i.e. there is not need for distributed transforms with global communications as in the plane wave based methods. The basis, however, is not orthogonal and one is left with solving a generalised eigenvalue problem for a Hermitian matrix. Depending on the required precision, between 50% and 10% of the eigenvalues/eigenvectors have to be computed. In extreme cases, where benchmark results used by other methods are computed, the chosen basis set may be so large than only about 1% of the eigenvalue/eigenvectors may be needed to compute the desired electronic structure.

In most other methods, the basis set is more efficient than in the plane wave and LAPW methods just discussed, although at the expense of precision. That is, for a given problem size, the Hamiltonian matrix will be smaller and a larger fraction of the eigenvalues/eigenvectors (typically 40-60%) has to be computed to describe the electronic structure.

Studies have shown [15] that in most cases direct solvers are preferred for electronic structure methods, although the algorithm of choice (e.g. divide an conquer, MRRR, etc.) may vary depending on the method and basis set size. Most electronic structure codes today simply rely

on LAPACK or ScaLAPACK for these solvers. In the coming age of multi-core and hybrid multi-core/GPU systems it is thus desirable to use new algorithms and library implementations that are adapted to highly multithreaded nodes. Furthermore, it will be equally important that the successful strategy to run standard electronic structure codes like VASP at ORNL on many thousands of MPI ranks be introduced into the main European plane wave pseudo potential codes, such as ABINIT and Quantum Espresso.

While the eigenvalue problems and dense linear solves play a fundamental role in almost all modern electronic structure packages, the introduction of TD-DFT and self-consistent GW methods to describe electronic excitations lead to the need to solve integral equations, which, as it turns out, lead to large distributed matrix multiplication problems. Kozhevnikov et al. [Kozhevnikov] demonstrated how the screened Coulomb interaction, W expressed within the LAPW basis set can be computed at scale with very high efficiency. While the work was originally intended at the computation of frequency dependent Hubbard interactions, the real interest in the implementation is its use for the computationally demanding part within self-consistent GW calculations in European software packages such as EXCITING. The combination of these novel implementations of the GW method with very large supercomputers will open up new avenues in electronic structure simulations.

3.3.3 Community involvement, expected outcomes and their impact

The electronic structure community in Europe has been organised in the PsiK network, which started in the early 1990s. While funding from the European Commission (EC) for the network ended many years ago, it continues to strive with contributions from individual chairs across Europe. A significant fraction of active developers of the European electronic structure community is participating in the EC funded European Theoretical Spectroscopy Facility (ETSF), which will be one of the main communities driving the activities of this part to WP8.

The ETSF is (see etsf.eu) “a knowledge centre for theoretical spectroscopy carrying out state-of-the-art research on theoretical and computational methods for studying electronic and optical properties of materials. The ETSF gathers the experience and know-how of more than 200 researchers, facilitating collaborations and rapid knowledge transfer. The ETSF offers its experience to researchers, industry and students in the form of collaborative projects, free scientific software and training.”

It is be this free software that will be targeted by the materials science service activities of the WP8 of PRACE 2IP, in particular the ABINIT package, which includes a PWPP code and BigDFT (a wavelet based code), and the EXCITING package, which is based on the LAPW method. Professors Xavier Gonze of the Catholic University of Louvain, Angel Rubio of the University of the Basque Country, and Claudia Ambrosch-Draxl of the University of Leoben are supporting the efforts from the ETSF side.

A second significant community has formed around Quantum ESPRESSO, a distributed multinational initiative, open to all those who want to contribute to develop it further, or simply use it. Current and past developers are mainly based at the Scuola Internazionale Superiore di Studi Avanzati (SISSA), the Abdus Salam International Centre for Theoretical Physics (ICTP), and CNR-IOM DEMOCRITOS Simulation Center in Trieste; at the Princeton University; at the University of Oxford; at the Ecole Polytechnique Fédérale de Lausanne (EPFL); at the Università di Udine; at the University of California at Davis; at the University of Minnesota; at the CINECA supercomputing centre in Bologna; at the ICHEC, Irish national supercomputing centre; at the Université Pierre et Marie Curie in Paris.

The QE user community consists of hundreds of users worldwide. Relevant person from representative institutions of the community such as Prof. Nicola Marzari, director of the Laboratory for Theory and Simulations of Materials (THEOS) at EPFL; Prof. Stefano Baroni

head of condense matter department at SISSA; Prof. Paolo Giannozzi from Udine University (Italy) and Prof. Stefano De Gironcoli from DEMOCRITOS expressed significant interest in the goals of this project.

The following potential activities will be further investigated and a work plan will be developed in the coming months:

(1) Scaling up of the PWPP codes in ABINIT and QE by reorganizing the way plane waves and orbitals are spread over distributed memory machines according to the prescription given by Kent [Kent]. The expected outcome is significant improvement in scalability of PWPP codes.

(2) Incorporate novel development in eigenvalue solver technology [Solcà], which has recently been implemented into MAGMA in an ongoing collaboration between the Innovative Computing Laboratory at University of Tennessee in Knoxville and the Computational Materials Group at ETH Zürich, into all above-mentioned electronic structure packages mentioned above. The expected outcome will be significant improvement of time to solution on all multi-core and hybrid multi-core/GPU platforms at all scales (from clusters to large supercomputers).

(3) Explore developing basic libraries for PWPP and LAPW with the goal of reducing future porting works of electronic structure software packages to new architectures. A further expected outcome will be that other LAPW based packages besides EXCITING can adopt the work done in this work package with reasonable programming efforts by the respective communities.

(4) Explore using the distributed matrix techniques used by Kozhevnikov and co workers to compute the screened Coulomb interaction in GW and TD-DFT methods that are implemented in all of the above-mentioned packages. The techniques are directly transferable to the EXCITING package. Applicability to ABINIT and QE will have to be studied. The expected outcome will be a drastically improved scalability and efficiency of excited state electronic structure simulations.

The work performed in this part of WP8 is not necessarily limited to the above mentioned software packages, and may be extended to other packages, depending on effort and available manpower. Furthermore, depending on work performed for the climate community, we will explore leveraging structured grid type solver libraries for PDEs in the OCTOPUS software package, which is a multi-grid based TD-DFT package within the ETSF.

3.3.4 Relevant Bibliography

[1] http://mml.materials.ox.ac.uk/Main/Top20

[2] Kohn, W., and Sham, L. J., Self-consistent equations including exchange and correlation effects, Phys. Rev. 140, A1133 (1965).

[3] Kohn, W., Nobel Lecture: electronic structure and matter – wave functions and density functionals, Rev. Mod. Phys. 71, 1253 (1998).

[4] Alvarez, G., Summers, S., Maxwell, D. E., Eisenbach, M., Meredith, J. S., Larkin, J. M., Levesque, J. M., Maier, T. A., Kent, P. R., D'Azevedo, E. and Schulthess, T. C., New algorithm to enable 400+ TFlop/s sustained performance in simulations of disorder effects in high-Tc superconductors, Proceedings of the ACM SC08 Conference (2008).

[5] Wang, L.-W., Lee, B., Shan, H., Zhao, Z. Meza, J., Strohmaier, E. Balley, D., Linear scaling divide-and-conquer electronic structure calculations for thousand atom nanostructures, Proceedings of the ACM SC08 Conference (2008).

[6] Eisenbach, M. Zhou, C., Nicholson, D. M., Brown, G., Larkin, J., and Schulthess, T. C., A scalable method for abinitio computation of free energies in nanoscale systems, Proceedings of the ACM SC09 Conference (2009).
[7] Apra, E., Harrison, R. J., de Jong, W. A., Rendell, A., Tipparaju, V., and Xanthess, S., Proceedings of the ACM SC09 Conference (2009).
[8] Kozhevnikov, A., Eguiliuz, A. G., and Schulthess, T. C., Toward first principles electronic structure simulations of excited states and strong correlations in nano- and materials science, Proceedings of the ACM SC10 Conference (2010).
[9] http://www.abinit.org
[10] http://www.quantum-espresso.org
[11] Martin, R. M., Electronic structure: basic theory and practical methods, Cambridge (2004).
[12] http://www.etsf.eu
[13] Singh, D. J. and Nordstrom, L., Planewaves, pseudopentials, and the LAPW method, Springer (2006).
[14] Kent, P. R. C., Computational Challenges of Large-Scale Long-Time First-Principles Molecular Dynamics, J. Phys.: Conf. Series 125, 012058 (2008).
[15] Auckenthaler, T., Blum, V., Bungartz, H. J., Huckle, T., Jahanni, R., Krämer, L., Lang, B., Lederer, H., and Willems, P. R., Parallel solution of partial symmetric eigenvalue problems from electronic structure calculations, preprint (2010).