Recent developments and applications of the real-space multigrid method

2008 J. Phys.: Condens. Matter 20 294205

(http://iopscience.iop.org/0953-8984/20/29/294205)

J. Phys.: Condens. Matter **20** (2008) 294205 (8pp) doi:10.1088/0953-8984/20/29/294205

Recent developments and applications of the real-space multigrid method

J Bernholc[1,2], Miroslav Hodak[1] and Wenchang Lu[1,2]

[1] Center for High Performance Simulation and Department of Physics, North Carolina State University, Raleigh, NC 27695-7518, USA
[2] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6367, USA

Received 1 February 2008, in final form 28 May 2008
Published 24 June 2008
Online at stacks.iop.org/JPhysCM/20/294205

Abstract
The salient features of the real-space multigrid method and its recent applications are described. This method is suitable for very large scale, massively parallel calculations of atomic and electronic structure, as well as quantum molecular dynamics. Its nearly O(N) implementation provides a compact, variationally optimized basis that is also very useful for fully O(N) calculations of quantum transport. Recently, we also developed a hybrid method for simulating biomolecules in solution, in which most of the solvent is inexpensively treated using an approximate density-functional method, while the biomolecule and its first solvation shells are described at the full Kohn–Sham level. Our calculations show excellent parallel efficiency and scaling on massively parallel supercomputers.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

It is well known that *ab initio* methods have evolved to a point where properties of materials can often be successfully predicted based solely on their atomic structure and without any experimental input. Conversely, if the atomic structure is only approximately known from, e.g., atomic force microscope imaging of the surface, the same methods can be used to 'optimize' the structure and to predict the precise atomic positions. Furthermore, theory can provide a 'theoretical microscope', by identifying the origins of the various properties of a given material and by uncovering principles that can be used to systematically enhance the desired characteristics, or to suppress the unwanted ones.

To apply *ab initio* methodology to systems with more than a few tens of atoms, efficient implementation is essential. Most *ab initio* implementations use plane-waves as a basis for representing wavefunctions. While this approach has been highly successful, achieving efficiency on massively parallel supercomputers is difficult. This is mostly due to the frequent use of Fourier transforms, which are global operations requiring large amounts of interprocessor communication. For modern massively parallel supercomputers with very fast microprocessors, communication between distant processors may become a bottleneck, limiting the parallelization efficiency.

Our group has developed a multigrid-based method for performing electronic structure calculations in real space [1]. Unlike the plane-wave approach, real-space methods are inherently local, and therefore can be efficiently used on parallel supercomputers. The multigrid iteration technique is employed to improve the convergence rate of ground-state wavefunctions. Our implementation enables the use of either norm-conserving [2] or ultrasoft [3] pseudopotentials and our tests show that systems with several thousands of atoms can be treated efficiently.

The real-space method enables effective implementation of O(N) techniques, which promise to reduce the asymptotic O(N^3) scaling of traditional electronic structure methods with respect to the number of electrons or atoms [4]. The O(N) method relies on the construction of a minimal set of localized orbitals, variationally optimized for the system at hand. The resulting compact basis is very useful for calculations of quantum transport properties of nanoscale systems, where expansion of various operators in a compact local basis is needed [5–7].

The real-space approach can also be used for simulations of biological systems. However, these simulations provide a unique challenge for *ab initio* methods for several reasons. First, the biological molecules are very large, often containing many thousands of atoms. Second, solvent plays an

important role in biosystems and thus it has to be included in the calculations. Because including full solvation in *ab initio* calculation is prohibitively expensive, simplified solvent models are used. Often, the solvent is modeled implicitly, as a polarizable continuum dielectric [8–10]. This approach was successfully used [11–14] in many instances, but evidence suggests that for a complete understanding of the behavior of many biological systems an explicit treatment of solvent is essential. For example, molecular mechanics simulations of proteins have shown that an implicit representation of solvent leads to incorrect predictions of native structures and folding pathways, while explicit treatment yields results in agreement with experimental observations [15, 16]. Also, many biomolecules have water molecules near their active sites, which are essential to their functions. To properly describe these water molecules, hydrogen bonds with other water molecules have to be included and this can only be done with an explicit solvent model.

To address this, we have developed a method that includes explicit full solvation at yet fully quantum level in *ab initio* simulations at low computational cost. In our method, solvent molecules are treated with a frozen-density orbital-free (FDOF) DFT method, while the rest of the system, including the biomolecule and its first solvation shells, is described at the full *ab initio* level. This approach enables the inclusion of thousands of explicit solvent molecules at the cost of about 1% of the time required for the *ab initio* calculation on the rest of the system.

2. Grid-based discretizations of the Kohn–Sham equations

The standard Kohn–Sham energy functional is

$$E_{\rm tot}[\{\phi_i\},\{\mathbf{R}_I\}] = \sum_i \langle\phi_i| -\tfrac{1}{2}\nabla^2 + V_{\rm NL}|\phi_i\rangle + \frac{1}{2}\int\int {\rm d}\mathbf{r}\,{\rm d}\mathbf{r}'\,\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\rm xc}[n] + \int {\rm d}\mathbf{r}\,V_{\rm loc}^{\rm ion}(\mathbf{r})n(\mathbf{r}) + E_{\rm ion}(\{\mathbf{R}_I\}), \quad (1)$$

where ϕ_i are the wavefunctions, $V_{\rm NL}$ is the non-local pseudopotential, $E_{\rm xc}$ provides the exchange–correlation energy, $V_{\rm loc}^{\rm ion}$ represents the local part of pseudopotential and $E_{\rm ion}$ accounts for the energy of interaction between ions in the system. In this expression, it is assumed that the Kleinman–Bylander [2] norm-conserving pseudopotentials are used to describe the ions. The Vanderbilt ultrasoft [3] pseudopotentials require modifications to the computational procedure, these will be discussed in section 3. The minimization of this functional requires the solution of the Kohn–Sham equations

$$H_{\rm KS}[\psi_n] = -\tfrac{1}{2}\nabla^2\psi_n + V_{\rm eff}\psi_n = \epsilon_n\psi_n. \quad (2)$$

The differential operator in the Kohn–Sham equations is approximated using a *generalized* eigenvalue form:

$$\mathbf{H}_{\rm mehr}[\psi_n] = \tfrac{1}{2}\mathbf{A}_{\rm mehr}[\psi_n] + \mathbf{B}_{\rm mehr}[V_{\rm eff}\psi_n] = \epsilon_n\mathbf{B}_{\rm mehr}[\psi_n], \quad (3)$$

where $\mathbf{A}_{\rm mehr}$ and $\mathbf{B}_{\rm mehr}$ are the components of the *Mehrstellen* discretization [17], which is based on Hermite's generalization of Taylor's theorem. It uses a weighted sum of the wavefunction and potential values to improve the accuracy of the discretization of the *entire* differential equation, not just the kinetic energy operator. In contrast to the central finite-differencing method, this discretization uses near-neighbor points along all directions, rather than just those along the Cartesian directions, to improve the accuracy of the discretization while reducing its range.

To efficiently solve equation (3), we have used multigrid iteration techniques that accelerate convergence by employing a sequence of grids of varying resolutions. The solution is obtained on a grid fine enough to accurately represent the pseudopotentials and the electronic wavefunctions. If the solution error is expanded in a Fourier series, it may be shown that iterations on any given grid level will quickly reduce the components of the error with wavelengths comparable to the grid spacing, but are ineffective in reducing the components with wavelengths that are large relative to the grid spacing [18, 19]. The solution is to treat the lower frequency components on a sequence of auxiliary grids with progressively larger grid spacings, where the remaining errors appear as high frequency components. This procedure provides excellent preconditioning for all length scales present in a system and leads to very rapid convergence rates. The operation count to converge one wavefunction with a fixed potential is $O(N_{\rm grid})$, compared to $O(N_{\rm grid}\log N_{\rm grid})$ for FFT-based approaches [20]. A subspace diagonalization is performed every few SCF steps (every 10–20 SCF steps is generally adequate) in order to unmix eigenstates that may be close in energy.

The computational effort in traditional electronic structure calculations must ultimately scale as $O(N^3)$, where N is the number of atoms. This is because the wavefunction of each electron can in general extend over the whole material, and therefore computing one wavefunction will take at least $O(N)$ operations. Since the number of electrons grows linearly with the number of atoms, the computational effort must grow at least as $O(N^2)$. Furthermore, the individual wavefunctions must be orthogonal to each other and the orthogonalization or diagonalization effort will ultimately dominate, since they scale as $O(N^3)$.

A number of ingenious methods have been proposed for evaluating the total energy in $O(N)$ operations. These methods usually make a *localization* approximation, which involves either the use of a localized, Wannier-like basis [21–27] or a neglect of off-diagonal elements of the density matrix $\rho(\mathbf{r},\mathbf{r}')$ for $|\mathbf{r}-\mathbf{r}'|$ greater than an appropriate cutoff radius [28–33]. Real-space methods are inherently local, and therefore suitable for imposing localization constraints on the basis functions that span the subspace of the both occupied and unoccupied orbitals.

Consider a trial basis of non-orthogonal normalized functions, $\Phi = \{\phi_1,\ldots,\phi_N\}$ [34], that will be refined by iterative updates, and at convergence will accurately describe the true Kohn–Sham ground state of the system. The $O(N)$

scaling in electronic structure part of the problem is achieved by minimizing the total energy functional

$$E[\{\phi_i\}_{i=1}^N] = 2\,\mathrm{Tr}(H^{(\phi)}\bar{\rho}^{(\phi)}) - \frac{1}{2}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' - \int \mu_{\mathrm{xc}}(\rho)\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} + E_{\mathrm{xc}}[\rho] \tag{4}$$

for a set of non-orthogonal orbitals $\{\phi_i\}_{i=1}^N$. Here $\bar{\rho}^{\Phi}$ is the density matrix in the basis $\{\phi\}$ and the $N \times N$ matrix $H^{(\Phi)}$ is defined by $H_{ij}^{(\Phi)} = \langle\phi_i|H|\phi_j\rangle$. The main advantage of this formulation is that the shapes of the localized orbitals are variationally optimized for a given atomic geometry. Full DFT accuracy can thus be achieved with very few orbitals per atom. The 'price' is significant delocalization of the optimized orbitals (up to 8–10 au), but the small size of the matrix equations is still a major advantage. In fact, it has been shown [35] that the dual, non-orthogonal Wannier-like basis is more localized than either an orthogonal Wannier basis or a density-support basis, thus resulting in the smallest possible number of non-zero overlaps.

The preconditioned steepest descent (PSD) direction in the non-orthogonal basis can be evaluated by

$$\delta\Phi = K(\Phi\Theta - H\Phi) \tag{5}$$

where $\Theta = S^{-1}H^{(\Phi)}$ and S, H^{Φ} are the overlap and Hamiltonian matrices. K is the multigrid preconditioning operator, which substantially accelerates convergence.

In actual calculations, the basis functions Φ are corrected at each iteration using the PSD directions (5) with Pulay mixing. A new electronic density $\rho(\mathbf{r})$ is then generated according to

$$\rho(\mathbf{r}) = \sum_{j,k} \rho_{jk}\phi_j(\mathbf{r})\phi_k(\mathbf{r}), \tag{6}$$

as well as the new Hartree and exchange–correlation potentials. In true O(N) calculations, the total energy is minimized iteratively until self-consistency is achieved. However, we often find that convergence can be accelerated further diagonalizing the Hamiltonian matrix. Using a highly parallel SCALAPACK solver, the diagonalization cost represents only a modest fraction of the total cost per iteration, while substantially reducing the number of iterations. Diagonalization is also required when the system contains partially occupied orbitals, or when unoccupied orbitals are needed. The latter is the case for our quantum transport calculations, which use this basis for the expansion of Green's functions and transfer matrices (see section 6).

3. Ultrasoft pseudopotentials

In Vanderbilt's ultrasoft pseudopotential scheme, the electron density consists of two parts: a soft contribution given by the squared moduli of the wavefunctions ϕ_i, and a hard contribution given by angular momentum projector functions β_n^{I} and augmentation functions Q_{nm}^{I}, which are localized at ionic cores:

$$n(\mathbf{r}) = \sum_i \left[|\phi_i(\mathbf{r})|^2 + \sum_{nm,I} Q_{nm}^{\mathrm{I}}(\mathbf{r}) \times \langle\phi_i(\mathbf{r})|\beta_n^{\mathrm{I}}(\mathbf{r})\rangle\langle\beta_m^{\mathrm{I}}(\mathbf{r})|\phi_i(\mathbf{r})\rangle \right]. \tag{7}$$

Relaxation of norm-conservation in UPPs leads to the generalized orthonormality condition

$$\langle\phi_i|S(\{\mathbf{R}_I\})|\phi_j\rangle = \delta_{ij}, \tag{8}$$

where the overlap operator S, which depends on ionic positions $\{\mathbf{R}_I\}$, is defined by

$$S = 1 + \sum_{nm,I} q_{nm}|\beta_n^{\mathrm{I}}\rangle\langle\beta_m^{\mathrm{I}}|, \tag{9}$$

with

$$q_{nm} = \int Q_{nm}(\mathbf{r})\,\mathrm{d}\mathbf{r}. \tag{10}$$

When norm-conserving pseudopotentials are used, the wavefunctions, the charge density, and the various components of the total potential are all represented on a single uniform grid. The UPP charge density is hard around the ionic cores, while the wavefunctions are much smoother. To take advantage of these properties, our ultrasoft real-space implementation [36] has to use two global grids: a fine one for the charge density and the $Q_{nm}(\mathbf{r})$ functions, and a coarse one for representing wavefunctions. The coarse wavefunction grid also accelerates the solution of the Kohn–Sham equations, which is usually the most time-consuming part of electronic structure calculations. Furthermore, it decreases computer memory requirements for storing wavefunctions, which is often the limiting factor for systems with many electrons. The fine and coarse grids are implemented in such a way that the coarse grid is a subset of the fine grid. Our tests show that using a fine grid with twice as many points in each direction as the coarse grid is usually adequate.

In the calculations, one needs to transfer quantities known on the fine grid to the coarse grid, and vice versa. Transferring from the fine to the coarse grid is straightforward; since the coarse grid is a subset of the fine one, it is sufficient to use injection, i.e., use its already known values on the coarse-grid points. However, the situation is more complicated when transferring from the coarse grid to the fine one. This is needed when calculating the electron density according to equation (7), since the wavefunctions $\phi_i(\mathbf{r})$ are only known on the coarse grid, while the electron density is represented on the fine one. An interpolation scheme is therefore required, and our tests show that it must be quite accurate. Since our RMG implementation uses regular rectangular grids, an interpolation in three dimensions can be obtained as a tensor product of one-dimensional interpolations. Two types of interpolation have been implemented: (i) a local cubic interpolation that uses four points on a coarse grid to find fine-grid values between the second and the third coarse-grid point, and (ii) a spline interpolation, using B-splines functions [37]. The RMG method can also employ non-orthogonal meshes. For example, a hexagonal discretization of the Kohn–Sham equations was described in [1].

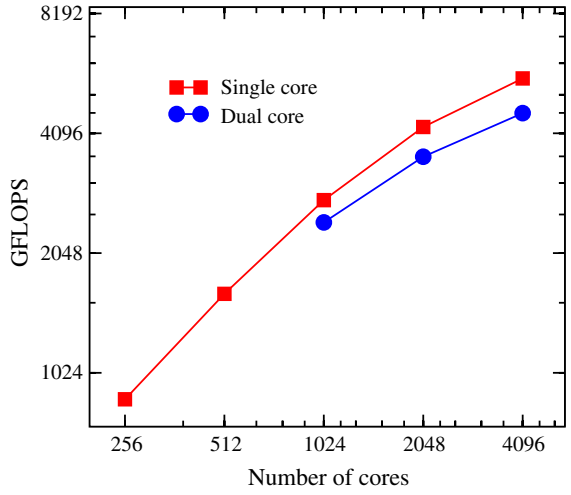

Figure 1. Performance data for a system of 819 water molecules on Cray XT3. Points represent data from calculations, lines are guides to an eye.

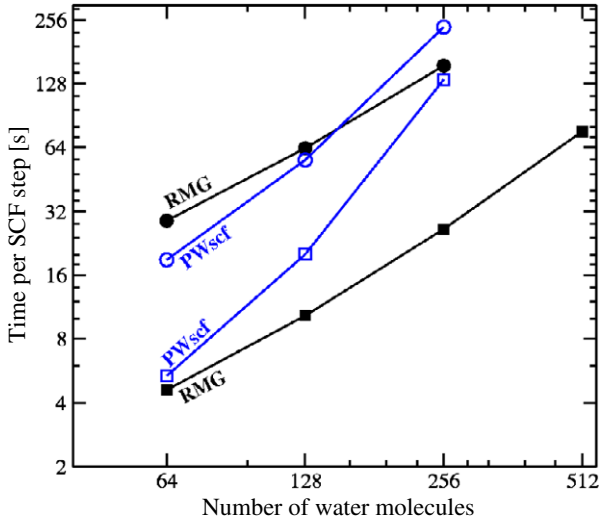

Figure 2. Times per SCF step in RMG and PWscf calculations for supercells of liquid water containing 64, 128, 256 and 512 molecules. Circles represent calculations in which the ratio of 8 cores/molecule was used, while the squares display calculations for the ratio of 1 core/molecule. The runs used a Cray XT3 supercomputer, consisting of dual-core 2.6 GHz AMD Opteron processors, with 2 GB of memory per core. We were not able to complete a PWscf run for 512 water molecules, hence this point is missing.

In Vanderbilt's UPP scheme, it is often required to perform inner products between the wavefunctions and the non-local projectors, or their derivatives with respect to the ionic positions. As mentioned above, the wavefunctions are smooth, while the non-local projectors usually vary rapidly in inside the core regions. To perform these inner products accurately and efficiently, the double grid technique [38] is used. In this technique, an inner product calculated on a grid that is denser than the original coarse grid is evaluated on the original grid with non-local projectors replaced by weight factors, which depend on the interpolation used to obtain values of wavefunctions on the denser grid.

Another important aspect of the real-space implementation of the UPPs is that since the spacing in the coarse grid is generally much larger than the one used when utilizing norm-conserving pseudopotentials (NCPPs), the fourth-order Mehrstellen discretization used with NCPPs [1] for the Kohn–Sham equations is not accurate enough when using UPPs. Instead of the fourth-order expression, we use a sixth-order Mehrstellen discretization, which samples the functions at 57 points.

4. Parallelization and scaling

Modern parallel supercomputers contain tens of thousands of processors and are capable of running at hundreds of TFLOPS. These numbers are expected to increase rapidly over the next few years. To utilize these resources for simulations of large systems, efficient parallelization over many processors is essential. To test parallelization efficiency, performance tests were run on two liquid water systems on the Cray XT3 supercomputer using up to 4096 processors. The XT3 supercomputer consist of 2.6 GHz dual-core AMD Opteron processors and 4 GB of memory per processor is available. The processors can be used in single-core or dual-core modes. In the dual-core mode, each core is utilized as an independent CPU using 2 GB memory, while in the single-core mode only one of the cores is active, allowing it to use 4 GB of memory.

The performance data for a calculation on on a system containing 819 water molecules are shown in the figure 1. The single-core calculation shows slightly better performance than the dual-core one. This is mainly due to the fact that in the dual-core mode the cost of interprocess communication is higher; our timings show that the time required for communication in the dual-core mode is about 40% larger than the time in the single-core node. The reason for this difference is sharing of the interconnect by the cores on a processor: when both cores use the same interconnect the communication is less efficient than when each core has its dedicated interconnect. On the new Cray XT4, our code performs about 20% faster than on the XT3, due to improved speed of memory to processor communications.

We have also made brief performance comparisons between the RMG code and a well-parallelized plane-wave-based code, PWscf [39]. While these tests were not extensive and focused only on a single system, liquid water, they show that both codes execute at comparable speed. This is not surprising, as the real and reciprocal spaces are conjugate to each other. The specific timings, shown in figure 2, were carried out in supercells containing 64, 128, 256 and 512 water molecules. The number of processor cores was increased with the number of molecules, so that the ratio of molecules per core remained constant. All tests used the PBE [40, 41] exchange–correlation, ultrasoft pseudopotentials and a kinetic energy cutoff of 36 Ryd. The data show that with increasing system size and number of cores the PWscf time per SCF step increases quite steeply, presumably due to the cost of Fourier transforms, while the RMG's time increases linearly.

At this time, the RMG code uses a simple, linear mixing scheme, which needs more SCF steps for convergence than the significantly more sophisticated algorithm used in PWscf. We made no attempt to optimize the codes for the system at hand, which would likely improve their performance for a given processor range and system size.

5. Hybrid Kohn–Sham DFT/frozen-density orbital-free DFT

In simulations on systems with biological relevance, it is usually necessary to include solvent molecules. In a typical biosimulation, the solvent molecules outnumber the solute and make treatment of such systems with *ab initio* methods prohibitively expensive. Therefore, there is great interest in developing efficient solvent models that can be used in *ab initio* calculations.

To this end, we developed [42] a simplified solvent model, in which the solvent molecules are described with a frozen-density orbital-free (FDOF) DFT method. These molecules are assumed to have rigid geometries and frozen electron densities. The energy is given as follows:

$$E_{\mathrm{FDOF}}[\varrho] = T_{\mathrm{OF}}[\varrho] + E_{\mathrm{H}}[\varrho] + E_{\mathrm{xc}}[\varrho] + E_{\mathrm{PPlocal}}[\varrho] + E_{\mathrm{II}} \tag{11}$$

where T_{OF} stands for an orbital-free (OF) kinetic energy (KE) functional, E_{H} is the Hartree energy, E_{xc} is the exchange correlation energy, E_{PPlocal} evaluates the energy of interaction between the electrons and the local part of the pseudopotentials, and finally E_{II} denotes the interaction energy between ions in the system. Any OF KE functional can be used in the formula above, but we have found that the LLP [43] functional is preferable in calculations on liquid water:

$$T_{\mathrm{LLP}} = C_{\mathrm{TF}} \int_{\Omega} \varrho^{5/3} \left(1 + \frac{0.0045x^2}{1 + 0.0253x \sinh^{-1} x}\right) \mathrm{d}\mathbf{r}, \tag{12}$$

where

$$x = 2^{1/3} \frac{|\nabla \varrho(\mathbf{r})|}{\varrho^{4/3}(\mathbf{r})}. \tag{13}$$

Note that only the local part of the pseudopotential is used in equation (11), the reason being that the non-local part of the pseudopotential is applied to orbitals rather than charge density. Neglecting the non-local contributions is well justified, because the present model only deals with intermolecular interactions (molecules are assumed to have frozen geometries) and the short-range non-local effects are negligible at typical intermolecular distances.

We chose to represent the charge density for each solvent molecule by Gaussians centered on the ionic positions. For the case of water the electron density of a molecule is described as

$$\varrho_{\mathrm{H_2O}}(\mathbf{r}) = G(q_{\mathrm{O}}, \alpha_{\mathrm{O}}, \mathbf{R}_{\mathrm{O}}, \mathbf{r}) + G(q_{\mathrm{O}}, \alpha_{\mathrm{O}}, \mathbf{R}_{\mathrm{H1}}, \mathbf{r}) + G(q_{\mathrm{O}}, \alpha_{\mathrm{O}}, \mathbf{R}_{\mathrm{H2}}, \mathbf{r}), \tag{14}$$

where G is a Gaussian function centered at the atomic position $\mathbf{R}$

$$G(q, \alpha, \mathbf{R}, \mathbf{r}) = q \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp\left(-\alpha |\mathbf{r} - \mathbf{R}|^2\right). \tag{15}$$

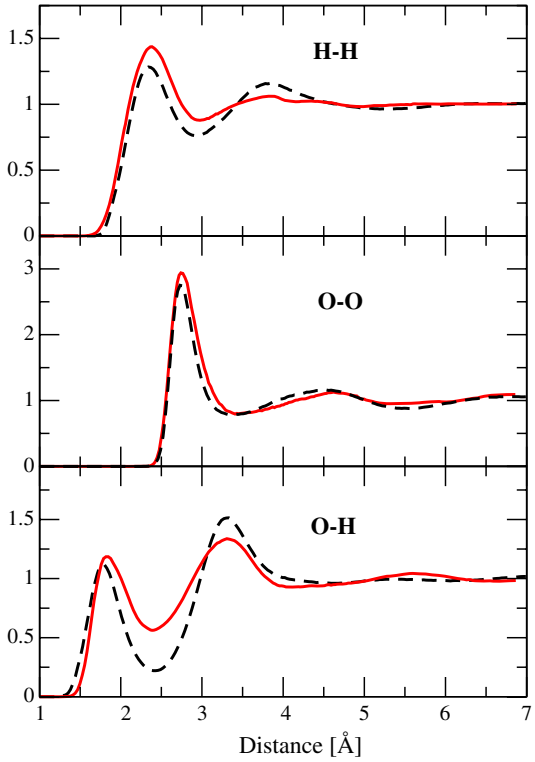

Figure 3. Radial distribution functions calculated using the frozen-density method (solid lines) for a system of 432 water molecules at 300 K. In each simulation, 10 000 molecular dynamics steps with a time step of 50 au were performed. The RDFs were calculated every 5 steps and averaged over the entire run. Experimental results [50] are also shown (dashed lines). For better comparison, intramolecular features in H–H and O–H plots were removed from the experimental data.

To perform calculations for liquid water, four parameters have to be determined: the charges on the oxygen and the hydrogens (q_{O} and q_{H}) and the Gaussian widths (α_{O} and α_{H}). The total molecular charge and dipole moment can be used on find those parameters. When using pseudopotentials, the total charge of a water molecule is 8 and therefore

$$q_{\mathrm{O}} + 2 * q_{\mathrm{H}} = 8. \tag{16}$$

The values of a dipole moment of a water molecule in a liquid phase range from 2.6 to 3.15 D [44–47]. In this work, a value of 3.0 D is used, which is close to the value of 2.95 D obtained from KS DFT studies [48, 49]. To obtain the correct molecular dipole moment the atomic charges have to be $q_{\mathrm{O}} = 7.05$ and $q_{\mathrm{H}} = 0.475$ electrons.

Gaussian width are adjusted by performing molecular dynamics simulations and matching the resulting radial distribution functions (RDFs) to experimental data. The values of $\alpha_{\mathrm{O}} = 0.75$ and $\alpha_{\mathrm{H}} = 0.8$ reproduce the experimental RDFs best. The obtained RDFs and their comparison to experimental results are shown in figure 3.

To define a hybrid calculation scheme combining the frozen-density, orbital-free method with the Kohn–Sham method, consider a system that contains two types of atoms: those that are treated with the FDOF method (FDOF atoms)

and those that are treated with KS DFT method (KS atoms). The FDOF atoms have fixed charge density, $\varrho_{\mathrm{FDOF}}(\mathbf{r})$, attached to them, while the charge density corresponding to the KS atoms, $\varrho_{\mathrm{KS}}(\mathbf{r})$, minimizes the energy functional for the hybrid system. The total charge density of the entire system is simply the sum of both densities

$$\varrho_{\mathrm{tot}} = \varrho_{\mathrm{FDOF}} + \varrho_{\mathrm{KS}}. \tag{17}$$

Since the Hartree and exchange–correlation energies are explicit functions of the charge density, it is straightforward to calculate these quantities for the hybrid system using ϱ_{tot}. The atom–atom interaction energy is also straightforward, being the sum over all atoms present in the hybrid system, regardless of their type. As for as the pseudopotential energy, the local and non-local parts are treated differently: each atom has a local potential associated with it and it acts on the total density ϱ_{tot}, while only the KS atoms have non-local potentials associated with them and these only act on the KS orbitals.

Defining KE of the hybrid system is somewhat more complicated. Taking the sum of the FDOF and KS KEs is not sufficient, since kinetic energies are non-additive. The non-additive part of KE can be estimated [51] as

$$T_{\mathrm{nadd}}[\varrho_{\mathrm{FDOF}}, \varrho_{\mathrm{KS}}] = T_{\mathrm{OF}}[\varrho_{\mathrm{tot}}] - T_{\mathrm{OF}}[\varrho_{\mathrm{FDOF}}] - T_{\mathrm{OF}}[\varrho_{\mathrm{KS}}]. \tag{18}$$

This gives the following functional for the total electron energy of a hybrid system:

$$\begin{aligned} &E_{\mathrm{Hybrid}}[\varrho_{\mathrm{FDOF}}, \varrho_{\mathrm{KS}}, \{\mathbf{R}_i\}_{\mathrm{FDOF}}, \{\mathbf{R}_i\}_{\mathrm{KS}}] \\ &\quad = T_{\mathrm{TF}}[\varrho_{\mathrm{TF}}] + T_{\mathrm{KS}}[\varrho_{\mathrm{KS}}] + T_{\mathrm{nadd}}[\varrho_{\mathrm{FDOF}}, \varrho_{\mathrm{KS}}] \\ &\qquad + E_{\mathrm{H}}[\varrho_{\mathrm{tot}}] + E_{\mathrm{xc}}[\varrho_{\mathrm{tot}}] + E_{\mathrm{PP}_{\mathrm{nonloc}}}[\varrho_{\mathrm{KS}}, \{\mathbf{R}_i\}_{\mathrm{KS}}] \\ &\qquad + E_{\mathrm{PP}_{\mathrm{loc}}}[\varrho_{\mathrm{tot}}, \{\mathbf{R}_i\}] + E_{\mathrm{II}}[\{\mathbf{R}_i\}] \end{aligned} \tag{19}$$

here $\{\mathbf{R}_i\}$ denotes all atoms in the system and $\{\mathbf{R}_i\}_{\mathrm{KS}}$ stands for atoms treated by the KS DFT. Note that in this functional ϱ_{FDOF} is rigidly tied to the positions of FDOF atoms, while ϱ_{KS} is determined by minimizing the energy functional above.

An important feature of our method is that it allows for for the flow of molecules across the interface between the regions treated by the different methods. Being able to deal with a change in the number of KS atoms is important, since in dynamical calculations the number of solvent molecules in first solvation shells or within the KS atomic cell can change. It is also possible that a solvent molecule originally treated by the KS method moves out of the KS atomic cell. Such molecule cannot be treated by the KS DFT anymore, since it is required that all KS atoms are inside the KS atomic cell, and the number of KS atoms has to be decreased. When such an event occurs, the KS/FDOF simulation can readily continue, since the total electron energy in KS/FDOF system (equation (19)) can be calculated for any number of KS and FDOF molecules. However, the exchange causes discontinuities in the total electron energy, due to the fact that the electron energy of a solvent molecule is different when calculated with KS or FDOF methods. Nevertheless, these discontinuities can be simply removed by subtracting differences in total energy before and after each exchange. This procedure yields essentially constant total energy and smooth kinetic energy curves as a function of simulation time.

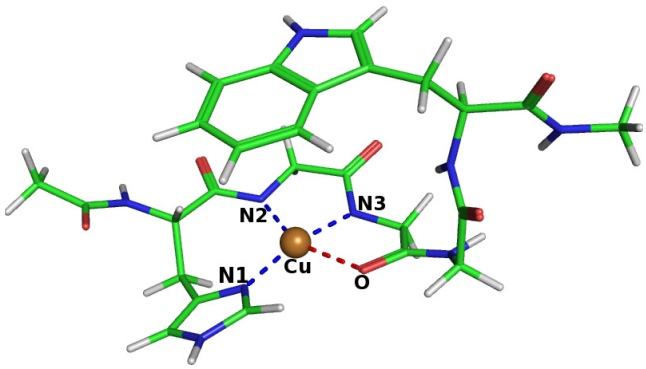

Figure 4. The relaxed structure of Cu^{2+} bound to a fragment of PrP protein. The copper ion is shown as a sphere and the binding atoms are labeled N1, N2, N3 (nitrogen) and O (oxygen).

The hybrid method was already used to study copper ion binding to prion protein, PrP, which is responsible for infectious neurodegenerative diseases such as the mad cow disease or the Creutzfeldt–Jakob disease. Experimental studies have determined that the fundamental copper binding site consists of aminoacids HGGGW [52]. In our simulation the PrP is modeled as a fragment consisting of these five aminoacids. It is solvated in a cubic box with 3101 water molecules. The water molecules closer than 6 Å to the copper ion are treated by KS DFT, which adds 12 water molecules to the KS subsystem, so that the total number of KS atoms is 108. The obtained relaxed structure is shown in figure 4. The calculated binding site geometry is quite similar to that obtained in [53]. The Cu–N bonds differ only by few hundreds of angstroms, while the Cu–O bond is 2.35 Å instead of 2.09 Å. Significant part of this difference appears to be due to the different treatment of solvent in [53], where only the first solvation shell was explicit and the rest of the solvent was modeled implicitly. In our calculation two water molecules are found in contact with the copper ion, which are stabilized by hydrogen bonding with other water molecules present in the system. No water molecules close to the copper are reported in [53]. When the calculation is performed without the water molecules in contact with copper, the difference in Cu–O bond distance decreases by more than a half to only 0.11 Å, while other bonds involving copper stay almost unchanged.

Other recent applications of the hybrid method include studies of multiple histidine coordination of Cu in prion proteins [54] and of Cu binding to the protein precursor of non-amyloid-beta component of Alzheimer disease amyloid plaque [55].

6. Quantum transport calculations

In evaluating quantum transport of electrons across a nanoscale constriction or device, one needs to consider open boundary conditions for the wavefunctions, since the nanodevice is part of a macroscopic circuit and is thus connected to leads that appear infinite on the quantum scale. The appropriate quantum-mechanical formulation needs to incorporate the scattering of the electrons using either the Lippmann–Schwinger or the Green's function equations [5, 56]. In

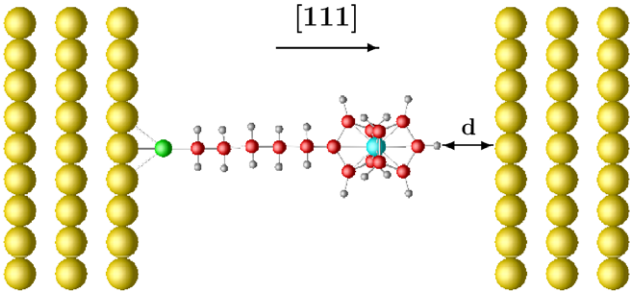

Figure 5. Schematic view of the Au(111)-Fc-C_5S-Au(111) junction. Semi-infinite Au(111) on left side represents the substrate, while that on right side represents the STM tip. d is the distance between the tip and the molecule.

order to include atomistic details of the infinite leads in either approach, it is convenient to use LCAO-type basis sets, whose finite range enables the calculation of the Green's functions or scattering matrices through matrix operations. The important advantage of our optimized-orbital method is the small number of orbitals per atom (e.g., 4 per Si atom), which leads to very small matrices. Furthermore, the various contributions to the Green's function and the transmission are evaluated in a layer-by-layer fashion, taking advantage of the localization of the basis and the small number of optimized orbitals per atom that are needed for full DFT accuracy. The infinite leads and the 'conductor' (device) region are split into layers that interact only with their nearest neighbors, leading to exact O(N) scaling of the quantum transport part [6]. Due to the efficient expansion of Hamiltonian and Green's function matrices, the dimensions of the relevant matrices are minimized and we can perform fully *ab initio*, non-equilibrium calculations of quantum transport for large systems [7, 57].

The quantum transport calculations are carried out using the non-equilibrium Green's function (NEGF) method [58, 59] in a basis of optimally localized orbitals [34, 6]. Both the leads and the conductor parts are treated at the state-of-the-art density-functional theory level.

In the NEGF SCF calculations, the density matrix ρ_{ij} is calculated by

$$\rho_{ij} = \frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{d}\epsilon\{[G(\epsilon)\Gamma_{\mathrm{L}}(\epsilon)G^{\dagger}(\epsilon)]_{ij} f(\epsilon-\mu_{\mathrm{L}}) + [G(\epsilon)\Gamma_{\mathrm{R}}(\epsilon)G^{\dagger}(\epsilon)]_{ij} f(\epsilon-\mu_{\mathrm{R}})\} \quad (20)$$

with complex energy contour integration. The charge density is calculated according to equation (6).

The Hartree potential is obtained by solving the Poisson equation with boundary conditions corresponding to each bias, i.e., the potential in the conductor region is self-consistently matched on each side with the chemical potentials of the leads. Finally, the transmission coefficients are calculated from

$$T(E,V) = \frac{2e^2}{h}\mathrm{Tr}[\Gamma_{\mathrm{L}}(E)G_{\mathrm{C}}^{+}(E)\Gamma_{\mathrm{R}}(E)G_{\mathrm{C}}^{-}(E)], \quad (21)$$

where $\Gamma_{\mathrm{L,R}}$ and $G_{\mathrm{C}}^{\pm}$ are the coupling functions for the left and right leads and the retarded and advanced Green's functions

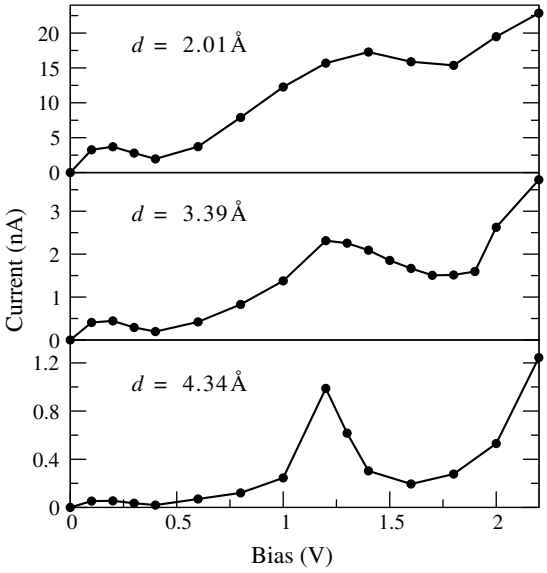

Figure 6. Current–voltage characteristics of the Au(111)-SC_5-Fc-Au(111) junction as a function of the STM-molecule distance.

of the conductor part, respectively. The current through the molecular junction is given by

$$I(V) = \int_{-\infty}^{\infty} T(E,V)[f(E-\mu_{\mathrm{L}}) - f(E-\mu_{\mathrm{R}})]\,\mathrm{d}E, \quad (22)$$

where f is the Fermi–Dirac distribution and $\mu_{\mathrm{L,R}}$ are the chemical potentials of the left and right leads, with $\mu_{\mathrm{L}} - \mu_{\mathrm{R}} = V$ being the effective applied bias.

Figure 5 shows the structure of ferrocenyl-alkanethiolate adsorbed on the Au(111) surface in a geometry-optimized atomic configuration. The Au(111) surface on the right side is used to represent the STM tip. Figure 6 shows the calculated I–V curves for positive bias at distances of 2.01, 3.39 and 4.34 Å, respectively. Their overall shapes are similar and the NDR positions are at about same biases. However, the absolute value of the current decreases exponentially with an increase of the distance. For example, at the bias of 0.6 V, we can fit the current by $I(d) = I_0 \mathrm{e}^{-\beta d}$ with the decay constant $\beta = 0.90$. The distance also affects the shape of the NDR region, or its 'strength.' Specifically, the NDR at 1.2 V is enhanced with increasing d. This indicates that a low current setpoint in the STM experiment is important for the observation of NDR at large biases. However, an increase of the distance has the opposite effect on NDR at low bias; the peak at 0.2 V is attenuated with increasing d. This result can explain why the low bias NDR is not seen in the STM experiment [60], because a high current setpoint would be needed for its observation. This 'NDR tuning' effect can be utilized in the design of molecular devices, for example by introducing spacer layers, which would adjust the strength of the NDR and of the switching current for optimal device performance.

Our other recent applications include a study of thiol and amine end group functionalization in alkane chains on gold [61] and of tunable NDR effects in porphyrin molecules sandwiched between doped Si leads [62].

7. Summary

We have briefly reviewed the salient features and recent applications of the real-space multigrid (RMG) method. This method is by now well established and has been applied to a large number of systems. It is fully parallelizable and scalable. The multigrid techniques provide convergence acceleration at all length scales, which is particularly useful for 'difficult' or large systems. Our codes, which have been especially adapted for massively parallel supercomputers, scale very well on the parallel platforms that we have access to, including the Cray XT3, XT4 and IBM Power 4/5 systems. The grid description is also very advantageous for applying O(N) techniques without significant loss of accuracy. Indeed, it is possible to reach full DFT accuracy while essentially preserving the linear scaling and reasonable computing cost. Furthermore, the variationally optimized localized basis for O(N) calculations is very suitable for quantum transport calculations, since it minimizes the basis set size needed to represent the various operators that are needed to obtain the electron transmission. For simulation of biomolecules in solution, it is possible to introduce a hybrid method where the biomolecule and the neighboring solvent shells are treated by full DFT, while the rest of the solvent is described by an approximate frozen-density orbital-free DFT method. The inclusion of the solvent into the RMG method is very cheap and O(N), enabling calculations with over 100 000 solvent molecules at minimal cost.

References

[1] Briggs E L, Sullivan D J and Bernholc J 1996 *Phys. Rev.* B **54** 14362–75
[2] Kleinman L and Bylander D M 1982 *Phys. Rev. Lett.* **48** 1425
[3] Vanderbilt D 1990 *Phys. Rev.* B **41** 7892
[4] Goedecker S 1999 *Rev. Mod. Phys.* **71** 1085
[5] Datta S 1995 *Electronic Transport in Mesoscopic Systems* (Cambridge: Cambridge University Press)
[6] Buongiorno Nardelli M, Fattebert J-L and Bernholc J 2001 *Phys. Rev.* B **64** 245423
[7] Lu W, Meunier V and Bernholc J 2005 *Phys. Rev. Lett.* **95** 206805
[8] Tomasi J and Persico M 1994 *Chem. Rev.* **94** 2027–94
[9] Rivail J L and Rinald D 1995 *Liquid State Quantum Chemistry: Computational Applications of the Polarizable Continuum Models* (Singapore: World Scientific)
[10] Cramer C J and Truhlar D G 1996 *Continuum Solvation Models* (Dordrecht: Kluwer)
[11] Fattebert J L and Gygi F 2002 *J. Comput. Chem.* **23** 662–6
[12] Begue D, Carbonniere P, Barone V and Pouchan C 2005 *Chem. Phys. Lett.* **416** 206–11
[13] Guimarães L, Avelino de Abreu H and Duarte H A 2007 *Chem. Phys.* **333** 10–7
[14] Bonifassi P, Ray P C and Leszczynski J 2006 *Chem. Phys. Lett.* **431** 321–5
[15] Zhou R and Berne B J 2002 *Proc. Natl Acad. Sci.* **99** 12777–82
[16] Rhee Y M, Sorin E J, Jayachandran G, Lindahl E and Pande V S 2004 *Proc. Natl Acad. Sci.* **101** 6456–61
[17] Collatz L 1960 *The Numerical Treatment of Differential Equations* (Berlin: Springer) p 164
[18] Brandt A 1977 *Math. Comput.* **31** 333
[19] Briggs W L 1987 *SIAM Tutorial (Philadelphia, PA)*
[20] Payne M C, Allan D C, Teter M P, Arias T A and Joannopoulos J D 1992 *Rev. Mod. Phys.* **64** 1045
[21] Galli G and Parrinello M 1992 *Phys. Rev. Lett.* **69** 3547
[22] Wang L W and Teter M P 1992 *Phys. Rev.* B **46** 12798
[23] Kohn W 1993 *Chem. Phys. Lett.* **208** 167
[24] Mauri F, Galli G and Car R 1993 *Phys. Rev.* B **47** 9973
[25] Ordejon P, Drabold D A, Grumbach M P and Martin R M 1993 *Phys. Rev.* B **48** 14646
[26] Stechel E B, Williams A R and Feibelman P J 1994 *Phys. Rev.* B **49** 10088
[27] Kim J N, Mauri F and Galli G 1995 *Phys. Rev.* B **52** 1640
[28] Li X P, Nunes R W and Vanderbilt D 1993 *Phys. Rev.* B **47** 10891
[29] Daw M S 1993 *Phys. Rev.* B **47** 10895
[30] Nunes R W and Vanderbilt D 1994 *Phys. Rev.* B **50** 17611
[31] Goedecker S and Colombo L 1994 *Phys. Rev. Lett.* **73** 122
[32] Kohn W 1996 *Phys. Rev. Lett.* **76** 3168
[33] Millam J M and Scuseria G E 1997 *J. Chem. Phys.* **506** 5569
[34] Fattebert J L and Bernholc J 2000 *Phys. Rev.* B **62** 1713
[35] He L and Vanderbilt D 2001 *Phys. Rev. Lett.* **86** 5341–4
[36] Hodak M, Wang S, Lu W and Bernholc J 2007 *Phys. Rev.* B **76** 085108
[37] De Boor C 1978 *A Practical Guide to Splines* (New York: Springer)
[38] Ono T and Hirose K 1999 *Phys. Rev. Lett.* **82** 5016–9
[39] Baroni S, Dal Corso A, de Gironcoli S and Giannozzi P http://www.pwscf.org.
[40] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
[41] Perdew J P, Burke K and Ernzerhof M 1997 *Phys. Rev. Lett.* **78** 1386
[42] Hodak M, Lu W and Bernholc J 2008 *J. Chem. Phys.* **128** 014101
[43] Lee H, Lee C and Parr R 1991 *Phys. Rev.* A **44** 768
[44] Coulsonm C and Eisenbero D 1966 *Proc. R. Soc.* A **291** 445
[45] Batista E, Xantheas S and Jonsson H 1998 *J. Chem. Phys.* **109** 4546
[46] Heggie M, Latham C, Maynard S and Jones R 1996 *Chem. Phys. Lett.* **249** 485
[47] Badyal Y, Saboungi M, Price D, Shastri S, Haeffner D and Soper A 2000 *J. Chem. Phys.* **112** 9206
[48] Silvestrelli P L and Parrinello M 1999 *Phys. Rev. Lett.* **82** 3308–11
[49] Silvestrelli P L and Parrinello M 1999 *J. Chem. Phys.* **111** 3572
[50] Soper A K 2000 *Chem. Phys.* **258** 121
[51] Wesolowski T and Warshel A 1993 *J. Phys.: Condens. Matter* **97** 8050
[52] Millhauser G 2004 *Acc. Chem. Res.* **37** 79
[53] Pushie M and Rauk A 2003 *J. Biol. Inorg. Chem.* **8** 53
[54] Hodak M, Lu W and Bernholc J 2008 in preparation
[55] Rose F, Hodak M and Bernholc J 2008 in preparation
[56] Ferry D K and Goodnick S M 1997 *Transport in Nanostructures* (Cambridge: Cambridge University Press)
[57] Wang S, Lu W, Zhao Q and Bernholc J 2006 *Phys. Rev.* B **74** 195430
[58] Larade B, Taylor J, Mehrez H and Guo H 2001 *Phys. Rev.* B **64** 075420
[59] Brandbyge M, Mozos J L, Ordejon P, Taylor J and Stokbro K 2002 *Phys. Rev.* B **65** 165401
[60] Gorman C B, Carroll R L and Fuierer R R 2001 *Langmuir* **17** 6923
[61] Kim G, Wang S, Lu W, Buongiorno Nardelli M and Bernholc J 2008 *J. Chem. Phys.* **128** 024708
[62] Ribeiro F, Lu W and Bernholc J 2008 *ACS Nano* at press