

# Finite difference methods for ab initio electronic structure and quantum transport calculations of nanostructures

*J.-L. Fattebert and M. Buongiorno Nardelli*

This article was submitted to Elsevier Science for publication in  
“Handbook of Numerical Analysis”

*U.S. Department of Energy*

Lawrence  
Livermore  
National  
Laboratory

**April 18, 2002**

Including proof corrections (January 2003).

## **DISCLAIMER**

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

# Finite difference methods for ab initio electronic structure and quantum transport calculations of nanostructures

Jean-Luc Fattebert<sup>a</sup> and Marco Buongiorno Nardelli<sup>b</sup>

<sup>a</sup>*Center for Applied Scientific Computing (CASC)  
Lawrence Livermore National Laboratory  
P.O. Box 808, L-561, Livermore, CA 94551*

<sup>b</sup>*Department of Physics, North Carolina State University, Raleigh, NC, and  
Center for Computational Sciences (CCS) and Computational Science and  
Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830*

*Key words:* Finite differences, Electronic structure calculation, First principles molecular dynamics, Preconditioning, Multigrid, Quantum conductance, Green's function, Carbon nanotubes

---

## 1 Introduction

Among the numerical discretization methods used to solve the equations of density functional theory (DFT), the most widely used are linear combination of atomic orbitals (LCAO) — usually Gaussian-type orbitals (GTO)—, plane-waves (PW) and finite differences (FD). Of these three methods, FD is the most recent and less common. General fully 3D grid-based electronic structure representation using finite differences as approximate numerical schemes for partial differential operators have started being widely used in the last ten years only. However, real-space finite difference approaches have already shown to be an efficient tool in a substantial number of large scale electronic structure calculations. Among its various applications, we can cite optical properties of surfaces (Schmidt et al., 2001), surface reconstruction (Ramamoorthy et al., 1998), properties of GaN surfaces (Bungaro et al., 1999), excitation energies and photoabsorption spectra of atoms and clusters (Vasiliev et al., 1999), diffusion of oxygen ions in SiO<sub>2</sub> (Jin and Chang, 2001), first-principles molecular dynamics of carbon nanotubes (Buongiorno Nardelli et al., 1998) and processes in solution (H.Takahashi et al., 2001; Fattebert and Gygi, 2002). In this

---

*Email address:* [fattebert1@llnl.gov](mailto:fattebert1@llnl.gov) (Jean-Luc Fattebert).

article, we review the numerical aspects of this approach for total-energy pseudopotential calculations and show the reasons why it is becoming a method of choice. We also illustrate the method with an application to calculations of electronic structure and conductance of carbon nanotube on a metallic contact. We will limit the discussion to ab initio Density Functional Theory (DFT) models, where no parameters are fitted to experimental data.

Traditionally, chemists have mostly used LCAO methods. Atomic orbitals expressed as linear combinations of Gaussian functions provide an efficient limited basis set that allows for a good description of the electronic structure of localized finite molecules. The wide spread commercial code GAUSSIAN (Gaussian, Inc.) is based on such an approach. On the other hand, plane waves (PW) (Payne et al., 1992; Galli and Pasquarello, 1993; Parrinello, 1997, e.g.) are very efficient to describe periodic systems. Since a plane wave represents a free electron, this approach has been very successful at describing systems with almost free electrons, like metals. By their nature, they have been mostly used and developed by solid state physicists. In this method, known as pseudo-spectral method in the mathematics community, the numerical basis set is completely independent of the positions of the atoms present in the simulation. It can be made as accurate as desired by systematically increasing the number of basis functions included in the basis set.

Like PW, the finite difference method is an alternative to the linear combination of atomic orbitals when highly accurate electronic wave functions are required. Both approaches try to estimate the electronic structure of a physical system without any assumption on where the atoms and electrons are located. While plane waves discretizations benefit from the long and extensive experience of many groups of solid state physicists, there are only a few groups around the world that have developed fully functional codes that allow first-principles molecular dynamics simulations on real-space grids.

In recent years, large parallel supercomputers have become an essential tool in first-principles molecular dynamics simulations. They allow not only calculations that would take months or years on single processor machines, but also calculations that would just not fit into the memory of a workstation or high-end PC. In a parallel environment, the electronic wave functions described in a PW approach can be distributed between the processors for example. Such an approach allows an efficient local application of the FFT algorithm to transform wave functions between real-space and reciprocal space (where the Laplacian is computed). However, every time a matrix element between two wave functions is required — in an orthogonalization process for example —, a huge traffic of data through the whole machine is necessary. In a real-space approach, all the expensive operations can be done locally, thanks to the real-space nature of the DFT Hamiltonian operator and the wave functions. To compute matrix elements between wave functions, local contributions are

computed on every processing element (PE) before being summed up at the end over all the PEs. This is one of the main advantages of FD over PW for nowadays simulations.

Another key element in the development of efficient grid-based FD approaches in large scale electronic structure calculations is the multigrid method (Brandt, 1977). Indeed, real-space large-scale ab initio calculations involve large sparse matrices, and multigrid methods, either as solvers or as preconditioners, allow to design very efficient scalable algorithms (Briggs et al., 1995, 1996; Fattebert, 1996, 1999; Ancilotto et al., 1999; Jin et al., 1999; Wang and Beck, 2000; Fattebert and Bernholc, 2000; Heiskanen et al., 2001).

More recently, in the context of the search for linear scaling algorithms (see Goedecker, this volume, e.g.), real-space methods have appeared to be appropriate for imposing natural localization constraints on the orbitals (Hernandez and Gillan, 1995; Hoshi and Fujiwara, 1997; Fattebert and Bernholc, 2000). Such an approach leads to a dramatic reduction in computer time and memory requirements for very large systems. However, since these algorithms are very recent and useful only for systems of sizes close to the limit of computing resources available today, these methods are still in development and many open questions remain.

Other advantages of FD over standard PW approaches include the possibility of introducing local mesh refinements (Gygi and Galli, 1995; Modine et al., 1996; Fattebert, 1999) and Dirichlet boundary conditions for nonperiodic systems in a very natural fashion. Some of these aspects are discussed for example in a recent review paper by Beck (2000). If local mesh refinement is a requirement for all electrons calculations, many pseudopotential calculations can be carried out with a perfectly regular mesh. Also, local mesh refinements involve many complications such as load balancing for implementation on parallel computers, or Pulay forces. To our knowledge, no large scale firstprinciples molecular dynamics with local mesh refinements have been carried out so far and this aspect of FD methods will not be discussed further here.

The computational advantages of real-space approaches have also motivated some research in 3D finite elements (FE) methods for electronic structure calculations (White et al., 1989; Murakami et al., 1992; Tsuchida and Tsukada, 1995; Kohn et al., 1997; Pask et al., 1999). However, since electronic structure calculations in general require computation domains of very simple shapes like rectangular cells, so far FE methods have not demonstrated real advantages over FD methods for regular grids like those currently used with pseudopotentials. On the other hand, FE methods can be considerably more expensive.

Advanced numerical methods like those presented in this article are useful only if they allow to treat real problems in solid state physics or physical chemistry.

Besides the foreseen improvement in performance for calculations of large scale systems, localized orbital adapted to their chemical environment offer the possibility of novel computational applications. In particular, the possibility of obtaining accurate localized orbital basis would be, in principles, a useful starting point for formal developments such as the semiclassical theory of electron dynamics or the theory of magnetic interactions in solids, in a natural extension of the Wannier representation of crystal wave functions (Ashcroft and Mermin, 1976, Chap.10, p.187 and ff.). Among the possible applications of localized orbital methods, one of increasing technological importance is the calculation of the quantum transport properties of nanostructures.

The current limits of semiconductor electronics and the challenges for future developments involve the continuous shrinking of the physical dimensions of the devices and the attainment of higher speeds. The drive to produce smaller devices has lead the current research towards a new form of electronics in which nanoscale objects, such as clusters or molecules, replace the transistors of today's silicon technology. However, the production and integration of nanoscale individual components into easily reproducible device structures presents many challenges, both experimentally and theoretically. From the theoretical point of view, the design of such devices requires explicit modelling of quantum propagation of electrons in nanoscale systems. The quantity to be calculated is the quantum conductance, that is the measure of the ease with which electrons will transmit through a conductor, or alternatively of the resistance that electrons will encounter in their flow. As we will see in the following sections, the evaluation of conductance in nanostructures requires an electronic structure calculation of the system under consideration, the computation of its Green's function, and an accurate treatment of the coupling to and scattering at the contacts.

Recent years have witnessed a great amount of research in the field of quantum conductance in nanostructures (Beenakker and Van Houten, 1991). These have become the systems of choice for investigations of electrical conduction at mesoscopic scale. The improvements in nanostructured material production have stimulated developments in both experiment and theory. In particular, the formal relation between conduction and transmission, the Landauer formula (Landauer, 1970), has enhanced the understanding of electronic transport in extended systems and has proven to be very useful in interpreting experiments involving the conductance of nanostructures.

The problem of understanding the transport behavior of nanoscale structures cannot be effectively solved without a fully ab initio methodology. Only the latter is able to accurately describe the behavior of the electrons in the highly inhomogeneous environment of the nanoscale device, as well as account for the charge transfer and the interactions within the nano-system. Most of the existing methods to compute conductance from ab initio methods are based on

the solution of the quantum scattering problem for the electronic wave functions through the conductor using a number of related techniques. Lippman-Schwinger and perturbative Green's function methods have been used to study conductance in metallic nanowires and recently in small molecular nanocontacts (Lang, 1995; Di Ventura et al., 2000). Conduction in nanowires, junctions and nanotube systems has been addressed using nonlocal pseudopotentials methods (Choi and Ihm, 1999; Choi et al., 2000) and through the solution of the coupled channels equations in a scattering-theoretic approach (Hirose and Tsukada, 1995; Kobayashi et al., 2000; Landman et al., 2000). The above methods compute *ab initio* transport using a plane wave representation of the electronic wave functions. This imposes severe restrictions on the size of the system because of the large number of basis functions necessary for an accurate description of the electron transmission process. Only recently real-space approaches been considered for a more efficient solution of the electronic transport problem. They are based on the use of LCAO (Yoon et al., 2001; Taylor et al., 2000) or Gaussian orbital bases (Yaliraki et al., 1999). These are combined with either a scattering state solution for the transmission (Yoon et al., 2001) or Green's function-based techniques (Yaliraki et al., 1999; Taylor et al., 2000).

In this article we do not intend to cover the wide variety of techniques that have been developed to compute quantum conductance, both from *ab initio* or from more phenomenological approaches. For such an exhaustive task, we refer the interested reader to the excellent monographs by Datta (1995) and Ferry and Goodnick (1997) and to the references at the end of this article. On the contrary, we have chosen to outline the main steps of the approach derived by us in the context of localized orbital methods (Buongiorno Nardelli et al., 2001). In this article we will limit the discussion to the linear response regime and thus to zero bias across the conductor-lead junctions.

This article is organized as follows. In Sec.2, we review the finite difference method in the context of the Kohn-Sham (KS) equations and pseudopotentials approach, also describing some specific features of iterative algorithms used to solve the KS equations in real-space. We also review the computation of the forces to optimize geometries and carry out first-principles molecular dynamics. We finish Sec.2 with some more advanced features designed to reduce the computational cost of the method in a localized orbitals representation. In Sec.3, we review a numerical method to compute quantum conductance through nanostructures. This method is based on a description of the electronic structure in a basis of localized orbitals. In Sec.4, we conclude this article by an illustration that brings together the localized grid-based orbitals and the quantum conductance algorithm to compute *ab initio* quantum conductance of a carbon nanotube on an Aluminium surface.

## 2 Electronic structure calculation by finite differences

### 2.1 Kohn-Sham equations

Kohn-Sham (KS) theory (Kohn and Sham, 1965) is a widely used model for first-principles calculations (see e.g. Cancès, this volume). It states that the electronic ground state of a physical system can be described by a system of orthogonal one-particle electronic wave functions  $\psi_j, j = 1, \dots, N$  that minimizes the KS total energy functional  $E_{KS}$ . To simplify the discussion we neglect here the spin of the electrons by allowing double occupations of the orbitals, so that the electronic density is defined as

$$\rho_e(\vec{r}) = \sum_{i=1}^N f_i |\psi_i(\vec{r})|^2 \quad (1)$$

where  $0 \leq f_i \leq 2, i = 1 \dots, N$  are the occupation numbers. We also assume that the system is neutral, *i.e.* the total charge of the electrons neutralizes exactly the nuclei charges.

For a molecule composed of  $N_a$  atoms located at positions  $\{\vec{R}_a\}_{a=1}^{N_a}$  in a computation domain  $\Omega$ , the KS energy functional is given by (in atomic units)

$$E_{KS} [\{\psi_i\}_{i=1}^N, \{\vec{R}_a\}_{a=1}^{N_a}] = \sum_{i=1}^N f_i \int_{\Omega} \psi_i^*(\vec{r}) \left( -\frac{1}{2} \nabla^2 \right) \psi_i(\vec{r}) d\vec{r} \\ + \frac{1}{2} \int_{\Omega} \int_{\Omega} \frac{\rho_e(\vec{r}_1) \rho_e(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 + E_{XC}[\rho_e] + \int_{\Omega} \psi_i^*(\vec{r}) (V_{ext} \psi_i)(\vec{r}) d\vec{r}. \quad (2)$$

The first term represents the kinetic energy of the electrons. The second represents the electrostatic energy of interaction between electrons that we will note  $E_{es}$ .  $E_{XC}$  models the exchange and correlation between electrons. In this paper, we will use the local density approximation (LDA), or the first-principles exchange-correlation functional proposed by Perdew-Burke-Ernzerhof (PBE) (Perdew et al., 1996) which often provides results in better agreement with experiments and is appropriate for a grid-based implementation. In the last term of Eq.(2), the potential  $V_{ext}$  represents the total potential produced by the atomic nuclei at positions  $\{\vec{R}_a\}_{a=1}^{N_a}$ .

The ground state of a physical system is represented by orbitals that minimize the energy functional (2) under the constraints that the  $\psi_j$  are orthonormal. This minima can be found by solving the associated Euler-Lagrange equations — Kohn-Sham equations (Kohn and Sham, 1965) —

$$H\psi_j = \left[ -\frac{1}{2}\nabla^2 + v_H(\rho_e) + \mu_{xc}(\rho_e) + V_{ext} \right] \psi_j = \epsilon_j \psi_j, \quad (3)$$

which must be solved self-consistently for the  $N$  lowest eigenvalues  $\epsilon_j$ , while imposing the orthonormality constraints  $\langle \psi_i | \psi_j \rangle = \delta_{ij}$ . We use the usual quantum mechanics notation  $\langle . | . \rangle$  for the  $L^2$  scalar product. The Hartree potential  $v_H$  represents the Coulomb potential due to the electronic charge density  $\rho_e$ , and  $\mu_{xc} = \delta E_{xc}[\rho_e] / \delta \rho_e$  is the exchange and correlation potential.

## 2.2 Finite differences approach

In order to discretize the KS equations, we introduce a real-space rectangular grid  $\Omega_h$  of mesh spacing  $h_x, h_y, h_z$  in the directions  $x, y, z$  that covers the computation domain  $\Omega$ . Let  $M$  be the number of grid points. The wave functions, potentials and the electronic density are represented by their values at the grid points  $\vec{r}_{ijk} = (x_i, y_j, z_k)$ . Integrals over  $\Omega$  are performed using the discrete summation rule

$$\int_{\Omega} u(\vec{r}) d\vec{r} \approx h_x h_y h_z \sum_{i,j,k \in \Omega_h} u(\vec{r}_{ijk}).$$

Given the values of a function  $u(\vec{r})$  on a set of nodes  $\vec{r}_{i,j,k}$  the traditional finite difference approximation  $w_{i,j,k}$  to the Laplacian of the function at a given node is expressed as a linear combination of values of the function at the neighboring nodes

$$w_{i,j,k} = \sum_{n=-p}^p c_n (u(x_i + nh_x, y_j, z_k) + u(x_i, y_j + nh_y, z_k) + u(x_i, y_j, z_k + nh_z)) \quad (4)$$

where the coefficients  $\{c_n\}$  can be computed from the Taylor expansion of  $u$  near  $\vec{r}_{i,j,k}$ . Such an approximation has an order of accuracy  $2p$ , that is for a sufficiently smooth function  $u$ ,  $w_{i,j,k}$  will converge at the rate  $O(h^{2p})$  as the mesh spacing  $h \rightarrow 0$ . For the second order approximation for example ( $p = 1$ ), we have  $c_0 = 2/h^2$  and  $c_1 = -1/h^2$ . High order versions of this scheme were first used in electronic structure calculations by Chelikowsky et al. (1994).

As an alternative, one can also use a compact finite difference scheme (also called *Mehrstellenverfahren* in Collatz (1966)). For example, a 4<sup>th</sup> order FD scheme for the Laplacian on a cubic grid is based on the relation

$$\frac{1}{6h^2} \left\{ 24u(\vec{r}_0) - 2 \sum_{\substack{\vec{r} \in \Omega_h, \\ \|\vec{r} - \vec{r}_0\| = h}} u(\vec{r}) - \sum_{\substack{\vec{r} \in \Omega_h, \\ \|\vec{r} - \vec{r}_0\| = \sqrt{2}h}} u(\vec{r}) \right\}$$

$$= \frac{1}{72} \left\{ 48(-\nabla^2 u)(\vec{r}_0) + 2 \sum_{\substack{\vec{r} \in \Omega_h, \\ \|\vec{r} - \vec{r}_0\| = h}} (-\nabla^2 u)(\vec{r}) + \sum_{\substack{\vec{r} \in \Omega_h, \\ \|\vec{r} - \vec{r}_0\| = \sqrt{2}h}} (-\nabla^2 u)(\vec{r}) \right\} \quad (5) \\ + O(h^4),$$

valid for a sufficiently differentiable function  $u(\vec{r})$ . For simplicity, we have assumed here that  $h_x = h_y = h_z$ , but this expression is easy to extend to the general case. This FD scheme requires only values at grid points within a sphere of radius  $\sqrt{2}h$ . Beside its good numerical properties, the compactness of this scheme reduces the amount of communications in a domain-decomposition based parallel implementation. While increasing the order of the FD scheme improves the accuracy for very fine grids, it may not be the case for a given computational grid. In practice, this compact 4<sup>th</sup> order scheme consistently improves the accuracy compared to a standard 4<sup>th</sup> order scheme, as illustrated in Fig.1.

**Remark 2.1** *The FD method is not variational, and by refining the mesh the total energy generally increases towards convergence.*

It is easy to see that a compact FD scheme like (5) leads to an eigenvalue problem of the form

$$(L_h + B_h V_h) \vec{\psi}_i = \epsilon_i B_h \vec{\psi}_i \quad (6)$$

where  $L_h$  represents the FD scheme on the left-hand side of Eq.(5) and  $B_h \in \mathcal{M}_M$  is a sparse well conditioned matrix that represents the FD-like scheme on the right-hand side of Eq.(5).  $V_h$  represents the potential on the grid. Let  $H_h = L_h + B_h V_h$ . One can show that  $B_h^{-1} H_h$  is symmetric ( $B_h$  and  $L_h$  commute) so that the eigenvalues  $\epsilon_i$  are real and the eigenvectors  $\vec{\psi}_i$  can be chosen orthogonal.

This type of compact FD scheme were simultaneously introduced in electronic structure calculations by Briggs et al. (1995) and Fattebert (1996). The simulations presented in this article are based on this scheme. However, to simplify the notations, we will drop the matrix  $B_h$  in the rest of this article. In general, from the point of view of computer time and memory requirements, FD schemes of order larger than 2 are clearly worthwhile since they allow to work with much coarser grids. This reduces the cost of operations like scalar products between trial eigenfunctions or linear combinations of trial eigenfunctions in the iterative solver (see Sec.2.5).

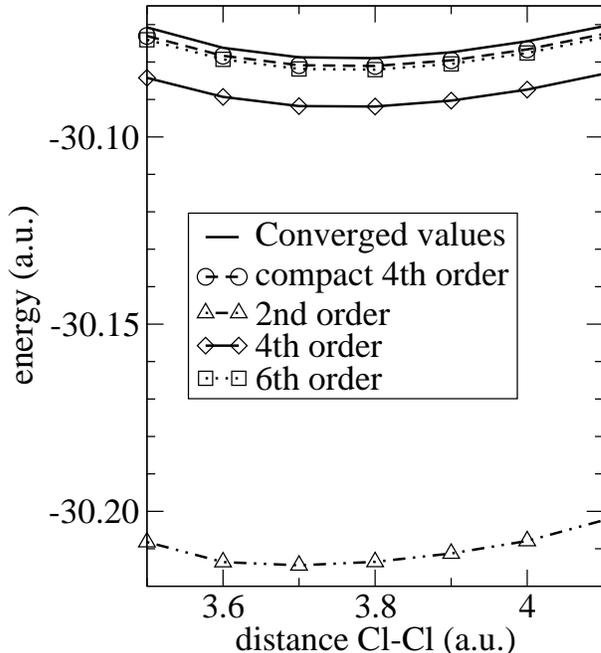


Fig. 1. Total energy for a  $Cl_2$  molecule as a function of the distance  $Cl - Cl$  for several finite difference schemes, using the PBE exchange and correlation functional and pseudopotentials by Hamann (1989). The grid spacing used ( $h = 0.34$ ) is sufficient to accurately compute the equilibrium bond length and binding energy of the molecule using the schemes of order 4 and 6, while it is clearly too coarse for the 2<sup>nd</sup> order scheme. The top line shows a fully converged result.

### 2.3 General formulation in nonorthogonal orbitals

Most of the time, we are not interested in the individual eigenfunctions solutions of the KS equations, but only in the subspace spanned by these functions. It means that one can represent the subspace of the electronic orbitals by a more general basis of nonorthogonal functions,  $\{\phi_1, \dots, \phi_N\}$ . We write these functions as vectors, columns of a matrix  $\Phi$ ,

$$\Phi = (\vec{\phi}_1, \dots, \vec{\phi}_N).$$

An orthonormal basis of approximate eigenfunctions (Ritz functions) can be obtained by a diagonalization in this subspace of dimension  $N$  (Ritz procedure). We denote by  $C \in \mathcal{M}_N$  the matrix that transforms  $\Phi$  into the basis  $\Psi$  of orthonormal Ritz functions,

$$\Psi = (\vec{\psi}_1, \dots, \vec{\psi}_N) = \Phi C. \quad (7)$$

The matrix  $C$  satisfies

$$CC^T = S^{-1},$$

where  $S = \Phi^T \Phi$  is the overlap matrix.

In the following, for an operator  $A$  we will use the notation

$$A^{(\Phi)} = \Phi^T A \Phi.$$

We then have the relation

$$A^{(\Psi)} = C^T A^{(\Phi)} C.$$

Equation (7) defines a transformation to a Ritz basis only if  $C$  is a solution of the generalized symmetric eigenvalue problem

$$H^{(\Phi)} C = S C \Lambda, \tag{8}$$

where  $\Lambda \in \mathcal{M}_N$  is a diagonal matrix that satisfies  $\Lambda = \Psi^T H \Psi$ . The matrix  $C$  can actually be decomposed as a product  $C = L^{-T} U$ , where  $L$  is the Cholesky factorization of  $S$ ,

$$S = L L^T,$$

and  $U$  is an orthogonal matrix. Knowing  $L$ , the generalized eigenvalue problem (8) is reduced to a standard symmetric eigenvalue problem

$$L^{-1} H^{(\Phi)} L^{-T} U = U \Lambda. \tag{9}$$

For a chemical potential  $\mu$ , let us define  $\Upsilon \in \mathcal{M}_N$  by its matrix elements

$$\Upsilon_{ij} = \delta_{ij} f[(\epsilon_i - \mu)/k_B T],$$

where  $f$  is a Fermi-Dirac distribution at temperature  $T$  and  $k_B$  is the Boltzmann constant. The density operator  $\hat{\rho}$  is then defined as

$$\hat{\rho} = \Psi \Upsilon \Psi^T = \Phi C \Upsilon C^T \Phi^T.$$

For  $T = 0$ ,  $\hat{\rho}$  is a projector onto the states of eigenvalues lower than  $\mu$ . The dimension of this density matrix is given by the number of degrees of freedom, *i.e.*, the number of grid points in a grid-based approach. This number is in general so large that it is impossible to apply numerical methods that require  $\rho(\vec{r}, \vec{r}')$  (matrix of size  $M \times M$ ). However, it is useful to represent  $\hat{\rho}$  in the basis  $\Phi$

$$\rho^{(\Phi)} = \Phi^T \hat{\rho} \Phi = C^{-T} \Upsilon C^{-1}.$$

Even more useful is the matrix  $\bar{\rho}^{(\Phi)}$

$$\bar{\rho}^{(\Phi)} = S^{-1} \rho^{(\Phi)} S^{-1} = C \Upsilon C^T. \quad (10)$$

This matrix appears naturally in the expression used to compute the expectation value  $\bar{A}$  of an operator  $A$  represented in the basis  $\Phi$ ,

$$\bar{A} = 2tr(\Upsilon A^{(\Psi)}) = 2tr(\bar{\rho}^{(\Phi)} A^{(\Phi)}).$$

In particular, the total number of electrons in the system is given by

$$N_e = 2tr(\bar{\rho}^{(\Phi)} S).$$

Also, the electronic density in a nonorthogonal orbitals formulation is simply given by

$$\rho_e(\vec{r}) = 2 \sum_{j,k=1}^N (\bar{\rho}^{(\Phi)})_{jk} \phi_j(\vec{r}) \phi_k(\vec{r}). \quad (11)$$

**Remark 2.2** *If all the computed states are fully occupied, we have  $\rho^{(\Phi)} = S$  and  $\bar{\rho}^{(\Phi)} = S^{-1}$ .*

## 2.4 Pseudopotentials

On regular grids, FD methods, like plane-waves, are not very efficient to describe singular atomic potentials accurately. In particular for microcanonical molecular dynamics, it is difficult to guarantee a good conservation of the total energy of the system. These singularities can however be removed by replacing the atomic nuclei and core electrons — which can be approximated as frozen in their atomic state — by pseudopotentials. In the pseudopotential approach, the electronic structure calculation problem is reduced to the computation of a cloud of valence electrons living in a background of positive ions represented by smooth non-singular pseudopotentials.

Accurate calculations can be performed by representing each atomic core by a nonlocal separable pseudopotential in its Kleinman-Bylander form (Kleinman and Bylander, 1982)

$$V_{ps} = V_{local} + V_{nl} = v_{ps}^{local}(\vec{r}) + \sum_{\ell=0}^{\ell_{max}} \sum_{m=-\ell}^{\ell} |v_{\ell}^m\rangle E_{\ell}^{KB} \langle v_{\ell}^m| \quad (12)$$

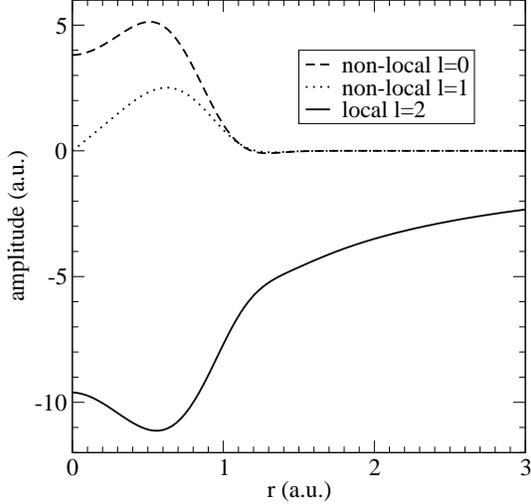


Fig. 2. Example of atomic pseudopotential: Chlorine for the PBE functional. The radial functions plotted are the radial components of the projector functions,  $v_\ell$  for  $\ell = 0, 1$  and the local pseudopotential  $v_{ps}^{local}$  ( $\ell = 2$ ).

where  $E_\ell^{KB}$  are normalization coefficients. The function  $v_{ps}^{local}$  contains the long range effects and is equal to  $-Z/r$  outside of the core. The functions  $v_\ell^m(\vec{r})$  are the product of a spherical harmonics  $Y_\ell^m$  by a radial function  $v_\ell(r)$  — centered on an atom — which vanishes beyond some critical radius. Being separable means that the matrix elements  $\langle \psi_i | V_{nl} | \psi_j \rangle$  can be computed efficiently according to

$$\langle \phi_j | V_{nl} | \phi_k \rangle = \sum_{\ell=0}^{\ell_{max}} \sum_{m=-\ell}^{\ell} E_\ell^{KB} f_{\ell m}(\phi_j) f_{\ell m}(\phi_k) \quad (13)$$

where the quantities  $f_{\ell m}(\phi_n) = \langle \phi_n | v_\ell^m \rangle$  can be computed independently of each other. Since the functions  $v_\ell^m$  are localized in real-space, the evaluation of all the matrix elements  $\langle \psi_j | V_{nl} | \psi_k \rangle$  for a system of  $N$  atoms scales like  $O(N^2)$ .

In the applications presented in this article, we use the pseudopotentials proposed by Hamann (1989). An example is represented in Fig. 2 (Chlorine).

**Remark 2.3** *All the atoms of the periodic table cannot be represented by pseudopotentials with the same degree of smoothness. To be represented accurately, atomic species like Oxygen or Nitrogen require finer discretization grids than Silicon for example. This makes the calculations of the electronic structure of a crystal of 64 Silicon atoms 5-10 times cheaper than the simulation of a cell of liquid water with 32 molecules, even if the number of valence electrons to compute is the same in both cases.*

Using periodic boundary conditions, the total energy of a system is in principle

invariant under spatial translations. Unlike in a PW approach, a real-space finite grid representation breaks this invariance (Briggs et al., 1995). However, a discretization grid fine enough ensures that the total energy is conserved under translation within the minimal accuracy required in the calculation. Also, in order to avoid problems related to energy variations under spatial translations, the pseudopotentials can be filtered (Briggs et al., 1996). The pseudopotentials (local part and projectors) are first transformed to Fourier space by the Fourier transform

$$V_{\ell m}(\vec{k}) = \frac{1}{(2\pi)^{3/2}} \int_{R^3} v_{\ell}^m(\vec{r}) e^{-i\vec{k}\vec{r}} d\vec{r} = C_{\vec{k},\ell,m} \int_0^{\infty} v_{\ell}(r) j_{\ell}(|\vec{k}|r) r^2 dr. \quad (14)$$

The index  $(\ell, m)$  denotes the symmetry of the functions  $v_{\ell}^m(\vec{r}) = v_{\ell}(r) Y_{\ell}^m(\theta, \phi)$ .  $j_{\ell}(r)$  is a spherical Bessel function of order  $\ell$  and  $C_{\vec{k},\ell,m}$  is a complex factor depending on  $\vec{k}, \ell, m$ . For  $|\vec{k}|$  larger than a cutoff  $k_{cut}$ , the coefficients  $V_{\ell m}(\vec{k})$  are filtered by a Gaussian function  $e^{-\beta(|\vec{k}|/k_{cut}-1)^2}$  before applying an inverse Fourier transform. One can use for example  $k_{cut} = 2\pi/3h$ . In practice, since the filtering depends only on  $|\vec{k}|$ , the coefficients  $C_{\vec{k},\ell,m}$  do not need to be computed and only the 1D radial integral is required for an appropriate range of  $|\vec{k}|$ . The pseudopotentials lose in general their localization properties in real-space after being filtered in Fourier space and a second filtering in real-space is required to ensure that the nonlocal projectors remain confined within a limited radius. This second filtering can also be done by a Gaussian function to avoid reintroducing too many high frequency components.

Of course this filtering procedure modifies the pseudopotentials to a degree set by the grid spacing. These filtered functions will however converge towards the true pseudopotentials — together with the wave functions — as one decreases the mesh spacing. For every atomic species, one should then carefully check what grid spacing is required to ensure that the physical quantities of interest are well converged.

## 2.5 Solving the eigenvalue problem

The Kohn-Sham equations discretized by FD result in a huge 3D eigenvalue problem. Fortunately, the matrices involved are very sparse and efficient iterative methods can be used to solve this problem. In this chapter, we are going to restrict the discussion to the particularities of eigenvalue solvers for finite difference discretizations. After introducing some general features of minimization processes, we describe an appropriate multigrid preconditioner for real-space discretizations.

The multigrid full-approximation scheme (FAS) originally proposed by Brandt (1977) is an efficient solver for nonlinear problems on a grid. In this algorithm, the entire problem has to be represented on all the grids, from the coarsest to the finest, in order to treat on equal footing all the length scales of the solution. Such an approach is not obvious to apply to the KS eigenvalue problem. Indeed, the numerous eigenfunctions of interest lose their meaning on very coarse grids, and one may be limited in the number of usable coarse grids. Only a few successful applications of the FAS algorithm for electronic structure calculations have been reported so far. They have been limited to purely academic problems (Costiner and Taasan, 1995) and static all electrons calculations of atoms and diatomic molecules (Wang and Beck, 2000). Recently, Heiskanen et al. (2001) proposed to use the Rayleigh quotient multigrid (RQMG) method (Mandel and McCormick, 1989) that avoids the coarse grid representation problem. So far it was applied only for static electronic structure calculations.

Various other methods are based on minimization schemes that make use of the steepest descent (SD) direction along which the energy functional decreases at the fastest rate. This direction is given by the gradient of the energy functional with the opposite sign. In the basis  $\Psi$ , this direction — in a space of dimension  $M \times N$  — is easy to compute since it is given by the negative residual of the Kohn-Sham equations (3) and can be expressed as an  $N \times M$  matrix

$$D^{(\Psi)} = \Psi\Lambda - H\Psi. \quad (15)$$

One verifies that this gradient satisfies the relation  $\Psi^T D^{(\Psi)} = 0$ .

For an optimum convergence rate, it is important to use the *true* steepest descent direction in algorithms expressed in non-orthogonal orbitals formulation. This direction can differ substantially from the derivative with respect to  $\Phi$  if the basis  $\Phi$  is highly non-orthogonal. This SD direction is easy to compute for the eigenfunctions  $\Psi$  and a simple way to obtain it in the basis  $\Phi$  is to use the matrix  $C$  (from Eq.(7)) to derive

$$D^{(\Phi)} = D^{(\Psi)}C^{-1} = (\Psi\Lambda - H\Psi)C^{-1} = \Phi\Theta - H\Phi, \quad (16)$$

where  $\Theta = S^{-1}H^{(\Psi)}$ . In the following, we consider a SD algorithm with a linear preconditioning operator  $K$ . The basis  $\Phi$  is updated according to

$$\Phi^{new} = \Phi + \eta K (\Phi\Theta - H\Phi) \quad (17)$$

where  $\eta$  is a pseudo-time step. In this algorithm all the trial wave functions are updated simultaneously. In the particular case  $K = Identity$ , (17) is equivalent to the method proposed in Galli and Parrinello (1992). Since by definition

$$\Phi^{new} = (\Psi + \eta K D^{(\Psi)}) C^{-1} = \Psi^{new} C^{-1},$$

the same subspace is generated at each iteration, independently of the choice of the basis  $\Phi$ . Note also that Eq.(17) does not depend on  $C$  and therefore does not require the solution of the eigenvalue problem (8).

**Remark 2.4** *Alternatively, the above SD directions can be used in a conjugate gradient (CG) method (Edelman et al., 1998). In the applications presented here, the convergence rate of the preconditioned steepest descent (PSD) algorithm with the preconditioning presented later in this Section is fast enough to make the implementation of the CG approach unnecessary.*

In actual calculations, the basis functions  $\Phi$  are corrected at each iteration using the PSD directions as in (17). A new electronic density  $\rho_e$  is then computed as well as new Hartree and exchange-correlation potentials. To avoid large oscillations of the charge distribution from step to step, these potentials can be mixed linearly with those used at the previous step. The basis  $\Phi$  is refined by iterative updates until self-consistency (SC) is achieved and, at convergence accurately describe the true Kohn-Sham ground state of the system.

**Remark 2.5** *For iterative algorithms based on the Ritz vectors, the transformation (7) is one of the most expensive part of the calculation for large scale simulations. In a nonorthogonal representation, this operation is not required anymore. The cost is however transferred to the computation of the SD directions (Eq.16).*

In the particular case of a linear Hamiltonian operator and if all the orbitals are fully occupied, the Kohn-Sham functional expressed in nonorthogonal orbitals can be written as a functional without constraints

$$E_{KS} = 2tr \left( S^{-1} \Phi^T H \Phi \right). \quad (18)$$

To motivate the introduction of a preconditioner, we first estimate the expected rate of convergence of iterative algorithms based on steepest descent directions to minimize Eq.(18). The rate of convergence of the SD method is determined by the condition number  $\chi(\mathcal{H})$  of the Hessian matrix  $\mathcal{H}$  associated to the problem. To estimate  $\chi(\mathcal{H})$  we compute the eigenvalues of  $\mathcal{H}$ . As in Pfrommer et al. (1999), we consider electronic states  $\{\phi_i\}_{i=1}^N$  expressed as a perturbation of the ground state eigenfunctions  $\{\psi_i\}_{i=1}^N$

$$\phi_i = \psi_i + \sum_{l=1}^M c_l^{(i)} \psi_l. \quad (19)$$

Inserting (19) into (18), we obtain to the second order in the coefficients  $c_l^{(i)}$

$$E_{KS} - E_0 = 2 \sum_{i=1}^N \sum_{k=m+1}^M (\epsilon_k - \epsilon_i) \left( c_k^{(i)} \right)^2. \quad (20)$$

The coefficients  $c_l^{(i)}$  for  $l \leq N$  correspond to directions in the parameter space along which the objective function is constant. In the complementary parameter space, we obtain that the condition number of the Hessian matrix is given by

$$\chi(\mathcal{H}) = \frac{\epsilon_M - \epsilon_1}{\epsilon_{N+1} - \epsilon_N}. \quad (21)$$

**Remark 2.6** *In general, as the number of atoms in a physical system increases, the spectrum of the Hamiltonian becomes denser, while the extreme eigenvalues ( $\epsilon_M$  and  $\epsilon_1$ ) remain about the same.*

As in PW calculations, real-space representations of the electronic wave functions require a very large number of degrees of freedom. In particular in the presence of atoms represented by very hard pseudopotentials. A very fine grid implies a quite large value for  $\epsilon_M$  in Eq.(21) that negatively affects the condition number of the Hessian matrix. If  $\Psi$  is corrected at each step according to a SD algorithm without preconditioning

$$\Psi^{new} = \Psi + \eta D^{(\Psi)},$$

the parameter  $\eta$  has to be very small — for numerical stability reasons — and the convergence can be very slow.

**Remark 2.7** *Equation (21) also points out at problems one can observe if the band gap ( $\epsilon_{N+1} - \epsilon_N$ ) is very small. In practice, one can overcome this limitation by including more eigenstates than needed in the search subspace  $\Psi$ , the highest eigenstates being empty or fractionally occupied.*

To introduce an appropriate preconditioner, we start by discussing the Rayleigh Quotient Iteration (RQI) method and some of its variants use in real-space electronic structure calculations. RQI is a very fast algorithm to compute one single eigenvalue of a matrix (e.g. Parlett, 1998, Chap.4). Here we describe this method in a different form that we find more suitable for matrices of size too large to make use of direct linear solvers. Starting from an approximate eigenpair  $(\epsilon^{(k)}, \vec{\psi}^{(k)})$  of a discretized Hamiltonian matrix  $H$ ,  $\epsilon^{(k)}$  given by the Rayleigh Quotient of  $\vec{\psi}^{(k)}$  at step  $k$ , we look for an improved approximation

$\vec{\psi}^{(k+1)}$ . Specifically, we write

$$\vec{\psi}^{(k+1)} = \frac{1}{\xi}(\vec{\psi}^{(k)} + \delta\vec{\psi}^{(k)})$$

where the correction  $\delta\vec{\psi}^{(k)}$  is chosen orthogonal to  $\vec{\psi}^{(k)}$ .  $\xi$  is a normalization factor. Improving  $\vec{\psi}^{(k)}$  by RQI requires then to find  $\delta\vec{\psi}^{(k)} \perp \vec{\psi}^{(k)}$  and  $\xi$  such that

$$\left(H - \epsilon^{(k)}\right) \frac{1}{\xi}(\vec{\psi}^{(k)} + \delta\vec{\psi}^{(k)}) = \vec{\psi}^{(k)}. \quad (22)$$

We eliminate  $\xi$  by projecting the whole equation onto  $(\vec{\psi}^{(k)})^\perp$  and then rewrite it as

$$(I - \vec{\psi}^{(k)}(\vec{\psi}^{(k)})^T) (H - \epsilon^{(k)}) \delta\vec{\psi}^{(k)} = - (H - \epsilon^{(k)}) \vec{\psi}^{(k)}. \quad (23)$$

By the properties of the Rayleigh Quotient, the projector on the right hand side of the equation has been omitted.

**Remark 2.8** Eq.(23) is the same equation used to define the iterative corrections in the Jacobi-Davidson method (Sleijpen and Van Der Vorst, 1996).

This algorithm can be generalized to the simultaneous search of  $N$  eigenfunctions, with possible degeneracy of some eigenvalues (Descloux et al., 1998; Fattebert, 1998). The idea is to replace the Rayleigh quotient by the Rayleigh-Ritz (RR) procedure (e.g. Parlett, 1998, Ch. 11) and to look for corrections orthogonal to the whole subspace  $\Psi^{(k)}$  of trial eigenfunctions. Denoting  $\Psi^{(k)} = (\vec{\psi}_1^{(k)}, \dots, \vec{\psi}_N^{(k)})$  as a matrix made of vector columns  $\vec{\psi}_j^{(k)}$ , we can write the following iterative algorithm:

**Algorithm 2.1** (1) Let  $\Psi^{(0)}$  be a trial subspace of dimension  $N$ .

(2) For  $k = 0, 1, 2, \dots$ , do:

(a) Rayleigh-Ritz for  $H$  in the subspace  $\Psi^{(k)} \rightarrow (\epsilon_j^{(k)}, \vec{\psi}_j^{(k)}), j = 1, \dots, N$

(b) For  $j = 1, \dots, N$ , compute  $\delta\vec{\psi}_j^{(k)} \perp \Psi^{(k)}$  solution of

$$(I - \Psi^{(k)}\Psi^{(k)T})(H - \epsilon_j^{(k)})\delta\vec{\psi}_j^{(k)} = -(H - \epsilon_j^{(k)})\vec{\psi}_j^{(k)}, \quad (24)$$

(c) Define  $\Psi^{(k+1)} = (\vec{\psi}_1^{(k)} + \delta\vec{\psi}_1^{(k)}, \dots, \vec{\psi}_N^{(k)} + \delta\vec{\psi}_N^{(k)})$ .

An exact solution of Eq.(24) would lead to a locally quadratic convergence rate close to the solution of the eigenvalue problem for a non self-consistent (i.e. linear) Hamiltonian as proved in Fattebert (1998). Such an algorithm

has actually been applied in Fattebert (1996, 1999), where the linear systems are solved by multigrid. Slightly different versions have been proposed by Jin et al. (1999) and Ancilotto et al. (1999) who omit in particular the projector in Eq.(24). These approaches have to deal with the difficult question of how to define meaningful potentials — and sometimes eigenfunctions — on very coarse grids. Briggs et al. (1996) proposed to keep only the Laplacian in the Hamiltonian operator on the coarse grids. In this approach, the multigrid V-cycles can be seen as a preconditioner or inexact solver. Such an approach can be very efficient for self-consistent Hamiltonians, since an accurate solution for Eq.(24) is not always useful when the operator changes at each iteration. This point of view has been more precisely formulated in Fattebert and Bernholc (2000) where the potential operator is used only once at the beginning of the multigrid cycle to compute the residual. The main advantage of the latter approach is that the operator in Eq.(24) then does not depend on  $j$  and can be used in any nonorthogonal representation  $\Phi$  of the trial subspace.

Let us now focus on the preconditioning approach. Looking at the correction equation (24), we note that the right hand side is the steepest descent direction for the minimization problem with orthonormality constraints associated with the KS eigenvalue problem (3). We note it  $(-\vec{r}_j^{(k)}, \vec{r}_j^{(k)})$  being the residual of the eigenvalue problem. In a SD approach,  $\delta\psi_j^{(k)}$  would be given by  $-\vec{r}_j^{(k)}/\epsilon_{max}$  where  $\epsilon_{max}$  is the largest eigenvalue of  $H$ . However, from the point of view of the inverse iteration method, an *optimal* correction is given by

$$\delta\psi_j^{(k)} = -((I - \Psi^{(k)}\Psi^{(k)T})(H - \epsilon_j^{(k)})\Big|_{\Psi^{(k)\perp}})^{-1}\vec{r}_j^{(k)}. \quad (25)$$

Thus one can consider a preconditioner  $K$  that approximates the operator

$$(I - \Psi^{(k)}\Psi^{(k)T})(H - \epsilon_j^{(k)})\Big|_{\Psi^{(k)\perp}}. \quad (26)$$

A close look at the Hamiltonian operator shows that for high energy states, the Laplacian is the dominant part, and the corresponding eigenfunctions are essentially similar to those of the Laplacian, *i.e.* plane waves perturbed by a relatively weak potential. It means that the operator  $-\frac{1}{2}\nabla^2$  is a good approximation of  $(H - \epsilon_j^{(k)})$  in  $\Psi^{(k)\perp}$  for  $1 \leq j \leq N$ , at least close to convergence, so that one can choose

$$K \sim (I - \Psi^{(k)}\Psi^{(k)T})(-\frac{1}{2}\nabla^2)\Big|_{\Psi^{(k)\perp}} = (I - \Phi^{(k)}S^{-1}\Phi^{(k)T})(-\frac{1}{2}\nabla^2)\Big|_{\Phi^{(k)\perp}}. \quad (27)$$

In real-space, one can associate frequencies with grid resolution. Applying a single grid iterative method — like Jacobi or Gauss-Seidel — to solve a Poisson problem, one essentially obtains the high frequency components of the

solution, the one that we cannot represent on a coarser grid. Using multigrid V-cycles based on such a *smoother*, we can solve the problem for components of lower frequencies by visiting coarser grids (Brandt, 1977). Furthermore, by choosing a limited number of grids, one can *select* the components that we want to solve for. Following this heuristic argument, we define the application of the preconditioner  $K^{-1}$  to  $\vec{r}_j^{(k)}$  as an iterative multigrid solver for the Poisson problem (Fattebert and Bernholc, 2000)

$$-\frac{1}{2}\nabla^2\delta\psi_j^{\vec{(k)}} = -\vec{r}_j^{\vec{(k)}} \quad (28)$$

limited to the finest grids. For practical calculations, using 2 coarse grids is often optimal. We start the process with an initial trial solution  $\delta\psi_j^{\vec{(k)}} = -\alpha\vec{r}_j^{\vec{(k)}}$ . Its main goal is to introduce some low frequency components in the correction  $\delta\psi_j^{\vec{(k)}}$ . The coefficient  $\alpha$  is defined by looking at the initial guess as the steepest descent correction one would make if the whole calculation was done on the coarse grids not visited during the V-cycles. As a smoother in the V-cycles, the Jacobi method is appropriate because of its inherent parallelism.

In Fig. 3, we present the convergence history of the error on the total energy for various discretization grids for an 8 atoms diamond cell self-consistent calculation. We use the PSD algorithm with the multigrid preconditioner described above, doing for each correction  $\delta\psi_j^{\vec{(k)}}$  and at each SC iteration 1 V-cycle with 2 pre-smoothing and 2 post-smoothing. The grid-independence of the convergence rate is observed. All the calculations use the same coarsest grid  $6 \times 6 \times 6$  for the multigrid preconditioning, and the same total number of states (16 occupied+ 8 unoccupied)  $N = 24$ . The initial trial functions are random functions.

**Remark 2.9** *Preconditioners based on a similar idea have been developed for PW calculations (Teter et al., 1989; Fernando et al., 1989; Chetty et al., 1995). Since the numerical basis functions in PW are eigenfunctions of the Laplacian operator, efficient simple diagonal preconditioners can be designed for PW.*

A different preconditioner was proposed by Saad et al. (1996) in conjunction with a Lanczos algorithm. Realizing that the eigenfunctions corresponding to the lowest eigenvalues are in general smoother than the others, they proposed to apply a low frequency filter directly to the trial eigenfunctions. This is done on a single grid in real-space by an averaging of the value of a function at every grid point with the values at its neighboring points. They note however that this preconditioner is probably not always sufficient, in particular when a large number of eigenfunctions is required and the highest eigenvalues of interest correspond to eigenfunctions presenting a lower degree of smoothness.

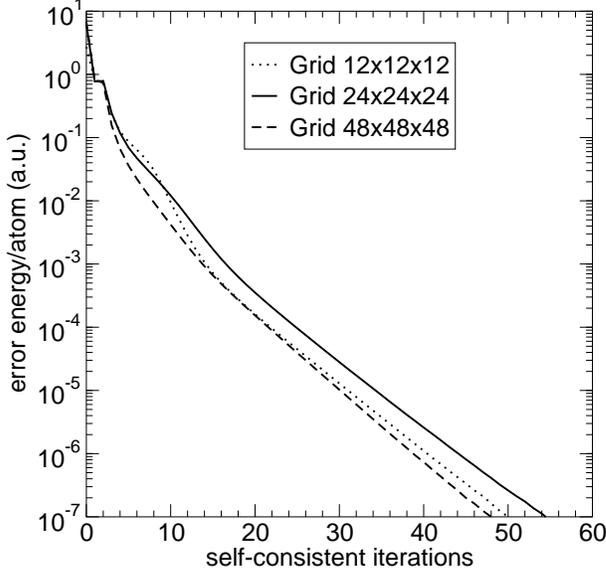


Fig. 3. Convergence rate for a diamond cell calculation (8 carbon atoms) for 3 different discretization grids.

## 2.6 Energy and forces

To optimize molecular geometries, or run molecular dynamics and measure physical quantities at finite temperature, it is important to be able to compute the forces acting on the atoms in any configuration. To derive expressions for these forces, we start by some considerations on the total energy of a physical system. In DFT, the total energy of a system can be expressed as the sum of three terms

$$E_t = E_{KS}[\Phi, \{\vec{R}_a\}_{a=1}^{N_a}] + E_{ions}[\{\vec{R}_a\}_{a=1}^{N_a}] + \frac{1}{2} \sum_{a=1}^{N_a} M_a \dot{\vec{R}}_a^2 \quad (29)$$

where  $E_{ions}$ , the electrostatic energy between ions of charges  $Z_i$ , is given by

$$E_{ions} = \frac{1}{2} \sum_{a,b=1, a \neq b}^{N_a} \frac{Z_a Z_b}{|\vec{R}_a - \vec{R}_b|}, \quad (30)$$

and  $M_a$  denotes the mass of the ion in  $\vec{R}_a$ . In the KS energy  $E_{KS}$ , the contribution due to the interaction between electrons and ions is given by the sum of two terms associated to the local and nonlocal parts of the pseudopotential

$$E_{ps} = \int_{\Omega} v_{ps}^{local}(\vec{r}) \rho_e(\vec{r}) d\vec{r} + Tr(\Phi^T V_{nl} \Phi \rho^{(\Phi)}) = E_{ps}^{local} + E_{ps}^{nl}. \quad (31)$$

It is computationally more efficient to compute the electrostatic term  $E_{es}$  by solving a Poisson problem. In order to deal with a neutral charge, it is a standard procedure to add to the system smeared core charges centered at atomic sites,

$$\rho_a(\vec{r}) = -\frac{Z_a}{(\sqrt{\pi}r_c^a)^3} \exp\left(-\frac{|\vec{r} - \vec{R}_a|^2}{(r_c^a)^2}\right). \quad (32)$$

The sum of these charges,  $\rho_s$ , neutralizes the electronic charge by generating a total potential

$$v_s(\vec{r}) = \sum_{a=1}^{N_a} \frac{-Z_a}{|\vec{r} - \vec{R}_a|} \operatorname{erf}\left(\frac{|\vec{r} - \vec{R}_a|}{r_c^a}\right). \quad (33)$$

We then compute the Hartree potential  $v_H$  as the solution of a Poisson problem for a neutral total charge  $\rho_e + \rho_s$ ,

$$-\nabla^2(v_H + v_s)(\vec{r}) = 4\pi(\rho_e + \rho_s)(\vec{r}) \quad (34)$$

with periodic or Dirichlet boundary conditions. This problem can be efficiently solved on the discretization grid in  $O(N)$  operations by the multigrid method (Brandt, 1977).

With the introduction of smeared neutralizing core charges, one can write

$$E_{ions} = \frac{1}{2} \sum_{a,b=1}^{N_a} \int_{R^3} \frac{\rho_a(\vec{r} - \vec{R}_a)\rho_b(\vec{r}' - \vec{R}_b)}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' - E_{self} + E_{diff} \quad (35)$$

where  $E_{self}$  is the self-interaction of the core charges,

$$E_{self} = \frac{1}{2} \sum_{a=1}^{N_a} \int_{R^3} \frac{\rho_a(\vec{r} - \vec{R}_a)\rho_a(\vec{r}' - \vec{R}_a)}{|\vec{r} - \vec{r}'|} d\vec{r} = \frac{1}{\sqrt{2\pi}} \sum_{a=1}^{N_a} \frac{Z_a^2}{(r_c^a)^2}, \quad (36)$$

and

$$\begin{aligned} E_{diff} &= \frac{1}{2} \sum_{a,b=1, a \neq b}^{N_a} \left[ \frac{Z_a Z_b}{|\vec{R}_a - \vec{R}_b|} - \int_{R^3} \frac{\rho_a(\vec{r} - \vec{R}_a)\rho_b(\vec{r}' - \vec{R}_b)}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' \right] \\ &= \sum_{a,b=1, a < b}^{N_a} \frac{Z_a Z_b}{|\vec{R}_a - \vec{R}_b|} \operatorname{erfc}\left(\frac{|\vec{R}_a - \vec{R}_b|}{\sqrt{(r_c^a)^2 + (r_c^b)^2}}\right). \end{aligned} \quad (37)$$

We then have, for  $r_c^a$  sufficiently small compared to  $\Omega$ ,

$$\begin{aligned}
& E_{es} + E_{ps}^{local} + E_{ions} \\
&= \frac{1}{2} \int_{\Omega} \frac{\rho_e(\vec{r})\rho_e(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' + \int_{\Omega} v_{ps}^{local} \rho_e(\vec{r}) d\vec{r} \\
&+ \frac{1}{2} \sum_{a,b=1, a \neq b}^{N_a} \int_{\mathbb{R}^3} \frac{\rho_a(\vec{r} - \vec{R}_a)\rho_b(\vec{r}' - \vec{R}_b)}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' - E_{self} + E_{diff} \\
&\approx \frac{1}{2} \int_{\Omega} \frac{(\rho_e(\vec{r}) + \rho_s(\vec{r}))(\rho_e(\vec{r}') + \rho_s(\vec{r}'))}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' - \int_{\Omega} \frac{\rho_e(\vec{r})\rho_s(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' \\
&+ \int_{\Omega} v_{ps}^{local}(\vec{r})\rho_e(\vec{r})d\vec{r} - E_{self} + E_{diff} \\
&= \frac{1}{2} \int_{\Omega} (\rho_e(\vec{r}) + \rho_s(\vec{r}))(v_H + v_s)(\vec{r})d\vec{r} \\
&+ \int_{\Omega} (v_{ps}^{local} - v_s)(\vec{r})\rho_e(\vec{r})d\vec{r} - E_{self} + E_{diff}. \tag{38}
\end{aligned}$$

Knowing the ground state electronic structure for a given atomic configuration  $\{\vec{R}_a\}_{a=1}^{N_a}$ , one can compute the internal force acting on the ion  $I$  by deriving the total energy with respect to the atomic coordinates  $\vec{R}_I$ ,

$$\vec{F}_I = -\frac{d}{d\vec{R}_I} E_t(\Phi, \{\vec{R}_a\}_{a=1}^{N_a}). \tag{39}$$

Using the property that  $\Phi$  is the minimum of the functional  $E$ , one shows that

$$\vec{F}_I = -\frac{\partial}{\partial \vec{R}_I} E_t(\Phi, \{\vec{R}_a\}_{a=1}^{N_a}) \tag{40}$$

(Hellmann-Feynman forces, Feynman (1939)). Since the electronic structure does not explicitly depend on the atomic positions, Eq.(40) means that the forces can be computed from a single ground state calculation, by deriving the atomic potentials only.

**Remark 2.10** *To obtain Eq. (40), we also use the fact that the numerical representation of  $\Psi$  does not explicitly depend on the atomic positions since the grid is atom independent. For atom-centered orbitals moving with the atoms, this is not true anymore and additional terms (Pulay forces) have to be included.*

From Eq.(38) and (40), using Eq. (34), we obtain the total force acting on atom  $I$  in the form

$$\begin{aligned} \vec{F}_I = & \int_{\Omega} (v_H(\vec{r}) + v_s(\vec{r})) \frac{d}{d\vec{R}_I} \rho_s d\vec{r} + \int_{\Omega} \frac{d}{d\vec{R}_I} (v_{ps}^{local} - v_s) \rho_e(\vec{r}) d\vec{r} \\ & + \frac{\partial}{\partial \vec{R}_I} Tr(\Phi^T V_{nl} \Phi \rho^{(\Phi)}) + \frac{d}{d\vec{R}_I} E_{diff}. \end{aligned} \quad (41)$$

Writing the forces in this form lets appear the functions  $(v_{ps}^{local} - v_s)$  and  $\rho_s$  which are localized in real-space. This can be directly used to reduce the complexity of the computation of the forces on a grid. In principle all the derivatives with respect to  $\vec{R}_I$  in Eq.(41) can be computed analytically. In practice, because of the filtering of the pseudopotentials, the derivatives have to be evaluated numerically on the filtered pseudopotentials.

**Remark 2.11** *The sum of the forces over all the atoms should be zero if no external force is applied. In practice, the use of a finite grid introduces small errors (see Sec.2.4) and provides an estimate of the accuracy of the forces.*

## 2.7 Born-Oppenheimer molecular dynamics

To perform Born-Oppenheimer molecular dynamics simulations of quantum systems described by the KS equations, we compute the forces acting on the ions according to Eq.(41) and let the system evolve accordingly. The ions evolve like classical particles surrounded by quantum electrons (Born-Oppenheimer approximation). The error in the energy is second order with respect to the error in the electronic wave functions, but the error in the forces is first order. It means that one should be particularly careful in the computation of the ground state of the KS energy functional for the each atomic configuration at each iteration. It is particularly important to have accurate forces to ensure a perfect conservation of the total energy of the system in a microcanonical simulation.

As shown by Jing et al. (1994) and Briggs et al. (1996), the computation of the forces in FD methods is accurate enough to allow for energy conserving microcanonical ab initio simulations. This is illustrated in Fig. 4 where we show the evolution of the energy during a molecular dynamics simulation of a  $\text{Si}_5$  cluster. To avoid any systematic drift of the energy due to the integration scheme, the equations of motions were integrated numerically using the time reversible Verlet's second order algorithm (e.g. Heermann, 1990, Chap.3)

$$\vec{R}_I^{(n+1)} = 2\vec{R}_I^{(n)} - \vec{R}_I^{(n-1)} + \vec{F}_I^{(n)} (\Delta t)^2 / M_I$$

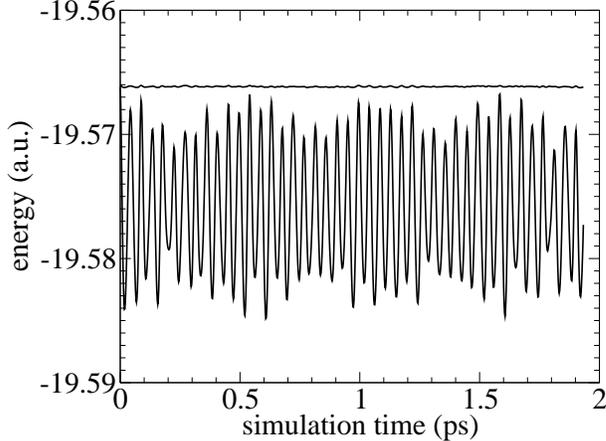


Fig. 4. Molecular dynamics simulation of a  $\text{Si}_5$  cluster. The total energy (top) shows fluctuations of the order  $10^{-4}a.u.$ , but no systematic drift. The KS energy is also plotted (bottom).

A grid spacing  $h = 0.56\text{Bohr}$  and a time step  $\Delta t = 80a.u.$  were used for this simulation.

## 2.8 Localized orbitals

Formulating the minimization problem in terms of non-orthogonal functions  $\Phi$  instead of the Ritz functions  $\Psi$ , one can formally impose localization constraints on  $\Phi$  to reduce the cost of the calculation. This is one of the most popular method to obtain a linear scaling of the computational cost with respect to the size of the system. These so called Order N methods are discussed in another article (Goedecker, this book, e.g.) and we will limit here the discussion to the grid-based method exposed in this article.

On a real-space grid, spatial localization can be imposed by forcing each orbital to be zero outside of a spherical region centered on a particular ion (Hoshi and Fujiwara, 1997; Fattebert and Bernholc, 2000). Such a truncation will linearize the computational cost of  $D^{(\Phi)}$  in Eq. (16) and  $\rho_e$  in Eq. (11), the most expensive operations in the minimization algorithm. It also reduces to  $O(N)$  the storage requirements for the wave functions. Of course, this reduction in computational cost does not come for free. It introduces some approximation error that one expects to keep within a certain tolerance. It means in particular that one cannot choose the localization regions arbitrarily small.

The application of the compact FD Laplacian operator to a wave function localized in a sphere of radius  $R_c$  generates a function localized in a sphere of radius  $R_c + h\sqrt{2}$ , which is used as the localization radius for  $H\phi_j$ . This truncation suppresses some components of  $H\phi_j$  that are generated by the non-local, short-range projectors of the pseudopotential operator. These components are

lost in the correction of the wave functions since the latter should remain localized. However, they are included exactly in the computation of the matrix  $\Theta$  and the total energy by writing  $H^{(\Phi)}$  as the sum of two matrices

$$H^{(\Phi)} = \Phi^T(H - V_{nl})\Phi + \Phi^T V_{nl}\Phi,$$

which isolates the non-local potential  $V_{nl}$  in the second term. This second term is easily computed in  $O(N)$  operations using only the nonzero terms  $\langle \phi_j | v_i \rangle$ . Therefore, the only approximation in this approach is the use of a localization radius to limit the spatial extent of each non-orthogonal orbital.

Since the eigenfunctions are in general not localized, the matrix  $C$  that solves Eq.(8) is not sparse and the computation of  $C$  requires  $O(N^3)$  operations. To linearize the cost of the whole calculation, one could impose localization constraints on the density matrix, requiring  $\bar{\rho}_{ij}^{(\Phi)} = 0$  if the localization regions of orbitals  $i$  and  $j$  are separated by a distance larger than a truncation radius  $R_\rho$  as in Hernandez and Gillan (1995). This can be imposed at each step of the iterative minimization in order to achieve linear scaling. Such a truncature is justified by the exponential decay of  $\rho(\vec{r}, \vec{r}')$  as  $|\vec{r} - \vec{r}'| \rightarrow \infty$  in insulators or metals at a finite temperature (Ismael-Beigi and Arias, 1999). However, for  $M \gg N$ , the full evaluation of  $S^{-1}$ ,  $\Theta$  or  $\bar{\rho}^{(\Phi)}$ , even with an order  $N^3$  algorithm, constitutes a small fraction of the total calculations for a large range of system sizes. Since a good accuracy can be obtained only by keeping the number of non-zero elements in  $\bar{\rho}$  much larger than in  $S$  (Millam and Scuseria, 1997), using the sparsity of  $\bar{\rho}$  does not lead to much gain in this context.

**Remark 2.12** *If an exact and explicit  $O(N^3)$  diagonalization is performed, in Eq.(8), partially occupied and unoccupied orbitals can be used, which permits calculations for metallic as well as semiconducting systems. However, calculations for metals may require more localized orbitals or larger localization radii for an accurate description of their electronic structure.*

For systems with  $N > 1000$ , solving Eq.(8) on a single processor becomes very expensive, if required at each self-consistent iteration. However, for fully parallel calculations, it is natural to also parallelize the  $N \times N$  submatrices operations. This can be done using standard libraries, such as PBLAS (Parallel Basic Linear Algebra Subprograms) and ScaLapack (Blackford et al., 1997).

**Remark 2.13** *According to PBLAS and ScaLapack requirements,  $S$ ,  $H^{(\Phi)}$ ,  $\Theta$  and  $\bar{\rho}^{(\Phi)}$  are stored as full  $N \times N$  matrices, distributed among the processors. Although most of the operations on these matrices can be optimized using their sparsity — except the diagonalization in Eq. (9)— the full storage approach is adequate for a substantial range of calculations. It is also the easiest implementation given the available standard numerical libraries for distributed memory multiprocessors computers. The solution of the eigenvalue problem (9) is clearly the dominant part of these  $O(N^3)$  operations.*

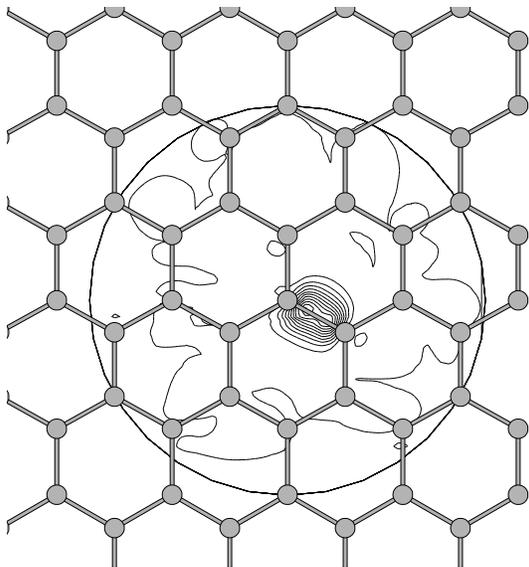


Fig. 5. Contour plot of the square of a typical localized orbital in the plane defined by the cylindrical surface of a (5,5) nanotube. The external circle shows the localization region (radius 8 Bohr).

In the iterative minimization of the KS energy functional, the truncature of the orbitals modifies the correction directions in a way that can slow down the convergence process. On the other hand, the localization constraints break the invariance in the representation of the occupied subspace and may generate multiple local minimas for  $E_{KS}$  (Goedecker, this book). One way to deal with these issues is to choose localization radii large enough so that one can easily end up in a minima of energy close enough to the *true* global minima — the one obtained without localization constraints. For example, Fattebert and Bernholc (2000) were able to compute accurately energy differences in a carbon nanotube using a localization radius of 8 Bohr.

Since the method described above allows to determine the eigenfunctions of the Kohn-Sham equations in a basis of localized functions — according to Eq.(7)—, it can be considered as a generalization of *ab initio* methods that use a linear combination of atomic orbitals (LCAO) to expand the eigenfunctions:  $\psi_j = \sum_i c_i \phi_i$ . The main difference is that grid-based local functions  $\phi_i$  are defined by their values on a grid and are variationally optimized according to their environment. In particular, the functions  $\phi_i$  have many more degrees of freedom and one can systematically increase the accuracy of the calculations by mesh refinement or expansion of the localization domain. An example of such an orbital computed for a (5,5) carbon nanotube is plotted in Fig.5. The total number of basis functions in high precision calculations is thus much smaller than in LCAO approaches and minimizes the  $O(N^3)$  part.

### 3 Quantum transport

#### 3.1 Electron transmission and Green's functions

Let us consider a system composed of a conductor,  $C$ , connected to two semi-infinite leads,  $R$  and  $L$ , as in Fig. 6. A fundamental result in the theory of electronic transport is that the conductance through a region of interacting electrons (the  $C$  region in Fig. 6) is related to the scattering properties of the region itself via the Landauer formula (Landauer, 1970)

$$\mathcal{C} = \frac{2e^2}{h} \mathcal{T}, \quad (42)$$

where  $\mathcal{T}$  is the transmission function and  $\mathcal{C}$  is the conductance. The former represents the probability that an electron injected at one end of the conductor will transmit to the other end. In principle, we can compute the transmission function for a coherent conductor<sup>1</sup> starting from the knowledge of the scattering matrix,  $\mathcal{S}$ . The latter is the mathematical quantity that describes the response at one lead due an excitation at another. In principle, the scattering matrix can be uniquely computed from the solution of the Schroedinger equation and would suffice to describe the transport processes we are interested in this work. However, it is a general result of conductance theory that the elements of the  $\mathcal{S}$ -matrix can be expressed in terms of the Green's function of the conductor (Datta, 1995; Fisher and Lee, 1981; Meir and Wingreen, 1992) which, in practice, can be sometimes simpler to compute.

Let us consider a physical system represented by an Hamiltonian  $H$ . Its Green's function for an energy  $E$  is defined by the equation

$$(E \pm i\eta - H)G(\vec{r}, \vec{r}') = \delta(\vec{r}, \vec{r}') \quad (43)$$

where  $i\eta > 0$  is an infinitesimal imaginary part added to the energy to incorporate the boundary conditions into the equation. The solution with  $+$  sign is the retarded Green's function  $G^r$ , while the solution with  $-$  sign is called advanced Green's function  $G^a$ . The transmission function can then be expressed in terms of the Green's functions of the conductors and the coupling of the conductor to the leads in a simple manner (see Datta, 1995, p.141 and ff.)

$$\mathcal{T} = \text{Tr}(\Gamma_L G_C^r \Gamma_R G_C^a), \quad (44)$$

---

<sup>1</sup> A conductor is said to be coherent if it can be characterized by a transmission matrix that relates each of the outgoing wave amplitudes to the incoming wave amplitudes at a given energy

where  $G_C^{\{r,a\}}$  are the retarded and advanced Green's functions of the conductor, and  $\Gamma_{\{L,R\}}$  are functions that describe the coupling of the conductor to the leads.

In the following we are going to restrict the discussion to discrete systems that we can describe by ordinary matrix algebra. More precisely, we are going to work with matrices representing a physical system in a basis of localized electronic orbitals centered on the atoms constituting the system. It includes in particular the tight-binding model.

For a discrete media, the Green's function is then solution of a matrix equation

$$(\epsilon - H)G = I \quad (45)$$

where  $\epsilon = E \pm i\eta$  with  $\eta$  arbitrarily small and  $I$  is the identity matrix. To simplify the notations, we drop the exponent  $\{a,r\}$  referring to advanced and retarded functions when implicitly defined by  $\epsilon$ . For an open system, consisting of a conductor and two semi-infinite leads (see Fig. 6), the above Green's function can be partitioned into sub-matrices that correspond to the individual subsystems

$$\begin{pmatrix} g_L & g_{LC} & g_{LCR} \\ g_{CL} & G_C & g_{CR} \\ g_{LRC} & g_{RC} & g_R \end{pmatrix} = \begin{pmatrix} (\epsilon - h_L) & -h_{LC} & 0 \\ -h_{LC}^* & (\epsilon - H_C) & -h_{CR} \\ 0 & -h_{CR}^* & (\epsilon - h_R) \end{pmatrix}^{-1}, \quad (46)$$

where the matrix  $(\epsilon - H_C)$  represents the finite "isolated" conductor (with no coupling elements to the leads),  $(\epsilon - h_{\{R,L\}})$  represent the semi-infinite leads, and  $h_{CR}$  and  $h_{LC}$  are the coupling matrices between the conductor and the leads, and  $h^*$  denotes the familiar conjugate transpose of  $h$ . As a convention, we use lower case letters for (semi-)infinite matrices and upper case for finite dimension matrices. In Eq.(46) we have made the assumption that there is no direct interaction between the left and right leads. From this equation it is straightforward to obtain an explicit expression for  $G_C$

$$G_C = (\epsilon - H_C - \Sigma_L - \Sigma_R)^{-1} \quad (47)$$

where the finite-dimension matrices

$$\Sigma_L = h_{LC}^*(\epsilon - h_L)^{-1}h_{LC}, \quad \Sigma_R = h_{RC}(\epsilon - h_R)^{-1}h_{RC}^* \quad (48)$$

are defined as the self-energy terms due to the semi-infinite leads. The self-energy terms can be viewed as effective Hamiltonians that arise from the



Fig. 6. A conductor described by the Hamiltonian  $H_C$ , connected to two semi-infinite leads  $L$  and  $R$ , through the coupling matrices  $h_{LC}$  and  $h_{CR}$ .

coupling of the conductor with the leads. The coupling functions  $\Gamma_{\{L,R\}}$  can then be obtained as (Datta, 1995)

$$\Gamma_{\{L,R\}} = i[\Sigma_{\{L,R\}}^r - \Sigma_{\{L,R\}}^a], \quad (49)$$

where the advanced self-energy  $\Sigma_{\{L,R\}}^a$  is the conjugate transpose of the retarded self-energy  $\Sigma_{\{L,R\}}^r$ . The core of the problem lies in the calculation of the self-energies of the semi-infinite leads.

It is well known that any solid (or surface) can be viewed as an infinite (semi-infinite in the case of surfaces) stack of principal layers with nearest-neighbor interactions (Lee and Joannopoulos, 1981a,b). This corresponds to transforming the original system into a linear chain of principal layers. For a lead-conductor-lead system, the conductor can be considered as one principal layer sandwiched between two semi-infinite stacks of principal layers. The next three sections are devoted to the computation of the self-energies using the principal layers approach for different geometries.

### 3.2 Transmission through a bulk system.

Within the principal layer approach, the matrix elements of Eq.(45) between layer orbitals yield a series of matrix equations for the Green's functions

$$\begin{aligned} (\epsilon - H_{00})G_{00} &= I + H_{01}G_{10} \\ (\epsilon - H_{00})G_{10} &= H_{01}^*G_{00} + H_{01}G_{20} \\ &\dots \\ (\epsilon - H_{00})G_{n0} &= H_{01}^*G_{n-1,0} + H_{01}G_{n+1,0} \end{aligned} \quad (50)$$

where the finite dimension matrices  $H_{nm}$  and  $G_{nm}$  are formed by the matrix elements of the Hamiltonian and Green's function between the layer orbitals. We assume that in a bulk system  $H_{00} = H_{11} = \dots$  and  $H_{01} = H_{12} = \dots$ . Following Lopez-Sancho et al. (1984, 1985), this chain can be transformed in order to express the Green's function of one individual layer in terms of the Green's function of the preceding (or following) one. This is done via the

introduction of the transfer matrices  $T$  and  $\bar{T}$ , defined such that

$$G_{10} = TG_{00}$$

and

$$G_{00} = \bar{T}G_{10}.$$

Using these definitions, we can write the bulk Green's function as (Garcia-Moliner and Velasco, 1992)

$$G(E) = (\epsilon - H_{00} - H_{01}T - H_{01}^*\bar{T})^{-1}. \quad (51)$$

The transfer matrix can be easily computed from the Hamiltonian matrix elements via an iterative procedure, as outlined in Lopez-Sancho et al. (1984). In particular  $T$  and  $\bar{T}$  can be written as

$$\begin{aligned} T &= t_0 + \tilde{t}_0 t_1 + \tilde{t}_0 \tilde{t}_1 t_2 + \dots + \tilde{t}_0 \tilde{t}_1 \dots \tilde{t}_{n-1} t_n \\ \bar{T} &= \tilde{t}_0 + t_0 \tilde{t}_1 + t_0 t_1 \tilde{t}_2 + \dots + t_0 t_1 \dots t_{n-1} \tilde{t}_n \end{aligned}$$

where  $t_i$  and  $\tilde{t}_i$  are defined via the recursion formulas

$$\begin{aligned} t_i &= (I - t_{i-1} \tilde{t}_{i-1} - \tilde{t}_{i-1} t_{i-1})^{-1} t_{i-1}^2, \\ \tilde{t}_i &= (I - t_{i-1} \tilde{t}_{i-1} - \tilde{t}_{i-1} t_{i-1})^{-1} \tilde{t}_{i-1}^2 \end{aligned}$$

and

$$\begin{aligned} t_0 &= (\epsilon - H_{00})^{-1} H_{01}^*, \\ \tilde{t}_0 &= (\epsilon - H_{00})^{-1} H_{01}. \end{aligned}$$

The process is repeated until  $t_n, \tilde{t}_n \leq \delta$  with  $\delta$  arbitrarily small. Usually no more than 5 or 6 terms are required to converge the above sums.

In the hypothesis of leads and conductors being of the same material (bulk conductivity), we can identify one principal layer of the bulk system with the conductor  $C$ , so that  $H_{00} \equiv H_C$ . If we compare Eq.(51) with Eq. (47), we obtain the expression for the self-energies of the conductor-leads system

$$\Sigma_L = H_{01}^* \bar{T}, \quad \Sigma_R = H_{01} T. \quad (52)$$

The coupling functions are then obtained from the sole knowledge of the transfer matrices and the coupling Hamiltonian matrix elements:  $\Gamma_L = -\text{Im}(H_{01}^* \bar{T})$

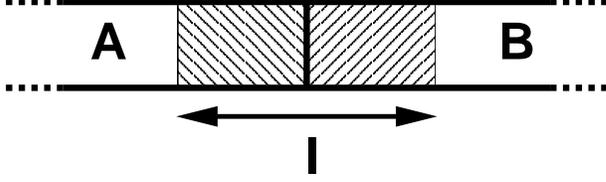


Fig. 7. Sketch of a system containing an interface between two media A and B.  $I$  is the interface region for which we need to compute the Green's function  $G_I$ .  $I$  is composed of two principal layers, one in each side of the interface (dashed area).

and  $\Gamma_R = -\text{Im}(H_{01}\bar{T})$  (Buongiorno Nardelli, 1999).

**Remark 3.1** *In the application of the Landauer formula, it is customary to compute the transmission probability from one lead to the other assuming that the leads are connected to a reflectionless contact whose electron energy distribution is known (see for instance Datta, 1995, p. 59 and ff.).*

### 3.3 Transmission through an interface.

The procedure outlined above can also be applied in the case of electron transmission through an interface between two different media A and B. To study this case we make use of the Surface Green's Function Matching (SGFM) theory, pioneered by Garcia-Moliner and Velasco (1991, 1992).

We have to compute the Green's function  $G_I$ , where the subscript  $I$  refers to the interface region composed of two principal layers — one in each media — (see Fig.7). Using the SGFM method,  $G_I$  is calculated from the bulk Green's function of the isolated systems,  $G_A$  and  $G_B$ , and the coupling between the two principal layers at the two sides of the interface,  $H_{AB}$  and  $H_{BA}$ . Via the calculation of the transmitted and reflected amplitudes of an elementary excitation that propagates from medium A to medium B, it can be shown that the interface Green's function obeys the equation (Garcia-Moliner and Velasco, 1992, Ch.4)

$$G_I = \begin{pmatrix} G_{AA} & G_{AB} \\ G_{BA} & G_{BB} \end{pmatrix} = \begin{pmatrix} \epsilon - H_{00}^A - (H_{01}^A)^* \bar{T} & -H_{AB} \\ -H_{BA} & \epsilon - H_{00}^B - H_{01}^B T \end{pmatrix}^{-1}. \quad (53)$$

Once the interface Green's function is known, we can compute the transmission function in terms of block super-matrices

$$\mathcal{T}(E) = \text{Tr}(\Gamma_A G_{AB}^r \Gamma_B G_{BA}^a)$$

with  $\Gamma_{\{A,B\}} = i[\Sigma_{\{A,B\}}^r - \Sigma_{\{A,B\}}^a]$ ,  $\Sigma_{\{A,B\}}$  given by the analogous of Eq.(52)

for the two semi-infinite sections, and  $G_{BA}^a = (G_{AB}^r)^*$  (Buongiorno Nardelli, 1999).

### 3.4 Transmission through a left lead-conductor-right lead (LCR) system.

Within the SGFM framework, the approach described in the previous section can be extended to the case of multiple interfaces and superlattices (Garcia-Moliner and Velasco, 1991, 1992) with little complication. For the calculation of conductances in realistic experimental geometry, the method can be expanded to the general configuration of a Left-lead-Conductor-Right-lead (LCR) systems — as displayed in Fig.6. In the language of block matrices and principal layers, outlined in the previous sections, the LCR Green's function obeys the equation

$$G_{LCR} = \begin{pmatrix} G_L & G_{LC} & G_{LR} \\ G_{CL} & G_C & G_{CR} \\ G_{RL} & G_{RC} & G_R \end{pmatrix} = \begin{pmatrix} \epsilon - H_{00}^L - (H_{01}^L)^* \bar{T} & -H_{LC} & 0 \\ -H_{CL} & \epsilon - H_C & -H_{CR} \\ 0 & -H_{RC} & \epsilon - H_{00}^R - H_{01}^R T \end{pmatrix}^{-1}. \quad (54)$$

where  $H_{nm}^{\{L,R\}}$  are the block matrices of the Hamiltonian between the layer orbitals in the left and right leads respectively, and  $T_{\{L,R\}}$  and  $\bar{T}_{\{L,R\}}$  are the appropriate transfer matrices. The latter are easily computed from the Hamiltonian matrix elements via the iterative procedure already described in the bulk case (Sec.3.2). Correspondingly,  $H_{LC}$  and  $H_{CR}$  are the coupling matrices between the conductor and the leads principal layers in contact with the conductor.

As in Sec.3.1, it is straightforward to obtain in the form of Eq.(47),  $G_C = (\epsilon - H_C - \Sigma_L - \Sigma_R)^{-1}$ , where  $\Sigma_L$  and  $\Sigma_R$  are the self-energy terms due to the semi-infinite leads, and identify (Buongiorno Nardelli and Bernholc, 1999)

$$\begin{aligned} \Sigma_L &= H_{LC}^* (\epsilon - H_{00}^L - (H_{01}^L)^* \bar{T}_L)^{-1} H_{LC}, \\ \Sigma_R &= H_{CR} (\epsilon - H_{00}^R - H_{01}^R T_R)^{-1} H_{CR}^*. \end{aligned} \quad (55)$$

The transmission function in the LCR geometry can then be derived from

Eq.(44) and (49).

**Remark 3.2** *The knowledge of the conductor's Green's function  $G_C$  gives also direct information on the electronic spectrum of the system via the spectral density of electronic states*

$$N(E) = -(1/\pi)\text{Im}[\text{Tr}(G_C(E))].$$

**Remark 3.3** *We have assumed a truly one-dimensional chain of principal layers, which is physical only for systems like nanotubes or quantum wires that have a definite quasi-one-dimensional character. The extension to a truly three-dimensional case is straightforward using Bloch functions wave vectors  $\vec{k}_{\parallel}$  parallel to the layers (in the directions perpendicular to the conduction). The introduction of the principal layer concept implies that along the direction of the conduction the system is described by an infinite set of wave vectors  $\vec{k}_{\perp}$ . The above procedure effectively reduces the three-dimensional system to a set of non-interacting linear-chains, one for each  $\vec{k}_{\parallel}$  (Lee and Joannopoulos, 1981a,b). We can then use the usual  $k$ -point summation techniques to evaluate, for instance, the quantum conductance*

$$T(E) = \sum_{\vec{k}_{\parallel}} w_{\vec{k}_{\parallel}} T_{\vec{k}_{\parallel}}(E)$$

where  $w_{\vec{k}_{\parallel}}$  are the relative weights of the different wave vectors  $\vec{k}_{\parallel}$  in the irreducible wedge of the surface Brillouin zone (Baldereschi, 1973).

### 3.5 Generalization to nonorthogonal orbitals.

In the previous sections we have assumed to have a Hamiltonian representation in terms of orthogonal orbitals. The expression for the Green's and transmission functions of a bulk system described by a general non-orthogonal localized-orbital Hamiltonian follows directly from the procedure outlined in Section 3.2. All the quantities can be obtained making the substitutions  $(\epsilon - H_{00}) \rightarrow (\epsilon S_{00} - H_{00})$  and  $H_{01} \rightarrow -(\epsilon S_{01} - H_{01})$ . Here, we introduce the matrices  $S$  that represent the overlap between the localized orbitals. With this recipe, the equation chain (50) now reads

$$\begin{aligned} (\epsilon S_{00} - H_{00})G_{00} &= I - (\epsilon S_{01} - H_{01})G_{10}, \\ (\epsilon S_{00} - H_{00})G_{10} &= -(\epsilon S_{01}^* - H_{01}^*)G_{00} - (\epsilon S_{01} - H_{01})G_{20}, \\ &\dots \\ (\epsilon S_{00} - H_{00})G_{n0} &= -(\epsilon S_{01}^* - H_{01}^*)G_{n-1,0} - (\epsilon S_{01} - H_{01})G_{n+1,0}. \end{aligned}$$

From here, via the same series of algebraic manipulations as in the orthogonal case, we obtain the Green's function

$$G = \left[ (\epsilon S_{00} - H_{00}) + (\epsilon S_{01} - H_{01})T + (\epsilon S_{01}^* - H_{01}^*)\bar{T} \right]^{-1},$$

and from the latter we can identify the self-energies

$$\Sigma_L = -(\epsilon S_{01}^* - H_{01}^*)\bar{T}, \quad \Sigma_R = -(\epsilon S_{01} - H_{01})T.$$

The above procedure can be extended to the case of the transmission through an interface or a LCR junction. For the latter case, we obtain

$$\begin{aligned} \Sigma_L &= (\epsilon S_{LC} - H_{LC})^* \\ &\quad [\epsilon S_{00}^L - H_{00}^L + (\epsilon S_{01}^L - H_{01}^L)^* \bar{T}_L]^{-1} (\epsilon S_{LC} - H_{LC}), \\ \Sigma_R &= (\epsilon S_{CR} - H_{CR}) \\ &\quad [\epsilon S_{00}^R - H_{00}^R + (\epsilon S_{01}^R - H_{01}^R) T_R]^{-1} (\epsilon S_{CR} - H_{CR})^*, \end{aligned} \tag{56}$$

where  $H_{nm}^{\{L,R\}}$  are the matrix elements of the Hamiltonian between layer orbitals in the left and right leads, respectively.  $S_{nm}^{\{L,R\}}$  are the corresponding overlap matrices and  $T_{\{L,R\}}$  and  $\bar{T}_{\{L,R\}}$  are the appropriate transfer matrices. The latter are easily computed from the Hamiltonian and overlap matrix elements via the usual iterative procedure (see Section 3.2) Correspondingly,  $H_{LC}$ ,  $H_{CR}$ ,  $S_{LC}$  and  $S_{CR}$  are the coupling and overlap matrices for the conductor-leads assembly.

## 4 Applications: Conductivity from *ab initio* local orbital Hamiltonian

### 4.1 Methodology

The procedure described in Section 3 requires the knowledge of the Hamiltonian and overlap matrix elements between layer orbitals of the conductor, and the left and right leads. In *ab initio* density-functional calculations, such matrix elements can be computed using the  $O(N)$ -like algorithm described in Section 2.8. In this context the numerical orbitals — defined on a uniform grid in real-space — are centered on atoms and localized in spherical regions of radius  $R_L$  around the respective atoms. Since the orbitals are variationally optimized on the grid according to their environment until they accurately

describe the ground state of the system, it allows us to use only a small number of orbitals per atom, much smaller than in LCAO-based calculations. The size of the matrices that enter in the quantum conductance calculation and the computational cost of the whole procedure are thus minimized. In order to ensure fast convergence and accuracy – even for metallic systems – we use both occupied and unoccupied orbitals.

The matrices that enter the electronic transport calculation of a LCR system are computed in two steps. In the first calculation, we compute the ground state of the bare leads in a supercell with periodic boundary conditions. From this calculation we extract the Hamiltonian in the basis of the localized nonorthogonal orbitals and the overlap matrices required for the computation of the self-energies. We then perform a second ground state calculation in a supercell with periodic boundary conditions containing the conductor and one principal layer of the leads. In this calculation, the orbitals in the leads are kept the same as in the bare lead calculation, in order to extract the matrix elements describing the coupling between the conductor and the leads. This procedure fully accounts for the electronic structure of the conductor and the interaction between the conductor and the leads, provided that the lead region is large enough to avoid spurious interactions between periodic images of the contacts. In order to have interactions between the nearest-neighbor principal layers only, the width of the layers has to be sufficiently large compared to the localization regions. On the other hand, the localization regions have to be large enough to ensure an accurate solution of the density-functional equations. Moreover, in the Green’s function matching procedure one has to carefully align the Fermi levels of both systems in order to avoid spurious bias effects. Provided that in the conductor-lead calculation the lead region is large enough to recover bulk-like behavior far from the interfaces, we align the macroscopic average of the electrostatic potentials in the bare lead and in the conductor-lead geometry. This ensures a seamless conductor-lead geometry and prevents the spurious bias. An equivalent procedure is often used to extract band offsets in superlattice calculations (Baldereschi et al., 1988; Buongiorno Nardelli et al., 1997).

**Remark 4.1** *If a principal layer is composed of  $N$  orbitals, the calculation of the Green’s function requires a matrix inversion that scales as  $O(N^3)$ . However, for very large systems, the localization of the orbitals allows us to divide a principal layer into thinner layers and compute the quantities of interest in largely  $O(N)$  fashion (Anantram and Govindan, 1998).*

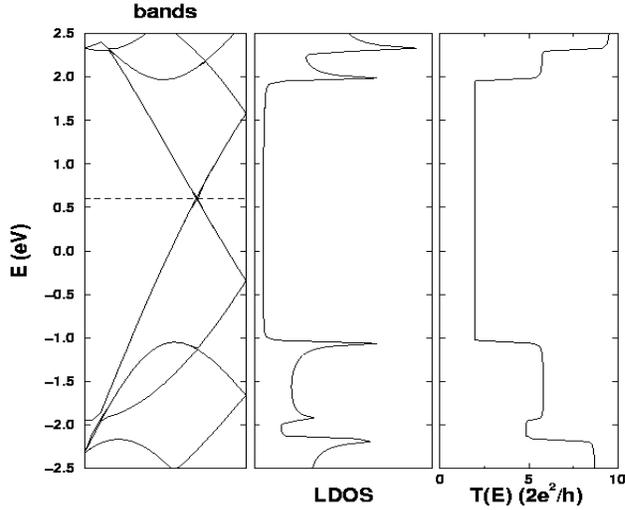


Fig. 8. Left: Electronic band structure of a metallic nanotube. Note the crossing of the bands at the Fermi energy. Middle: corresponding Density of States. Right: Quantum conductance spectrum. Note the metallic plateau of conductance equal to  $2G_0$ .

#### 4.2 Example: carbon nanotube on metallic contacts

To illustrate the above *ab initio* methodology we use the example of transport behavior of nanotube-metal contacts studied by Buongiorno Nardelli et al. (2001). The problem of contacts in metal-carbon nanotubes assemblies is a crucial issue for technological development, and determines much of the nanoscale device characteristics. A perfect metallic nanotube behaves like a ballistic conductor: every electron injected into the nanotube at one end should come out at the other end. The basic electronic properties of metallic nanotubes imply the existence of two propagating modes for electronic transmission, independent of the diameter (Bernholc et al., 2002). The electronic conductance is then expected to be twice the fundamental quantum of conductance,  $G_0 = 2e^2/h = 1/12.9 (k\Omega)^{-1}$ . At higher energies, the electrons are able to probe different sub-bands, which gives rise to an increase in  $G$  that is proportional to the number of additional bands available for transport. Hence,  $G$  for ideal nanotubes is expected to consist of a series of "down-and-up" steps as a function of the electron energy, in which the position of the steps correlate with the band edges. An illustration of this behavior is displayed in Fig.8. These considerations would suggest that nanotubes should behave as ideal device elements because of their electrical properties.

However, one of the fundamental problems that hinder a broader technological application of carbon nanotubes is the observation that most carbon nanotube

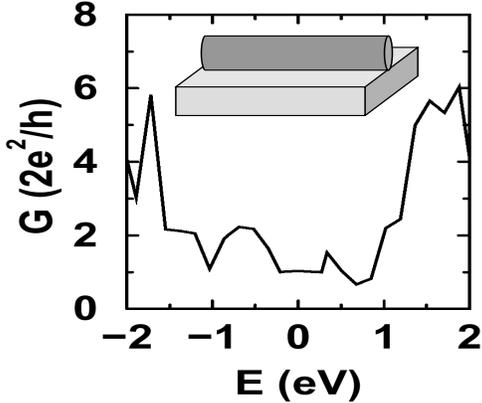


Fig. 9. The geometry and conductance spectrum of an infinite (5,5) nanotube deposited on Al(111). Adapted from Buongiorno Nardelli et al. (2001).

devices display contact resistances of the order of  $M\Omega$  (Tans et al., 1997, 1998; Martel et al., 1998; Bachtold et al., 1998), rather than  $k\Omega$ , as one would expect.

What is the physical origin behind the very high contact resistance for carbon nanotube systems? As a prototypical example, we consider the transport properties of a metallic (5,5) nanotube deposited on an Al (111) surface in an idealized side-contact geometry, as shown in the inset of Fig.9. In order to accurately account for the highly inhomogeneous environment of the nanowire-metal junction, and to account for the charge transfer occurring at the interface between these two dissimilar materials, it is important to use the accurate and self-consistent *ab initio* description we have previously discussed. The main characteristics of the electronic response of the system is a marked transfer of charge from the nanotube to the metal that allows the valence band edge of the nanotube to align with the Fermi level of the metal electrode (Xue and Datta, 1999).

This charge transfer, which has been already observed for other experimental systems (Tans et al., 1998; Wildoer et al., 1998; Martel et al., 1998) and calculations (Xue and Datta, 1999; Rubio et al., 1999; Kong et al., 1999), leads to enhanced conductivity along the tube axis and gives rise to a weak ionic bonding between the tube and the metal. The conductance spectrum for the coupled nanotube is displayed in Fig. 9. Although the metal contact increases the resistance by a factor of two as compared to ideal isolated tubes, the transmission through the system is still substantial. To further analyze the contact resistance, we have calculated the *eigenchannels* (Brandbyge et al., 1997). Among the conducting channels, *i.e.*, those with a significant nonzero transmission coefficient, we observe a clear distinction between channels that are localized in the metal and those on the nanotube itself. This result reflects the clear separation of the individual electronic wave functions of each of the components of the system. In particular, the eigenchannel corresponding to the

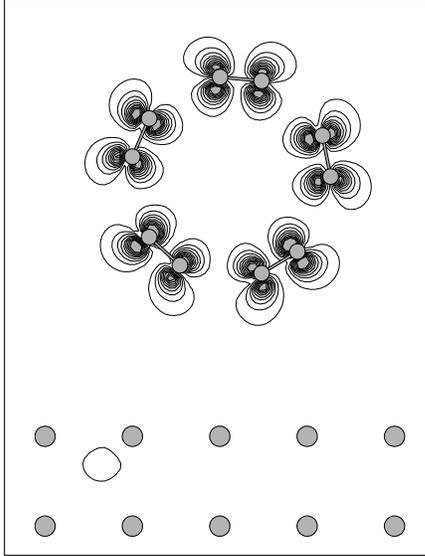


Fig. 10. Cross section of the square modulus of the electronic wave function corresponding to the only open eigenchannel at the Fermi level that has a sizable component on the nanotube for the system represented in Fig.9. The other wave functions at the Fermi level are mostly localized on the metal. Adapted from Buongiorno Nardelli et al. (2001).

plateau of conductance around the Fermi energy corresponds to an individual wave function, reproduced in Fig.10, almost fully localized on the nanotube. This implies that there is very little hybridization and intermixing between the nanotube and the metal in the channel responsible for conduction at the Fermi level. Thus, the conduction electron transfer between the tube and the metal in the idealized side-contact geometry considered here is very inefficient, which can explain the high contact resistance observed in nanotube-metal contacts.

This initial investigation has been extended to a geometry that more closely resembles an experimental two-terminal device, with two semi-infinite contacts connected by a nanotube bridge, 1.5 nm long. In this geometry, the system recovers the ideal conductance of an isolated tube with two conductance channels at the Fermi energy, as shown in Fig. 11. This behavior is induced by the alignment of the valence band edge of the nanotube with the Fermi energy of the metal contacts, triggered by the charge transfer in the lead regions. In this particular geometry, these conditions restore the two original eigenchannels of the nanotube and thus conserve the number of conducting channels throughout the system. It is important to note that the weak nanotube-metal interaction, responsible for the pathologically high resistance of the nanotube-metal assembly, is not strengthened.

**Remark 4.2** *In this calculation, the conductor  $C$  is made of a (5,5) carbon nanotube composed of 120 atoms, while a principal layer of the leads is made of the same 120 atoms carbon nanotube and of an aluminium surface of 100 atoms. The width of a principal layer is  $14.7 \text{ \AA}$  and a localization radius of*

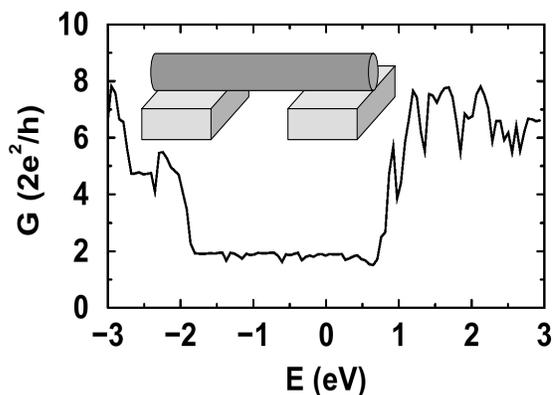


Fig. 11. The conductance spectrum of an ideal two-terminal device, as shown in the inset. The Fermi level is taken as a reference. From Buongiorno Nardelli et al. (2001).

*5.3 Å was used for the orbitals. The LDA exchange and correlation functional was used with the pseudopotentials by Hamann (1989).*

**Remark 4.3** *The occurrence of a single channel in the first case is due to the idealized geometry of an infinite nanotube on an infinite metallic surface. The conservation of the total number of channels is ensured by the channels localized on the metal.*

These examples clearly demonstrate that the weak nanotube-metal coupling is mostly responsible for the weak electron transport in the combined system, and that wave vector conservation is not a significant factor (Tersoff, 1999; Delaney et al., 1999). The weak distributed coupling also explains why the measured contact resistance is inversely proportional to the contact length (Tans et al., 1997; Frank et al., 1998; Anantram et al., 2000). Although the nanotube behaves as an ideal ballistic conductor, the bonding characteristics of the nanotube-metal system prevent an efficient electron transfer mechanism from the nanotube to the Al contact. Indeed, inducing defects in the contact region, *e.g.*, by localized electron bombardment (Bachtold et al., 1998), drastically increases the bonding strength of the nanotube-metal assembly and greatly improve the performance of the device. Alternatively, mechanically pushing the nanotube closer to the Al surface by a small amount ( $\approx 1$  Å, with an energy cost of  $\approx 10$  meV/atom) more than doubles the transmission efficiency between the metal and the nanotube. The mechanical deformation induces a small inward relaxation of the Al surface in the contact region, facilitating stronger hybridization between the nanotube and the metal contact in the conducting channels and therefore leading to a higher transmission rate.

## Acknowledgements

The authors are very grateful to Prof. J. Bernholc for pointing them towards the problems of linear scaling and electronic transport, and for numerous stimulating discussions about these problems. A portion of the work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

## References

- Anantram, M., Govindan, T., 1998. Conductance of carbon nanotubes with disorder: a numerical study. *Phys. Rev. B* 58 (8), 4882–4887.
- Anantram, M. P., Datta, S., Xue, Y. Q., 2000. Coupling of carbon nanotubes to metallic contacts. *Phys. Rev. B* 61 (20), 14219–14224.
- Ancilotto, F., Blandin, P., Toigo, F., 1999. Real-space full-multigrid study of the fragmentation of  $\text{Li}_{11}^+$  clusters. *Phys. Rev. B* 59 (12), 7868–7875.
- Ashcroft, N., Mermin, N., 1976. *Solid State Physics*. Holt-Saunders.
- Bachtold, A., Henny, M., Tarrier, C., Strunk, C., Schonenberger, C., Salvetat, J. P., Bonard, J. M., Forro, L., 1998. Contacting carbon nanotubes selectively with low-ohmic contacts for four-probe electric measurements. *Appl. Phys. Lett.* 73 (2), 274–276.
- Baldereschi, A., 1973. Mean-value point in the Brillouin zone. *Phys. Rev. B* 7, 5212.
- Baldereschi, A., Baroni, S., Resta, R., 1988. Band offsets in lattice-matched heterojunctions: A model and first-principles calculations for GaAs/AlAs. *Phys. Rev. Lett.* 61, 734.
- Beck, T., 2000. Real-space mesh techniques in density-functional theory. *Rev. Mod. Phys.* 72 (4), 1041–1080.
- Beenakker, C., Van Houten, H., 1991. Quantum transport in semiconductor nanostructures. *Solid State Phys.* 44, 1–228.
- Bernholc, J., Brenner, D., Buongiorno Nardelli, M., Meunier, V., Roland, C., 2002. Mechanical and electrical properties of nanotubes. *Annual Reviews in Materials Science* .
- Blackford, L. S., Choi, J., Cleary, A., D’Azevedo, E., Demmel, J., Dhillon, I., Dongarra, J., Hammarling, S., Henry, G., Petitet, A., Stanley, K., Walker, D., Whaley, R. C., 1997. *SCALAPACK User’s Guide*. SIAM, Philadelphia.
- Brandbyge, M., Sorensen, M., Jacobsen, K., 1997. Conductance eigenchannels in nanocontacts. *Phys. Rev. B* 56, 14956.
- Brandt, A., 1977. Multilevel adaptative solutions to boundary-value problems. *Math. Comp.* 31 (138), 333–390.
- Briggs, E. L., Sullivan, D. J., Bernholc, J., 1995. Large scale electronic struc-

- ture calculations with multigrid acceleration. *Phys. Rev. B* 52 (8), R5471–5474.
- Briggs, E. L., Sullivan, D. J., Bernholc, J., 1996. Real-space multigrid-based approach to large-scale electronic structure calculations. *Phys. Rev. B* 54 (20), 14362–14375.
- Bungaro, C., Rapcewicz, K., Bernholc, J., 1999. Surface sensitivity of impurity incorporation: Mg at GaN(0001) surfaces. *Phys. Rev. B* 59 (15), 9771–9774.
- Buongiorno Nardelli, M., 1999. Electronic transport in extended systems: Application to carbon nanotubes. *Phys. Rev. B* 60 (11), 7828–7833.
- Buongiorno Nardelli, M., Bernholc, J., 1999. Mechanical deformations and coherent transport in carbon nanotubes. *Phys. Rev. B* 60 (24), R16338–R16341.
- Buongiorno Nardelli, M., Fattebert, J.-L., Bernholc, J., 2001. An  $O(N)$  real-space method for ab initio quantum transport calculations: application to carbon nanotube-metal contacts. *Phys. Rev. B* 64, 245423.
- Buongiorno Nardelli, M., Rapcewicz, K., Bernholc, J., 1997. Strain effects on the interface properties of nitride semiconductors. *Phys. Rev. B* 55, R7323.
- Buongiorno Nardelli, M., Yakobson, B., Bernholc, J., 1998. Mechanism of strain release in carbon nanotubes. *Phys. Rev. B* 57 (8), R4277–R4280.
- Chelikowsky, J. R., Trouiller, N., Saad, Y., 1994. Finite-difference-pseudopotential method: Electronic structure calculations without a basis. *Phys. Rev. Lett.* 72 (8), 1240–1243.
- Chetty, N., Weinert, M., Rahman, T. S., Davenport, J. W., 1995. Vacancies and impurities in aluminium and magnesium. *Phys. Rev. B* 52 (9), 6313–6326.
- Choi, H., Ihm, J., 1999. Ab initio pseudopotential method for the calculation of conductance in quantum wires. *Phys. Rev. B* 59, 2267.
- Choi, H., Ihm, J., Louie, S., Cohen, M., 2000. Defects, quasibound states, and quantum conductance in metallic carbon nanotubes. *Phys. Rev. Lett.* 84, 2917.
- Collatz, L., 1966. *The Numerical Treatment of Differential Equations*. Springer, Berlin.
- Costiner, S., Taasan, S., 1995. Simultaneous multigrid techniques for nonlinear eigenvalue problems: Solutions of the nonlinear Schrödinger-Poisson eigenvalue problem in two and three dimensions. *Phys. Rev. E* 52 (1), 1181–1192.
- Datta, S., 1995. *Electronic transport in mesoscopic systems*. Cambridge University Press.
- Delaney, P., Di Ventura, M., Pantelides, S. T., 1999. Quantized conductance of multiwalled carbon nanotubes. *Appl. Phys. Lett.* 75 (24), 3787–3789.
- Descloux, J., Fattebert, J.-L., Gygi, F., 1998. RQI (Rayleigh Quotient Iteration), an old recipe for solving modern large scale eigenvalue problems. *Comput. Phys.* 12 (1), 22–27.
- Di Ventura, M., Pantelides, S., Lang, N., 2000. First-principles calculation of transport properties of a molecular device. *Phys. Rev. Lett.* 84, 979.
- Edelman, A., Arias, T., Smith, S., 1998. The geometry of algorithms with

- orthogonality constraints. *SIAM J. Matrix Anal. Appl.* 20 (2), 303–353.
- Fattebert, J.-L., 1996. An inverse iteration method using multigrid for quantum chemistry. *BIT* 36 (3), 509–522.
- Fattebert, J.-L., 1998. A block Rayleigh Quotient Iteration with local quadratic convergence. *ETNA* 7, 56–74.
- Fattebert, J.-L., 1999. Finite difference schemes and block Rayleigh Quotient Iteration for electronic structure calculations on composite grids. *J. Comput. Phys.* 149, 75–94.
- Fattebert, J.-L., Bernholc, J., 2000. Towards grid-based  $O(N)$  density-functional theory methods: optimized non-orthogonal orbitals and multigrid acceleration. *Phys. Rev. B* 62 (3), 1713–1722.
- Fattebert, J.-L., Gygi, F., 2002. Density functional theory for efficient ab initio molecular dynamics simulations in solution. *J. Comput. Chem.* 23, 662–666.
- Fernando, G. W., Quian, G.-X., Weinert, M., Davenport, J. W., 1989. First-principles molecular dynamics for metals. *Phys. Rev. B* 40 (11), 7985–7988.
- Ferry, F., Goodnick, S., 1997. *Transport in nanostructures*. University Press, Cambridge.
- Feynman, R., 1939. Forces in molecules. *Phys. Rev.* 56, 340–343.
- Fisher, D., Lee, P., 1981. Relation between conductivity and transmission matrix. *Phys. Rev. B* 23, 6851.
- Frank, S., Poncharal, P., Wang, Z. L., Heer, W. A. D., 1998. Carbon nanotube quantum resistors. *Science* 280 (5370), 1744–1746.
- Galli, G., Parrinello, M., 1992. Large scale electronic structure calculations. *Phys. Rev. Lett.* 69 (24), 3547–3550.
- Galli, G., Pasquarello, A., 1993. First-principles molecular dynamics. In: Allen, M., Tildesley, D. (Eds.), *Computer Simulation in Chemical Physics*. Kluwer Academic Publishers, pp. 261–313.
- Garcia-Moliner, F., Velasco, V., 1991. Matching methods for single and multiple interfaces: discrete and continuous media. *Phys. Rep.* 200 (3), 83–125.
- Garcia-Moliner, F., Velasco, V., 1992. *Theory of Single and Multiple Interfaces*. World Scientific, Singapore.
- Gygi, F., Galli, G., 1995. Real-space adaptive-coordinate electronic-structure calculations. *Phys. Rev. B* 52 (4), 2229–2232.
- Hamann, D. R., 1989. Generalized norm-conserving pseudopotentials. *Phys. Rev. B* 40 (5), 2980–2987.
- Heermann, D., 1990. *Computer simulation methods*, 2nd Edition. Springer-Verlag, Berlin Heidelberg.
- Heiskanen, M., Torsti, T., Puska, M., Nieminen, R., 2001. Multigrid method for electronic structure calculations. *Phys. Rev. B* 63, 245106.
- Hernandez, E., Gillan, M. J., 1995. Self-consistent first-principles technique with linear scaling. *Phys. Rev. B* 51 (15), 10157–10160.
- Hirose, K., Tsukada, M., 1995. First-principles calculation of the electronic structure for a bielectrode junction system under strong field and current. *Phys. Rev. B* 51, 5278.
- Hoshi, T., Fujiwara, T., 1997. Fully-selfconsistent electronic-structure calcu-

- lation using nonorthogonal localized orbitals within a finite-difference real-space scheme and ultrasoft pseudopotential. *J. Phys. Soc. Jpn.* 66 (12), 3710–3713.
- H. Takahashi, Hori, T., Hashimoto, H., Nitta, T., 2001. A hybrid QM/MM method employing real space grids for QM water in the TIP4P water solvent. *J. Comp. Chem.* 22 (12), 1252–1261.
- Ismael-Beigi, S., Arias, T. A., 1999. Locality of the density matrix in metals, semiconductors, and insulators. *Phys. Rev. Lett.* 82 (10), 2127–2130.
- Jin, Y.-G., Chang, K., 2001. Mechanism for the enhanced diffusion of charged oxygen ions in SiO<sub>2</sub>. *Phys. Rev. Lett.* 86 (9), 1793–1796.
- Jin, Y.-G., Jeong, J.-W., Chang, K., 1999. Real-space electronic structure calculations of charged clusters and defects in semiconductors using a multigrid method. *Physica B* 273–274, 1003–1006.
- Jing, X., Troullier, N., Dean, D., Binggeli, N., Chelikowsky, J., 1994. Ab initio molecular dynamics simulations of Si clusters using higher-order finite difference-pseudopotential method. *Phys. Rev. B* 50 (16), 12234–12237.
- Kleinman, L., Bylander, D., 1982. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.* 48 (20), 1425–1428.
- Kobayashi, N., Brandbyge, M., Tsukada, M., 2000. Conduction channels at finite bias in single-atom gold contact. *Phys. Rev. B* 62, 8430.
- Kohn, S., Weare, J., Ong, M., Baden, S., 1997. Software abstractions and computational issues in parallel structured adaptive mesh methods for electronic structure calculations. In: Proc. of Workshop on structured adaptive mesh refinement grid methods. Minneapolis.
- Kohn, W., Sham, L. J., 1965. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A* 140, 1133–1138.
- Kong, K., Han, S., Ihm, J., 1999. Development of an energy barrier at the metal-chain-metallic- carbon-nanotube nanocontact. *Phys. Rev. B* 60 (8), 6074–6079.
- Landauer, R., 1970. Electrical resistance of disordered one-dimensional lattices. *Philos. Mag.* 21, 863–867.
- Landman, U., Barnett, R., Scherbakov, A. G., Avouris, P., 2000. Metal-semiconductor nanocontacts: silicon nanowires. *Phys. Rev. Lett.* 85, 1958.
- Lang, N., 1995. Resistance of atomic wires. *Phys. Rev. B* 52, 5335.
- Lee, D., Joannopoulos, J., 1981a. Simple scheme for surface-band calculations. I. *Phys. Rev. B* 23 (10), 4988.
- Lee, D., Joannopoulos, J., 1981b. Simple scheme for surface-band calculations. II. the Green's function. *Phys. Rev. B* 23 (10), 4997.
- Lopez-Sancho, M., Lopez-Sancho, J., Rubio, J., 1984. Quick iterative scheme for the calculation of transfer matrices: application to Mo(100). *J. Phys. F: Metal Phys.* 14, 1205–1215.
- Lopez-Sancho, M., Lopez-Sancho, J., Rubio, J., 1985. Highly convergent schemes for the calculation of bulk and surface Green functions. *J. Phys. F: Metal Phys.* 15, 851–858.
- Mandel, J., McCormick, S., 1989. A multilevel variational method for  $Au =$

- $\lambda Bu$  on composite grids. *J. Comput. Phys.* 80 (2), 442–452.
- Martel, R., Schmidt, T., Shea, H. R., Hertel, T., Avouris, P., 1998. Single- and multi-wall carbon nanotube field-effect transistors. *Appl. Phys. Lett.* 73 (17), 2447–2449.
- Meir, Y., Wingreen, N., 1992. Landauer formula for the current through an interacting electron region. *Phys. Rev. Lett.* 68 (16), 2512–2515.
- Millam, J. M., Scuseria, G. E., 1997. Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations. *J. Chem. Phys.* 506 (13), 5569–5577.
- Modine, N. A., Zumbach, G., Kaxiras, E., 1996. Adaptive-coordinate real-space electronic-structure calculations on parallel computers. *Solid State Comm.* 99 (2), 57–61.
- Murakami, H., Sonnad, V., Clementi, E., 1992. A three-dimensional finite element approach towards molecular SCF computations. *Int. J. Quant. Chem.* 42, 785–817.
- Parlett, B. N., 1998. *The Symmetric Eigenvalue Problem*. SIAM, Philadelphia.
- Parrinello, M., 1997. From Silicon to RNA: the coming age of *ab initio* molecular dynamics. *Sol. State Comm.* 102 (2–3), 107–120.
- Pask, J. E., Klein, B. M., Fong, C. Y., Sterne, P. A., 1999. Real-space polynomial basis for solid-state electronic structure calculations: A finite-element approach. *Phys. Rev. B* 59 (19), 12352–12358.
- Payne, M. C., Teter, M. P., Allan, D. C., Arias, T., Joannopoulos, J., 1992. Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* 64 (4), 1045–1097.
- Perdew, J., Burke, K., Ernzerhof, M., 1996. Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77 (18), 3865–3868.
- Pfrommer, B., Demmel, J., Simon, H., 1999. Unconstrained energy functionals for electronic structure calculations. *J. Comput. Phys.* 150, 287–298.
- Ramamoorthy, M., Briggs, E., Bernholc, J., 1998. Chemical trends in impurity incorporation into Si(100). *Phys. Rev. Lett.* 81 (8), 1642–1645.
- Rubio, A., Sanchez-Portal, D., Artacho, E., Ordejon, P., Soler, J. M., 1999. Electronic states in a finite carbon nanotube: a one-dimensional quantum box. *Phys. Rev. Lett.* 82 (17), 3520–3523.
- Saad, Y., Stathopoulos, A., Chelikowsky, J., Wu, K., Ogut, S., 1996. Solution of large eigenvalue problems in electronic structure calculations. *BIT* 36 (3), 563–578.
- Schmidt, W., F. B., Bernholc, J., 2001. Terrace and step contributions to the optical anisotropy of Si(001) surfaces. *Phys. Rev. B* 63 (4), 045322.
- Sleijpen, G. L. G., Van Der Vorst, H. A., 1996. A generalized Jacobi-Davidson iteration method for linear eigenvalue problem. *SIAM Matrix Anal. and Appl.* 17 (2), 401–425.
- Tans, S. J., Devoret, M. H., Dai, H. J., Thess, A., Smalley, R. E., Geerligs, L. J., Dekker, C., 1997. Individual single-wall carbon nanotubes as quantum wires. *Nature* 386 (6624), 474–477.

- Tans, S. J., Verschueren, A. R. M., Dekker, C., 1998. Room-temperature transistor based on a single carbon nanotube. *Nature* 393 (6680), 49–52.
- Taylor, J., Guo, H., Wang, J., 2000. Ab initio modeling of open systems: Charge transfer, electron conduction, and molecular switching of a C60 device. *Phys. Rev. B* 63, 121104.
- Tersoff, J., 1999. Response to comment on contact resistance of carbon nanotubes "[*appl. phys. lett.* 75, 4028 (1999)]". *Appl. Phys. Lett.* 75 (25), 4030–4030.
- Teter, M. P., Payne, M. C., Allan, D. C., 1989. Solution of Schrödinger's equation for large systems. *Phys. Rev. B* 40 (18), 12255–12263.
- Tsuchida, E., Tsukada, M., 1995. Electronic-structure calculations based on the finite-element method. *Phys. Rev. B* 52 (8), 5573–5578.
- Vasiliev, I., Ogut, S., Chelikowsky, J., 1999. Ab initio excitation spectra and collective electronic response in atoms and clusters. *Phys. Rev. Lett.* 82 (9), 1919–1922.
- Wang, J., Beck, T., 2000. Efficient real-space solution of the Kohn-Sham equations with multiscale techniques. *J. Chem. Phys.* 112 (21), 9223–9228.
- White, S. R., Wilkins, J. W., Teter, M. P., 1989. Finite-element method for electronic structure calculations. *Phys. Rev. B* 39 (9), 5819–5833.
- Wildoer, J. W. G., Venema, L. C., Rinzler, A. G., Smalley, R. E., Dekker, C., 1998. Electronic structure of atomically resolved carbon nanotubes. *Nature* 391 (6662), 59–62.
- Xue, Y. Q., Datta, S., 1999. Fermi-level alignment at metal-carbon nanotube interfaces: application to scanning tunneling spectroscopy. *Phys. Rev. Lett.* 83 (23), 4844–4847.
- Yaliraki, S., Roitberg, A. E., Gonzalez, C., Mujica, V., Ratner, M., 1999. The injecting energy at molecule/metal interfaces: Implications for conductance of molecular junctions from an ab initio molecular description. *J. Chem. Phys.* 111, 6997.
- Yoon, Y.-G., Mazzoni, M., Choi, H., Ihm, J., Louie, S., 2001. Structural deformation and intertube conductance of crossed carbon nanotube junctions. *Phys. Rev. Lett.* 86, 688.