The Siesta method. Developments and applicability

Emilio Artacho[1]‡ E Anglada[2], O Diéguez[3], J D Gale[4],
A García[3], J Junquera[5], R M Martin[6], P Ordejón[7],
J M Pruneda[3], D Sánchez-Portal[8] and J M Soler[2]

[1] Department of Chemistry, University of California, Berkeley CA 94720, USA, and Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK
[2] Departamento de Física de la Materia Condensada, C-III, Universidad Autónoma, 28049 Madrid, Spain
[3] Instituto de Ciencia de Materiales de Barcelona (ICMAB-CSIC), Campus de Bellaterra, 08193 Barcelona, Spain
[4] Nanochemistry Research Institute, Curtin University of Technology, Perth, Western Australia 6845, Australia
[5] Departamento de Ciencias de la Tierra y Física de la Materia Condensada, Universidad de Cantabria, Avda de Los Castros s/n, 39005 Santander, Spain
[6] Department of Physics, University of Illinois, Urbana, IL 61801, USA
[7] Centre d'Investigació en Nanociència i Nanotecnologia (CSIC-ICN), Campus de Bellaterra, 08193 Barcelona, Spain
[8] Centro Mixto CSIC-UPV/EHU, Facultad de Química, Apartado 1072, 20080 San Sebastián, Spain

Abstract. Recent developments in and around the Siesta method of first-principles simulation of condensed matter are described and reviewed, with emphasis on (i) the applicability of the method for large and varied systems, (ii) efficient basis sets for the standards of accuracy of density functional methods, (iii) new implementations, and (iv) extensions beyond ground-state calculations.

1. Introduction and overview

Within the last two decades, the first-principles simulation of condensed matter systems has expanded spectacularly, from physics and chemistry into life, earth, nano and materials sciences. This success has been based on both the steady growth of computing power and the development of methods based on density-functional theory (DFT). However, deeply rooted algorithms within the standard methods demand computer resources that grow too rapidly (scaling as N^3) when increasing the number of atoms in a simulation box, N. Although cube-scaling algorithms are still used in most of the DFT calculations done today (up to about 1000 atoms on the best supercomputers), it is obvious that in the long run linear-scaling algorithms will be of advantage.

‡ Permanent address at the University of Cambridge

The early nineties witnessed considerable activity in the search for algorithms that could solve one-particle-like Hamiltonians in linear-scaling operations [1]. After several methods were proposed and tested on empirical Hamiltonians, the attention soon shifted to the problem of how to perform full DFT calculations with a linear-scaling effort, since the calculation of the Kohn-Sham Hamiltonian also displayed inefficient scaling. The SIESTA method started in 1995 [2] by merging Sankey's finite-support atomic orbitals [3], with a 3D real-space grid representation of the density and the basis functions. The grid offered a natural way of dealing with long-range electrostatics, which does not allow for localised treatments (localisation being the key for linear scaling). This method was coupled to several linear-scaling solvers, and coded in the SIESTA program in 1996 [4], a pioneer in linear-scaling DFT.

Other linear-scaling efforts started around that time in both the Quantum Chemistry (QC) and condensed matter physics communities. When overviewing the status of linear-scaling DFT, the first point to emphasise is that we are referring to Kohn-Sham based DFT, as opposed to the so-called orbital-free DFT [5], which computes the total energy directly from the electron density with approximations to the kinetic energy functional in addition to exchange and correlation. This approach is not only linear-scaling, but extremely efficient, and best suited for metals, where linear-scaling Kohn-Sham approaches still fail. The main weakness of orbital-free DFT is that the kinetic energy functionals proposed so far offer good approximations only for systems close to the homogeneous electron liquid, i.e. mostly simple metals [5].

Within Kohn-Sham approaches, several groups in QC have obtained and implemented methods for building the Hamiltonian in linear-scaling operations, using their traditional gaussian basis sets, while the linear-scaling Hamiltonian solvers have not been widely introduced so far. Head-Gordon and collaborators started the line within QC [6] and their methods have been implemented in the QCHEM program [7]. Challacombe's contributions [8] are implemented in the MondoSCF package [9], and the work of Scuseria and coworkers [10] have been incorporated into the GAUSSIAN program [11]. A characteristic of the QC approach to linear scaling is that the key localization approximations are normally done by "thresholding", i.e. by neglecting the matrix elements with values below a given threshold, while other approaches impose localisation regions for the relevant functions from the outset. Another important difference with respect to tendencies in physics is that these methods obtain non-local exact exchange routinely, with an effort similar to the computation of the Hartree term, given that they calculate both within the atomic-orbital representation by computing large amounts of bi-electronic integrals over Gaussian functions. They have thus linear-scaling access to Hartree-Fock and hybrid-functional Hamiltonians at the cost of possibly larger scaling pre-factors for comparable basis sizes (quantitative comparisons are beyond the scope of this work).

Among physicists, the early proposal of finite-elements calculations (the so-called blip functions, distributed in a cartesian way around atoms) together with grid integrals [12] has been implemented in the CONQUEST code [13]. The scalability of

this program for parallel computing is outstanding. It offers an unbiased basis set that can be improved systematically just by introducing more, smaller blips homogeneously in space. The same advantage is offered by the much more recent develement of the ONETEP scheme [14]. It works on a real-space discretisation in a similar fashion as a finite-difference method would, but computes the kinetic energy and other matrix elements with fast Fourier transforms on different "boxes", i.e. regions of space ample enough to offer a good approximation to the matrix element at hand. A strictly finite-difference approach to linear-scaling DFT has also been developed by Fattebert and coworkers [15], and there are also other developments happening around finite-element methods [16]. The capacity for an unbiased convergence of the basis set, and, not least, for having a clear and unique procedure to ensure such convergence comes at a price, which is again the higher computational cost reflected in a higher pre-factor to the linear scaling. The pre-factor scales with the number of basis functions per atom, with powers that depend on the particular method and implementation. Wavelets provide basis sets with localisation in both real and reciprocal space, and are also used in this context [17].

Finally, there are hybrid methods that use a mix of technologies from both the physics and chemistry communities: atomic-orbital basis sets and integration grids, or plane waves as auxiliary basis set. This is the case of the CP2K program (formerly QUICKSTEP) [18], which uses the Gaussian bases of the chemists, and calculates some integrals as in QC, some using plane waves [19]. SIESTA as a method also falls within this family, but using numerical atomic orbitals of finite support instead of Gaussians. This implies several differences. Gaussian basis sets have been developed over decades of QC work, and are tabulated, and thus can be used mainly off the shelf. This is not the case for numerical orbitals of finite support, which are generated by other means (see below). On the other hand, numerical orbitals offer flexibility in the choice of radial shapes of the basis functions at no cost, while varying the radial shapes in Gaussian bases requires adding more Gaussian primitives, which increases the computing effort needed for some of the Hamiltonian matrix elements. The other main difference is in the fact that finite support offers matrix elements that are exactly zero when their support regions do not overlap, while Gaussians tails do not strictly vanish, and matrix elements have to be neglected, as for the "thresholding" described above. This difference is only important if enforcing tight localisation: it is a valid approximation in both cases, but in the finite-support case, the calculations remain exactly within the defined Hilbert space, while in the other case, the operators deviate from it, with possible numerical instabilities for large thresholds.

The SIESTA method has been evolving since its inception, and has been applied by an ever expanding user community to a large variety of problems in many different fields. Note that we are distinguishing here between the SIESTA method and its program implementation: The former is the set of algorithms and ideas as published (see [20] for a detailed account), and thus what defines the calculations, while there are different implementations by other groups in addition to the SIESTA program developed by the SIESTA team. In this paper we report on the present status of the method and its main
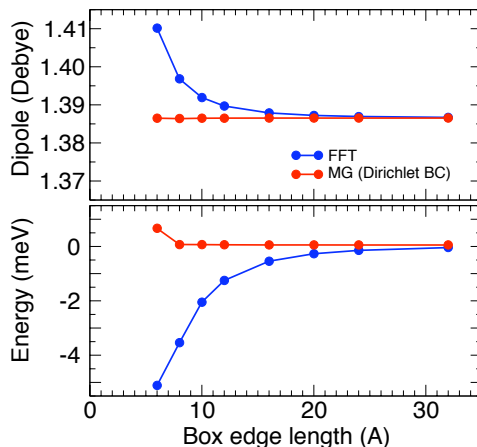
Figure 1. Convergence with box size of energy and dipole moment for a water molecule in a cubic box, using FFT or multigrid (MG) as the Poisson solver.

implementation, other developments that interface with it, and its applicability.

2. Recent developments in SIESTA

2.1. Multigrid solver for electrostatics

The SIESTA method as described before [20] and used nowadays relies on periodic boundary conditions (PBC) in the three directions of space, and so, clusters, molecules or point defects (0D), chains, tubes or line defects (1D), and surfaces, interfaces or plane defects (slabs, 2D) are treated with 3D supercells. The only algorithm that imposes these boundary conditions is the fast Fourier transform (FFT) used for the Hartree term, which scales as $N\log N$. To replace FFTs, a multigrid solver [21] for the Poisson equation on the grid has been implemented [22]. It allows the SIESTA method to become strictly linear-scaling, and offers substantial flexibility with the boundary conditions, allowing for truly 0D, 1D, 2D and 3D calculations. Dirichlet boundary conditions are used on single clusters or molecules (0D), whereby the cluster is not repeated, but the electrostatic potential at the box boundaries is calculated according to the lower moments of the charge density of the system. The multigrid method is compatible with 3D PBC as well as hybrid BC (open in some directions, periodic in others), allowing for unrepeated tubes and slabs.

For a molecule or cluster, the absence of spurious interactions between replicas gives rapid convergence with box size, substantially reducing the computational cost (Fig. 1). In addition, charged clusters or molecules can be treated exactly, without compensating backgrounds. The present multigrid implementation in its PBC form is around five times slower than FFTs for small box sizes and taking the same box for both, although the difference tends to disappear and even revert with size. The smaller sizes needed
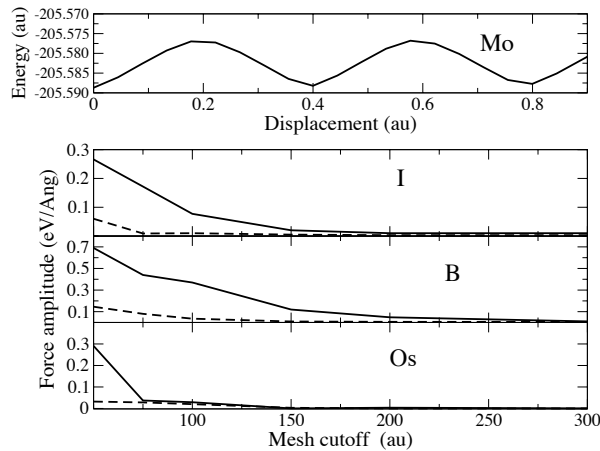
Figure 2. Top panel: Eggbox effect (energy variation with position) for a Mo atom on a 70 Ry grid. Lower panels: Eggbox amplitude versus grid cutoff for I, B, and Os, with (dashed line) and without filtering (continuous line).

by multigrid compensates for the difference in many cases. However, the computational effort devoted to this is not an issue since either of them represents a minute amount (~ 1 %) of the global effort.

2.2. Fourier filtering onto the 3D grid

Grid methods are known to suffer from space rippling problems: the homogeneity of space is lost by the discretisation, and spurious forces appear. This is most apparent in the oscillation of the energy when a single atom is moved across the simulation box ("eggbox" effect, Fig. 2). The problem diminishes for finer grids, but convergence to acceptable levels can be slow. A significant improvement is obtained by grid-cell sampling [20], by averaging over shifted positions of the grid points.

In strictly grid-based (finite differences) methods, a very good solution is provided by Fourier filtering the pseudopotentials to the plane-wave cutoff corresponding to the grid [23]. Although SIESTA has a smaller rippling than finite-difference methods since some key matrix-element calculations do not use the grid, a filtering of the neutral-atom potential, the partial-core density, and the basis functions has been proposed [24]. It preserves strict confinement in real space while maximising the confinement in reciprocal space (a goal shared with wavelet methods). Fig. 2 shows the size of the eggbox versus grid cutoff for different atoms, with and without filtering. The effect is dramatic. Although the filtered basis functions are slightly different from the original ones, they offer better energy conservation in ab initio molecular dynamics (AIMD), more efficient relaxations, and much smoother energy surfaces for finite-difference derivatives.

2.3. Parallelisation

The parallelisation of the SIESTA code means parallelising (i) linear-scaling solvers, (ii) 3D grid operations, and (iii) diagonalisation (maintained as an option for those situations not amenable to linear scaling, such as metals or spectral studies). The first two parts are parallelised using a real-space domain decomposition, which allows good scalability. Benchmarks with a linear-scaling solver [25] for crystalline Si and a minimal basis went up to 524,288 atoms in 256 nodes of an SGI Altix machine in 2001 [26].

The diagonalisation does not scale as well. It is parallelised using ScaLAPACK [27] based on a block cyclic distribution of orbitals over nodes [26]. A substantial improvement was obtained by J Hein in the British HPCx supercomputer centre, by using a 2D distribution of orbitals instead of the original 1D scheme [28]. The great efficiency of the 2D implementation of ScaLAPACK more than compensates for the extra manipulation it requires. A realistic calculation of a pilin protein of 944 atoms, with diagonalisation and a good double-ζ polarised basis, shows speed-up factors 2, 1.82, 1.67 and 1.53, when going from 8 to 16, 16 to 32, 32 to 64, and 64 to 128 processors, respectively [29], on the Cambridge HPCF supercomputer, where an average time step (with 10 SCF steps) takes 25 minutes in 32 processors. In the absence of high-quality communications the practical concurrency limit is 8-16 processors. A revision of the parallelisation strategies is being carried out at the Barcelona Supercomputer Centre [30], in which, in addition to an optimisation of domain decompositions, instances of global communications are being minimised and substituted by the minimum set of point-to-point communication calls, as obtained from graph-theory analysis.

3. Developments around SIESTA

3.1. Phonons, polarisation, effective charges and infrared spectra

Normal modes of vibration are calculated by obtaining the dynamical matrix with a finite-difference derivation of the forces [31], automatised in the VIBRA utility. Phonons at any point in the BZ can be computed using suitable supercells. This is instead of the more popular way of calculating phonons, which uses density-functional perturbation theory (linear response) [32]. The latter represents a more elegant method which can treat any phonon independently from the others. The finite-difference approach scales better with system size, however, allowing for linear-scaling calculations of the phonon modes, including the electronic computation [33]. Linear-response DFT has also been implemented on SIESTA for the calculation of phonons [34]. There is no fundamental difference in the accuracy achievable by both methods, which is limited by the underlying DFT theory, but finite-difference calculations should be done with enough precision to get sensible second derivatives: the egg-box should be minimised as much as possible, and the structure used as reference should be at the energy minimum within a much tighter tolerance than that used for a structural study.

The infrared activity of the modes is obtained from the Born effective charges

(BEC; derivatives of the dielectric polarisation with respect to atom displacements). They are also obtained from finite differences by calculating the polarisation at every displacement used in the force calculations for the dynamical matrix. The BECs allow the calculation of the splitting of the LO and TO phonon bands [35]. The calculation of the polarisation [36] uses by default the Berry-phase formalism, but it becomes very demanding for large systems. An efficient alternative uses the centres of the localised solution functions obtained by the linear-scaling solver of Ordejón *et al.* [25], using the fact that they are Wannier-like functions [37].

3.2. Ballistic transport

Keldysh's method of calculating ballistic currents based on non-equilibrium Green's functions has been implemented on top of the SIESTA method [38], exploiting the suitability of local bases for transport calculations. It allows for finite currents through junctions or interfaces, beyond the linear regime, by obtaining self-consistency in the presence of the bias voltage and current. There are at least four different implementations of this method: TRANSIESTA, as a utility of SIESTA, TRANSIESTAC, distributed commercially by other authors, SMEAGOL [39], and TRANSAMPA [40]. Recent developments include the possibility of inelastic scattering of electrons by phonons [41], as well as the calculation of electromigration effects [42]. It should be remembered, however, that the Kohn-Sham spectrum, with its known limitations, is behind the electron transport properties obtained. Otherwise it overcomes many of the limitations of previous approaches to transport.

3.3. Time-dependent DFT

Improved electronic spectra within the SIESTA context are obtained within the framework of time-dependent DFT (TDDFT) in the frequency domain [43]. It takes advantage of the localisation of the response functions in real space, as discussed in [44]. TD-DFT was also implemented in the time domain [45], and was used initially to obtain the non-linear optical response of C_{60}. It has been recently used to study the non-adiabatic response of insulators to ions moving through them at high velocity [46], the first ab initio calculation of electronic stopping power in insulators. Time-evolving TD-DFT has recently been merged with the new treatment of finite electric fields in periodic systems [47] to study the response of dielectric media to time-dependent electric fields.

4. Basis sets

The SIESTA method requires the use of finite support atomic-like basis functions, i.e., functions composed of a spherical harmonic and a radial function that becomes zero beyond a radius r_c. The user has absolute freedom in everything else: where to center them (on atomic nuclei or not), how many angular momentum channels around any given center, how many radial functions for a given channel, what r_c and what radial
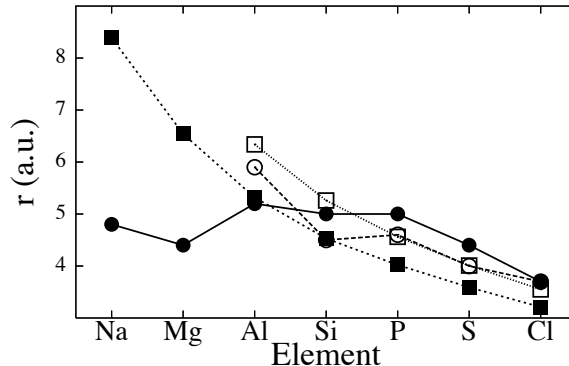
Figure 3. Matching radii for the second-ζ of the $3s$ (filled symbols) and $3p$ (open symbols) orbitals across period III of the periodic table. Circles correspond to bases obtained variationally for Si(diamond), AlP(zinc-blende), MgS(NaCl), and NaCl. Squares refer to standard bases with a 0.1 eV energy shift [48]. Percentile deviations of their respective lattice parameters are 0.0 (1.0), 0.7 (1.3), 0.2 (1.2), and 0.3 (0.0), for the variational (standard) bases, with respect to plane-wave results.

shape to use for each basis orbital. Wisely chosen basis sizes and shapes optimise the efficiency-accuracy dichotomy. There is a vast literature on this in the QC community, but given the finite-support constraint and the numerical flexibility in the radial shape, the SIESTA community builds their own bases following QC strategies. The SIESTA code incorporates different ways to introduce basis functions, and also offers preset algorithms and criteria to define reasonable basis sets automatically [48].

More accurate finite-support bases can be obtained variationally [49, 50]. Although a systematic prescription has not been found, some practical rules have emerged from experience: (*i*) The r_c values for the first-ζ orbitals should not be shorter than 5 Bohr, and there is normally no point in their being longer than 7 Bohr. The energy-shift criterion [48] gives too short values for orbitals of anions and for internal orbitals (e.g. in partially filled $3d$ shells) but unnecessary large for the valence orbitals of alkalis. A value of 6.5 Bohr is sensible in general. (*ii*) Smooth is preferable to hard confinement [49], with an onset between 80% and 90% of r_c, and V_0 between 50 Ry and 100 Ry. Their variations are not critical, except for orbitals that are (close to) unbound in the free atom. (*iii*) The quality of the basis is quite sensitive to the cutoff (matching) radii r_m for second and higher ζ's. Fig. 3 compares optimal values for r_m with the ones obtained by default by SIESTA across the third period of the periodic table. The values towards the right are reasonably reproduced by the standard procedures, if slightly lower, but deviate strongly towards the left. r_m values stay in the neighborhood of 5 Bohr.

The preceding lines refer to varying parameters as in [49], which is a highly non-linear way of changing orbital shapes. Ozaki [51] uses a linear procedure, by obtaining the basis orbitals as linear combinations of a set of eigenstates of the soft-confined atomic problem. An alternative is the expansion of the desired orbitals in terms of the eigenstates of the filtering projectors of ref. [24]. This will be presented elsewhere.

5. Applicability

SIESTA is being used by around a thousand groups worldwide, on a great variety of systems [52, 53]. We give here just a few illustrative examples. SIESTA has been used in nano-science since its inception [2], with many calculations of tubes [54], wires, and clusters, as low connectivity increases SIESTA's efficiency. A very recent study [55] shows results for tapered Si wires with 4044 atoms, obtained with up to only four processors.

Although they demand high accuracy in addition to their inherent size and complexity, biomolecules were treated from the early days [56]. Hydrogen bonds are successfully reproduced and the dynamics of liquid water has been shown to be amenable within the method [57]. For dispersion interactions a non-local functional [58] has been implemented into SIESTA, by means of an algorithm based on a factorisation of the integration kernel, and FFTs, well suited to the grid structure in the method. Details on the implementation and performance will be presented elsewhere [59]. Presently, short AIMD runs or relaxations of proteins of a few thousand atoms can be done. The pilin of *geobacter sulfurreducens* [60] is being calculated in solution, with 4545 atoms, a typical time step (10 SCF steps) taking 5 hours in 128 processors [29]. Alternatively, hybrid QM/MM (quantum mechanics / molecular mechanics) approaches have also been developed for SIESTA, allowing for smaller quantum active sites embedded within a much larger classical environment [61].

Complex oxides frequently appear in materials and earth sciences, including transition-metal and rare-earth ions with highly correlated electrons. Even the systems/properties that can be reliably computed within present GGAs impose computational demands significantly higher than for other materials (a relaxation of a slab of manganite became very heavy for only 224 atoms [62]). Many highly-correlated materials are not well described by GGAs. Other functionals and corrections have been thus implemented, namely, LDA+U [63] and the pseudo-form of the self-interaction correction [64]. Exact exchange for Hartree-Fock and hybrid functionals is being tested, although the computational effort implied is substantially larger.

Many studies demand a chemical analysis of their results. Crystal orbital overlap and Hamiltonian populations [65] have been recently implemented [66], as well as Voronoi deformation density charges, and Mulliken and Hirshfeld populations [67]. Finally, of increasing importance for any method is the interoperability between different codes. SIESTA participates in various efforts, especially in data standarisation. In addition to CML [68] (chemical markup language) input/output [69], pseudopotential formats have been standarised [70] for the joint use of SIESTA and ABINIT [71].

6. Conclusions and outlook

We have shown how SIESTA is developing, and how many new techniques are appearing around it exploiting its technology. However, calculations using full linear scaling are still a small fraction of the total, in spite of ten years of history. The main reason for

this is the wall encountered behind the N^3 barrier: the configurational complexity of the nuclear degrees of freedom. It presently limits the systems that can be addressed to sizes that rarely benefit from linear scaling. Regardless of foreseeable progress addressing nuclear complexity, much more efficient DFT methods are needed for complex systems. We thus face the challenge of substantially reducing the linear-scaling prefactor, while keeping standards of accuracy and robustness. Notwithstanding the possibility of new ideas, we think there is scope for substantial progress in the exploration of the many ideas already proposed, and of variants and synergies thereof.

Acknowledgments

References

[1] S Goedecker, Rev. Mod. Phys. **71**, 1085 (1999); P Ordejón, Phys. Stat. Sol (b) **217**, 335 (2000).
[2] P Ordejón, E Artacho and J M Soler, Phys. Rev. B **33**, 10441 (1996).
[3] O F Sankey and D J Niklewski, Phys. Rev. B **40**, 3979 (1989).
[4] D Sánchez-Portal, P Ordejón, E Artacho and J M Soler, Int. J. Quantum Chem. **65**, 453 (1997).
[5] M Pearson, E Smargiassi and P A Madden, J. Phys.: Condens. Matter **5**, 3221 (1993).
[6] C A White and M Head-Gordon, J. Chem. Phys. **101**, 6593, (1994).
[7] Y. Shao *et al.* Phys. Chem. Chem. Phys. **8**, 3172 (2006).
[8] V Weber and M Challacombe, J. Chem. Phys. **125** 104110 (2006), and references therein.
[9] http://www.t12.lanl.gov/home/mchalla/MondoSCF.html
[10] M C Strain, G E Scuseria and M J G Frisch, Science **271**, 51 (1996).
[11] http://www.gaussian.com
[12] E Hernandez and M J Gillan, Phys. Rev. B **51**, 10157 (1995).
[13] D R Bowler, R Choudhury, M J Gillan and T Miyazaki, Phys. Stat. Sol. b **243**, 989 (2006).
[14] C-K Skylaris, P D Haynes, A A Mostofi and M C Payne, J. Chem. Phys. **122**, 084119 (2005).
[15] J L Fattebert and F Gygi, Phys. Rev. B **73**, 115124 (2006).
[16] E Tsuchida, J. Phys. Soc. Jpn. **76**, 034708 (2007).
[17] A I Neelov and S Goedecker, J. Comput. Phys. **217**, 312 (2006).
[18] J VandeVondele, M Krack, F Mohamed, M Parrinello, T Chassaing and J Hutter, Comp. Phys. Comm. **167**, 103 (2005).
[19] G Lippert, J Hutter and M Parrinello, Mol. Phys. **92**, 477 (1997).
[20] J M Soler, E Artacho, J D Gale, A García, J Junquera, P Ordejón and D Sánchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).
[21] WL Briggs, VE Henson and SF McCormick, *A Multigrid Tutorial*, SIAM, Norwood MA (2000).
[22] O Diéguez and P Ordejón, unpublished (2007).
[23] E L Briggs, D J Sullivan and J Bernholc, Phys. Rev. B **54**, 14362 (1996).
[24] E Anglada and J M Soler, Phys. Rev. B **73**, 115122 (2006).
[25] P Ordejón, D A Drabold, M P Grumbach and R M Martin, Phys. Rev. B **48**, 14646 (1993).

[26] J D Gale, unpublished.

[27] http://www.netlib.org/scalapack

[28] J Hein, Technical Report of HPCx, HPCxTR0410 (2004).

[29] G T Feliciano, G Reguera and E Artacho, to be published.

[30] See http://www.bsc.es

[31] G J Ackland, M C Warren and S J Clark, J. Phys. Condens. Matter **9**, 7861 (1997).

[32] S Baroni, P Gianozzi and A Testa, Phys. Rev. Lett. **58**, 1861 (1987).

[33] P Ordejón, D A Drabold, R M Martin and S Itoh, Phys. Rev. Lett. **75**, 1324 (1995).

[34] J M Pruneda, SK Estreicher, J Junquera, J Ferrer and P Ordejón, Phys. Rev. B **65**, 075210 (2002).

[35] D Fernandez-Torre, R Escribano, T D Archer, J M Pruneda and E Artacho, J. Phys. Chem. A **108**, 10535 (2004), and references therein.

[36] D Sánchez-Portal, I Souza and R M Martin, in *Fundamental Physics of Ferroelectrics 2000*, ed. R E Cohen, pp. 111-120, AIP conference proceedings, vol. **535** (2000), and references therein.

[37] N Marzari and D Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[38] M Brandbyge, J L Mozos, P Ordejón, J Taylor and K Stokbro, Phys. Rev. B **65**, 165401 (2002).

[39] A R Rocha, V García-Suarez, S W Bailey, C J Lambert, J Ferrer and S Sanvito, Nature Materials **4**, 335 (2005).

[40] F D Novaes, A J R da Silva and A Fazzio, Brazilian J. Phys. **36**, 799 (2006).

[41] T Frederiksen, M Brandbyge, N Lorente and A Jauho, Phys. Rev. Lett. **93**, 256601 (2004).

[42] M Brandbyge, K Stokbro, J Taylor, J L Mozos and P Ordejón, Phys. Rev. B **67**, 193104 (2003).

[43] X Blase and P Ordejón, Phys. Rev. B **69**, 085111 (2004).

[44] X Blase, A Rubio, S G Louie and M L Cohen, Phys. Rev B **52**, R2225 (1995).

[45] A Tsolakidis, D Sánchez-Portal and R M Martin, Phys. Rev. B **66**, 235416 (2002).

[46] J M Pruneda, D Sánchez-Portal, A Arnau, J I Juaristi and E Artacho, arXiv.org:0706.1803 (2007).

[47] I Souza, J Íñiguez and D Vanderbilt, Phys. Rev. B **69**, 085106 (2004).

[48] E Artacho, D Sánchez-Portal, P Ordejón, A García and J M Soler, Phys. Stat. Sol. (b) **215**, 809 (1999).

[49] J Junquera, O Paz, D Sánchez-Portal and E Artacho, Phys. Rev. B **64**, 235111 (2001).

[50] E Anglada, J M Soler, J Junquera and E Artacho, Phys. Rev. B **66**, 205101 (2002).

[51] T Ozaki, Phys. Rev. B **67**, 155108 (2003).

[52] D Sánchez-Portal, P Ordejón and E Canadell, Struct. Bond. **113**, 103 (2004).

[53] http://www.uam.es/siesta

[54] M S C Mazzoni, H Chacham, P Ordejón, D Sánchez-Portal, J M Soler and E Artacho, Phys. Rev. B **60**, R2208 (1999).

[55] Z Wu, J B Neaton and J C Grossman, submitted.

[56] E Artacho, M Machado, D Sánchez-Portal, P Ordejón and J M Soler, Mol. Phys. **101**, 1587 (2003).

[57] M V Fernandez-Serra and E Artacho, J. Chem. Phys. **121**, 11136 (2004).

[58] M Dion, H Rydberg, E Schroder, D C Langreth and B I Lundqvist, Phys. Rev. Lett. **92**, 246401 (2004); ibid **95**, 109902 (2005).

[59] J M Soler *et al.*, to be published.

[60] G Reguera, K D McCarthy, T Mehta, J S Nicoll, M T Tuominen and D R Lovley, Nature **435**, 1098 (2005).

[61] A Crespo, D A Scherlis, M A Marti, P Ordejón, A E Roitberg, D A Estrin, J. Phys. Chem. B. **107**, 13728 (2003).

[62] J M Pruneda, V Ferrari, R Rurali, P B Littlewood, N A Spaldin and E Artacho, arXiv.org: 0705.0335 (2007).

[63] S Riikonen and D Sánchez-Portal, to be published.

[64] C D Pemmaraju, T Archer, D Sánchez-Portal and S Sanvito, Phys. Rev. B **75**, 045101, (2007).

[65] R Dronskowski and P E Blochl, J. Phys. Chem. **97**, 8617 (1993).

[66] M Llunell, A García, P Alemany and E Canadell, unpublished.

[67] C F Guerra, J W Handgraaf, E J Baerends and F M Bickelhaupt, J. Comp. Chem. **25** 189 (2003).

[68] R Guha *et al.*, J. Chem. Inf. Mod. **46**, 991 (2006).

[69] http://www.eminerals.org/tools/xml.html .

[70] J Junquera, M Verstraete, X Gonze and A García, unpublished.

[71] X. Gonze *et al.*, Comp. Mat. Sci. **25**, 478 (2002).