Towards grid-based $O(N)$ density-functional theory methods: Optimized nonorthogonal orbitals and multigrid acceleration

J.-L. Fattebert* and J. Bernholc
Department of Physics, North Carolina State University, Raleigh, North Carolina 27695-8202
(Received 13 September 1999)

We have formulated and implemented a real-space *ab initio* method for electronic structure calculations in terms of nonorthogonal orbitals defined on a grid. A multigrid preconditioner is used to improve the steepest descent directions used in the iterative minimization of the energy functional. Unoccupied or partially occupied states are included using a density matrix formalism in the subspace spanned by the nonorthogonal orbitals. The freedom introduced by the nonorthogonal real-space description of the orbitals allows for localization constraints that linearize the cost of the most expensive parts of the calculations, while keeping a fast convergence rate for the iterative minimization with multigrid acceleration. Numerical tests for carbon nanotubes show that very accurate results can be obtained for localization regions with radii of 8 bohr. This approach, which substantially reduces the computational cost for very large systems, has been implemented on the massively parallel Cray T3E computer and tested on carbon nanotubes containing more than 1000 atoms.

I. INTRODUCTION

The relative simplicity and accuracy of density-functional theory (DFT) has enabled great progress in electronic structure calculations.[1,2] The fact that the exchange and correlation interactions between electrons can be quite accurately described by a potential that depends only on the electronic density and possibly its gradient [the local-density and generalized gradient approximations (LDA) and (GGA), respectively] is an enormous simplification that led to *ab initio* calculations for systems containing a fairly large number of atoms. However, the interest in complex and/or technological materials and structures is stimulating the development of methods that can substantially enlarge the number of atoms that can be handled by DFT techniques.

The largest calculations usually employ pseudopotentials, which eliminate the need to explicitly consider core electrons in the Kohn-Sham (KS) equations. This enables the description of the electronic wave functions in a plane wave basis, which offers substantial advantages: (i) plane waves do not depend on the atomic positions, which simplifies the calculation of the atomic geometry and makes the results more precise, and (ii) the accuracy is determined by a single parameter, the highest kinetic energy of the waves included in the calculations. The structure of the DFT equations is also advantageously exploited in the plane-wave basis: the kinetic energy is diagonal in Fourier space, while the potential terms are diagonal in real space. The transformation between the two spaces occurs via the well-known fast Fourier transform (FFT) algorithm, which is very fast on vector supercomputers and modern workstations. However, its performance slows down on massively parallel computers since FFT is a global operation.

Recently, there has been substantial interest in the so-called real-space methods, which offer the promise of even larger calculations. There are several reasons for this, the most obvious being easier parallelization, since specific regions of space could just be assigned to particular processors. More subtly, multiscale convergence acceleration techniques,[3] which can substantially reduce the number of steps in an iterative solution, are usually formulated in real space. Finally, and most importantly for the present article, a real-space formulation is necessary for the implementation of the $O(N)$ 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.

Real-space bases that employ atom-centered functions, e.g., Gaussians or localized atomic orbitals, have long been used in electronic structure calculations. However, large numerical bases are also very interesting since they allow one to attain any desired numerical accuracy by systematically increasing the number of degrees of freedom. The numerical bases used in recent calculations include finite elements,[4–8] grids,[9–17] and wavelets.[18–20] These bases allow for natural implementation of mesh refinement,[9,11,12,14,15] cluster boundary conditions, and efficient domain decomposition approaches on massively parallel computers.[12,13,15]

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 wave function of each electron can, in general, extend over the whole material, and therefore computing one wave function 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 wave functions must be orthogonal to each other and the orthogonalization or diagonalization effort will ultimately dominate, since they scale as $O(N^3)$.

Recently, 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] In the wave function-based approaches the main idea is to rewrite the total energy expression with the help of

basis functions that are unknown at the outset and are to be variationally optimized in the process of minimizing the total energy. Each function is localized, i.e., confined to a given region in space, but the confinement regions for the various functions overlap. Since overlaps between functions localized in regions sufficiently far away vanish, the number of nonzero overlaps is $O(N)$. By exploring this sparsity and avoiding the calculation of individual eigenfunctions, it is possible to evaluate the total energy in $O(N)$ operations. In the density-matrix approaches, when formulated for full density-functional calculations with large numerical bases, it is necessary to expand the density operator in a basis of localized orbitals,[34–36] in order to reduce its effective dimension. Other methods, based on the divide-and-conquer algorithm,[37,38] energy renormalization group for the density matrix,[39] or calculation of the Green's function,[40] have also been proposed. Recently, fully self-consistent $O(N)$ DFT calculations that used atomiclike orbitals as a basis have been reported.[41] A more comprehensive discussion and list of references is given in recent reviews by Galli[42] and Goedecker.[43]

The electronic structure of a physical system is easiest to describe in the basis of eigenfunctions, but any set of functions spanning the same subspace is also appropriate. However, if one imposes *localization* constraints, which require that a given function is zero outside of a prescribed localization region, these different choices are no more equivalent. When searching for a basis set that most accurately approximates the solution with the localization constraints, it is important to work with a general nonorthogonal basis set. All quantities of interest need to be expressed in this set: the total energy, the steepest descent (SD) directions, preconditioners, and forces. This excludes the direct use of methods based on the search for the eigenvectors, e.g., Lanczos', Davidson's, or inverse iteration methods.

Real-space methods are inherently local, and therefore suitable for imposing localization constraints on the basis functions that span the subspace of the occupied orbitals. In the present paper, we investigate the performance of $O(N)$ techniques in the context of grid-based DFT calculations. Let N denote the number of wave functions needed and M the number of points of the grid used to discretize the KS problem. In general, N is much smaller than M. The computational cost for tractable systems is thus dominated by those parts of the algorithm that grow like $M\times N^2$. These operations usually consist of the construction of $N\times N$ matrices (overlap, Hamiltonian, etc.) in the basis of the trial wave functions, and the update of the trial wave functions in an orthogonalization or Ritz procedure. This cost can be largely reduced if spatial localization can be imposed on the orbitals that describe the ground state of the system. Localized orbitals also reduce the memory requirement, which scales as $N\times M$, in general. Note that when $M\gg N$, a good preconditioner is also required for an efficient iterative minimization. Otherwise a convergence slowdown would be observed as the system size increases, which would make a nonaccelerated $O(N)$ method slower than a preconditioned N^3 technique.

The method proposed in this paper is closely related to several recently described techniques, particularly that of Galli and Parinello,[21] where the Kohn-Sham equations are rewritten in terms of nonorthogonal, localized orbitals expanded in a plane-wave basis. Their basic idea is to use the nonuniqueness of the representation of the ground state of the system in terms of nonorthogonal orbitals to find an optimally localized basis that accurately describes the electrons. Our approach is more general and real-space oriented: we include a multigrid preconditioner and allow for unoccupied or partially occupied states. Indeed, numerical tests on realistic systems show that the inclusion of some unoccupied states substantially accelerates convergence. The unoccupied states are included using techniques similar to the density-matrix approach for nonorthogonal orbitals.[30,34,35] However, we do not impose additional localization constraints on the density matrix for reasons of accuracy and numerical efficiency, which results in a small $O(N^3)$ part for $N\times N$ matrix operations. This approach is efficient for a wide range of system sizes. For example, in our massively parallel implementation on the Cray T3E supercomputer, the $O(N^3)$ part does not exceed 20% of the computational cost for systems containing over 1100 atoms (see Sec. IV).

In our real-space calculations on a grid, we use a compact and accurate *Mehrstellen* finite difference (MFD) scheme of order 4,[44] which has been successfully introduced in electronic structure calculations a few years ago.[12,14] It discretizes the entire differential equation, as opposed to the Laplacian, and only uses the first- and second-neighbor points on the three-dimensional grid. As a result, zero boundary conditions are easy to impose at the boundaries and the short range of the discretization also aids parallelization. However, while the methodological discussion in this paper explicitly uses the MFD discretization, most of the arguments are general and can be applied to any other real-space method. For simplicity, the discussion is also restricted to real wave functions, but extensions to complex wave functions and **k** points are straightforward.

The optimized localized orbitals are also very useful apart from facilitating exceedingly large calculations. They have already enabled efficient calculations of quantum transport properties, which rely on the expansion of Hamiltonian and Green's function matrices in terms of localized orbitals.[45] Since one can use very few short-range localized functions per atom, e.g., three for the case of carbon atoms, the dimensions of the relevant matrices are minimized, which has led to *ab initio* calculations of quantum conductance for large nanotube structures.[46]

The remainder of this paper is organized as follows. In Sec. II, the general method for minimization of the energy functional in a nonorthogonal basis is described. The multigrid acceleration scheme and the MFD discretization are also discussed. Section III focuses on the implementation of localization constraints for nonorthogonal orbitals and their effect on the scaling of the calculations. Section IV presents numerical test results for long carbon nanotubes, containing up to 1120 atoms. Some mathematical details are given in the Appendix.

II. GENERAL ITERATIVE ALGORITHM FOR NONORTHOGONAL ORBITALS

A. Basis-invariant matrix formulation

This subsection describes a general matrix formulation for iterative density-functional calculations, where the iterative updates and thus the number of iterations do not depend on the choice of the basis.

Consider the standard problem of minimizing the Hohenberg-Kohn total energy functional[1]

$$E_t=\sum_{j=1}^{N} f_j\langle\psi_j|-\tfrac{1}{2}\nabla^2|\psi_j\rangle+\tfrac{1}{2}\int V_H(\rho)\rho\,d\mathbf{r}+\int V_{ion}(\rho)\rho\,d\mathbf{r}+E_{xc}[\rho]+E_{ion-ion} \quad (1)$$

for N-occupied orbitals in the system, each represented by a one-body wave function ψ_j, with the electron charge density $\rho(\mathbf{r})=\Sigma_{j=1}^{N} f_j|\psi_j(\mathbf{r})|^2$. The first term represents the kinetic energy, the second the electrostatic repulsion due to the classical Hartree potential $V_H(\mathbf{r})=\int\rho(\mathbf{r}')/|\mathbf{r}-\mathbf{r}'|d\mathbf{r}'$, the third the interactions of the electrons with all the nuclei, the fourth the exchange and correlation contributions, and the fifth the classical electrostatic interactions between all nuclei.

The orbitals ψ_j satisfy the Kohn-Sham equations[2]

$$H\psi_j=[-\tfrac{1}{2}\nabla^2+V_{ion}+V_H(\rho)+\mu_{xc}(\rho)]\psi_j=\epsilon_j\psi_j, \quad (2)$$

which must be solved self-consistently for the N lowest eigenvalues ϵ_j, while imposing the orthonormality constraints $\langle\psi_i|\psi_j\rangle=\delta_{ij}$. $\mu_{xc}=\delta E_{xc}[\rho]/\delta\rho$ is the exchange and correlation potential. Using Eq. (2), the total energy expression is easily rewritten as

$$E_t=\sum_{j=1}^{N} f_j\epsilon_j-\tfrac{1}{2}\int V_H(\rho)\rho\,d\mathbf{r}-\int\mu_{xc}(\rho)\rho\,d\mathbf{r}+E_{xc}[\rho]+E_{ion-ion}. \quad (3)$$

In the remainder of this subsection we will assume that each orbital is fully occupied, i.e., $f_j\equiv 2$.

In practice we use norm-conserving nonlocal pseudopotentials in the Kleinman-Bylander form

$$V_{ion}=V_{loc}+V_{nl},$$

where

$$V_{nl}=\sum_{i,\alpha}|\varphi_{KB}^{i,\alpha}\rangle\langle\varphi_{KB}^{i,\alpha}|. \quad (4)$$

The index i refers to the atoms and α to the projectors of the nonlocal pseudopotentials.

Consider a trial basis of normalized functions, $\{\phi_1,\ldots,\phi_N\}$, that will define the subspace of the occupied orbitals. This basis will be refined by iterative updates, and at convergence will accurately describe the true Kohn-Sham ground state of the system. In order to derive basis-invariant expressions for the updates, it is convenient to use matrix notations. The functions ϕ_j will thus be written as columns of a matrix Φ

$$\Phi=(\phi_1,\ldots,\phi_N).$$

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 the $N\times N$ matrix that transforms Φ into the basis Ψ of orthonormal Ritz functions,

$$\Psi=(\psi_1,\ldots,\psi_N)=\Phi C. \quad (5)$$

This matrix 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

$$A^{(\Psi)}=\Psi^T A\Psi,$$

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

We then have the relation

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

In order for Eq. (5) to define a transformation to a Ritz basis, C has to be a solution of the generalized symmetric eigenvalue problem

$$H^{(\Phi)}C=SC\Lambda, \quad (6)$$

where Λ is the diagonal, real $N\times N$ matrix that satisfies

$$\Lambda=\Psi^T H\Psi.$$

The matrix C can actually be decomposed as a product $C=G^{-T}U$, where G is the Cholesky factorization of S,

$$S=GG^T,$$

and U is an orthogonal matrix. Knowing G, the generalized eigenvalue problem (6) is reduced to a standard symmetric eigenvalue problem

$$G^{-1}H^{(\Phi)}G^{-T}U=U\Lambda. \quad (7)$$

In the basis Ψ, the SD directions, along which the energy functional decreases at the fastest rate, are easy to compute. They are given by the negative residuals of the Kohn-Sham equations (2), with the current potential kept fixed, and can be expressed as the $N\times M$ matrix

$$D^{(\Psi)}=\Psi\Lambda-H\Psi, \quad (8)$$

which satisfies the relation

$$\Psi^T D^{(\Psi)}=0.$$

In the following, we consider a SD algorithm with a linear preconditioning operator K. Without preconditioning, the convergence rate can be very slow, especially if $M\gg N$. For real-space finite difference (FD) discretizations, the multigrid preconditioner described in Sec. II D greatly improves the convergence. Introducing a pseudo-time-step η, Ψ should thus be corrected at each step according to

$$\Psi^{new}=\Psi+\eta KD^{(\Psi)}.$$

Note that in this algorithm all the trial wave functions are updated simultaneously. Without preconditioning, η has to be very small (for numerical stability reasons) and the convergence can be very slow.

For optimum convergence rate, it is important to preserve the *true* steepest descent direction when working with the

nonorthogonal orbitals Φ. This direction can differ substantially from the derivative with respect to $\phi_j(\mathbf{r})$ when the basis Φ is highly nonorthogonal. Since the direction $D^{(\Psi)}$ is easy to compute, the simplest way to obtain the SD direction in the basis Φ is to use the matrix C

$$D^{(\Phi)} = D^{(\Psi)} C^{-1} = (\Psi\Lambda - H\Psi) C^{-1} = \Phi\Theta - H\Phi, \tag{9}$$

where

$$\Theta = S^{-1} H^{(\Phi)}. \tag{10}$$

The preconditioned steepest descent (PSD) direction in the nonorthogonal basis is thus

$$\delta\Phi = K D^{(\Psi)} C^{-1} = K(\Phi\Theta - H\Phi), \tag{11}$$

and the basis Φ is updated according to $\Phi^{new} = \Phi + \eta\delta\Phi$. In the particular case $K =$ Identity, Eq. (11) is equivalent to the result given in Ref. 21. Note that $\delta\Phi$ does not depend on C and therefore does not require the solution of the eigenvalue problem (6).

Note that 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 basis, if the same pseudotime step is used. In principle, the above SD directions could also be used together with a conjugate gradient (CG) method,[47] but the convergence rate of the preconditioned algorithm described below was fast enough, in all applications to date, to make the implementation of the CG approach unnecessary.

In actual calculations, the basis functions Φ are corrected at each iteration using the PSD directions (11). A new electronic density $\rho(\mathbf{r})$ is then generated according to

$$\rho(\mathbf{r}) = 2 \sum_{j,k=1}^{N} (S^{-1})_{jk} \phi_j(\mathbf{r}) \phi_k(\mathbf{r}),$$

as well as the new Hartree and exchange-correlation potentials. However, the old and the new potentials are then mixed linearly. The functional (1) is minimized iteratively until self consistency is achieved.

At each step, the kinetic energy E_{kin} can be evaluated in the nonorthogonal basis Φ as

$$E_{kin} = 2 \sum_{j,k=1}^{N} (S^{-1})_{jk} \langle \Phi_j | -\tfrac{1}{2}\nabla^2 | \Phi_k \rangle.$$

To obtain the ground-state energy E_t, the first term on the right-hand side of Eq. (3) can be expressed as a trace of Θ,

$$\sum_{j=1}^{N} 2\epsilon_j = 2\,\mathrm{Tr}(\Lambda) = 2\,\mathrm{Tr}(\Theta).$$

B. Computations with unoccupied orbitals

We have found that the inclusion of unoccupied or partially occupied orbitals substantially enhances the convergence rate (see Sec. IV). The PSD algorithm described above can be used to improve the trial subspace of N computed orbitals, regardless whether they are occupied or empty. In the computation of the electron density and the total energy, however, one needs to account for the occupations. To this end, we introduce a density-matrix formalism in the nonorthogonal basis.

For a chemical potential μ, define the $N \times N$ matrix Υ by its matrix elements

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

where f is a Fermi-Dirac distribution. The density operator $\hat{\rho}$ is then given by

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

For $T = 0$, the operator $\hat{\rho}$ is a projector onto the states of eigenvalues lower than μ. In a grid-based approach, the dimension of this discrete density matrix is given by the number of degrees of freedom, i.e., the number of points on the real-space grid. This number is, in general, so large that it is impossible to apply iterative methods that require $\rho(\mathbf{r},\mathbf{r}')$ (matrix of size $M \times M$). However, one can represent $\hat{\rho}$ in the basis Φ

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

As in Ref. 30, we introduce the matrix $\overline{\rho}^{(\Phi)}$

$$\overline{\rho}^{(\Phi)} = S^{-1} \rho^{(\Phi)} S^{-1} = C \Upsilon C^T. \tag{12}$$

This matrix can be used to obtain the expectation value $\overline{A}$ of an operator A represented in the basis Φ

$$\overline{A} = 2\,\mathrm{Tr}(\Upsilon A^{(\Psi)}) = 2\,\mathrm{Tr}(\overline{\rho}^{(\Phi)} A^{(\Phi)}).$$

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

$$N_e = 2\,\mathrm{Tr}(\overline{\rho}^{(\Phi)} S).$$

Note that if all the computed states are fully occupied, we have $\rho^{(\Phi)} = S$ and $\overline{\rho}^{(\Phi)} = S^{-1}$.

Using the solution C of the generalized symmetric eigenvalue problem (6), one can compute the matrix $\overline{\rho}^{(\Phi)}$ using Eq. (12). The electronic density is then given by

$$\rho(\mathbf{r}) = 2 \sum_{j,k=1}^{N} (\overline{\rho}^{(\Phi)})_{jk} \phi_j(\mathbf{r}) \phi_k(\mathbf{r}). \tag{13}$$

The first term on the right-hand side of Eq. (3), commonly referred to as the band-structure energy, can be computed as

$$2 \sum_{j=1}^{N} \epsilon_j f[(\epsilon_j - \mu)/k_B T] = 2\,\mathrm{Tr}(\Lambda \rho^{(\Psi)}) = 2\,\mathrm{Tr}(H^{(\Phi)} \overline{\rho}^{(\Phi)}). \tag{14}$$

Note that in the present approach $\overline{\rho}^{(\Phi)}$ is used only to obtain $\rho(\mathbf{r})$ and the energy E_t, unlike the method of Ref. 34, where it is also explicitly used to define the correction direction $\delta\Phi$. Furthermore, the dimension of the subspace Φ remains fixed during the calculations, in contrast to the meth-

ods of Refs. 27 and 36, where the rank of the overlap matrix S decreases to the number of occupied orbitals at convergence.

In the calculations of the forces on the ions, the only term that differs from calculations with eigenfunctions is the one associated with the nonlocal part of the pseudopotential. The energy term for this part is given by

$$E_{nl}=\mathrm{Tr}(\Phi^T V_{nl}\Phi\overline{\rho}^{(\Phi)}).$$

At present, we obtain the forces on the ions by evaluating a numerical derivative of the energy with respect to the ionic positions. This is done by computing changes in the total energy for small virtual displacements of the ions, without the need for full self-consistent calculations. The associated computational cost is thus insignificant.

C. *Mehrstellen* finite difference discretization

To compute an approximate numerical solution to the KS equations we use a *Mehrstellen* finite difference (MFD) method. Upon discretization, one obtains a linear algebra eigenvalue problem

$$H\boldsymbol{\psi}=\epsilon B\boldsymbol{\psi},$$

where H and B are real, sparse $M\times M$ matrices, with M being the number of points on the real-space grid. The details of the MFD methodology and its implementation on massively parallel computers using domain decomposition are described in Ref. 12.

In the above notation, Ψ and Φ denote matrices $M\times N$, the columns of which correspond to the values of wave functions on the discretization grid. The steepest descent directions are given by

$$D^{(\Psi)}=B\Psi\Lambda-H\Psi, \tag{15}$$

and

$$\begin{aligned}D^{(\Phi)}&=D^{(\Psi)}C^{-1}\\&=B\Phi\Theta-H\Phi,\end{aligned} \tag{16}$$

where

$$\Theta=(B^{(\Phi)})^{-1}H^{(\Phi)}.$$

For the MFD scheme $H^{(\Phi)}$ should be replaced in Eqs. (6), (7), and (14) by

$$H_B=S(B^{(\Phi)})^{-1}H^{(\Phi)}.$$

As shown in the Appendix, H_B is a symmetric matrix at convergence. In practice, before reaching convergence, we explicitly symmetrize H_B.

In the above approach, the electron density at the point $\mathbf{r}_i$ is given by

$$\rho_{ii}=2\sum_{j,k}\phi_j(\mathbf{r}_i)\phi_k(\mathbf{r}_i)(\overline{\rho}^{(\Phi)})_{jk}.$$

The implementation issues of this algorithm are discussed in Sec. IV.

D. Multigrid preconditioning

Obtaining $O(N)$ scaling for the most expensive parts of the calculations is useful only if its prefactor is not much larger than for $O(N^3)$ calculations. It is thus important to have a preconditioner that works well for nonorthogonal orbitals. The importance of an efficient preconditioner for electronic structure calculations with large numerical bases has been recognized in Ref. 48, in the context of the plane-wave method. For real-space electronic structure calculations several efficient iterative solvers have been developed recently,[12,14,16,17,49] based on multilevel acceleration techniques.[3] For $O(N)$ methods, the importance of preconditioning has recently been discussed in Ref. 50.

A general mathematical analysis of parallel multilevel preconditioners for discretized partial differential equations has been given in Ref. 51. Here we introduce a specific multigrid preconditioner appropriate for electronic structure calculations in a nonorthogonal basis. Given a SD direction $\mathbf{d}$, a preconditioned SD direction $\mathbf{u}=K\mathbf{d}$ is computed. The operator K represents an iterative multigrid solver (V cycle)[3] for the Poisson equation

$$-\tfrac{1}{2}\nabla^2\mathbf{u}=\mathbf{d}, \tag{17}$$

with an initial $\mathbf{u}$ given by $\mathbf{u}_0=2\alpha\mathbf{d}$, and limited to the grids with a number of points larger than N, where Jacobi iterations are used as a smoother. The coefficient α is an estimate of the smallest eigenvalue of $(H-\epsilon_j)^{-1}$ on the coarsest grid,

$$\alpha=\frac{1}{\dfrac{2}{4^L h^2}+|V_{max}-\epsilon_1|},$$

where L denotes the number of coarse grids used, V_{max} the maximum of the total potential, and ϵ_1 the lowest eigenvalue of H_B. The role of the preconditioner is to systematically reduce high-frequency components in the SD directions. This preconditioner has the important property that its efficiency does not depend on the basis we work in and is very cheap (less than 20% of a self-consistent step). Moreover, it can be applied in the same manner to localized or delocalized orbitals.

The purpose of the iterative corrections of the wave functions is to reduce as fast as possible the components orthogonal to the subspace one is solving for. Since the above preconditioner deals much better with components associated with the highest eigenvalues of the Hamiltonian than with the ones associated with the lowest eigenvalues—due to the similarity between the eigenfunctions associated with the largest eigenvalues of the Hamiltonian and those of the Laplacian operator—the larger the N the better it works. In practice, taking into account cost and efficiency, it is advantageous to introduce some unoccupied states (10–30 %) in Φ.

Figure 1 shows the convergence of the multigrid-preconditioned SD algorithm for the C_{60} molecule for different numbers of coarse grid-levels, starting from random initial functions. The convergence rate without the coarse grids, with just Jacobi smoothing iterations for Eq. (17), is already good compared to a simple SD algorithm (not plotted here).

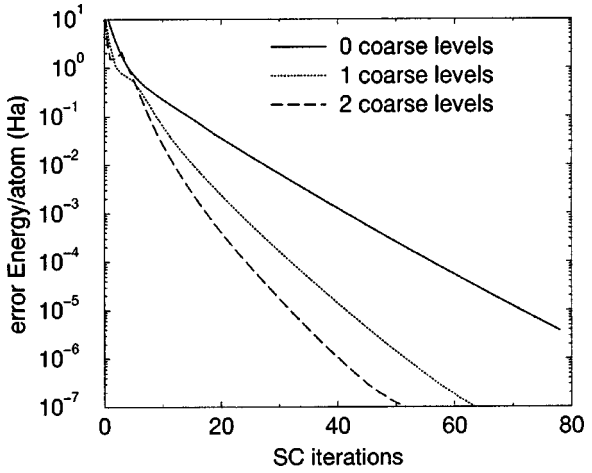

FIG. 1. Convergence rate of the multigrid-preconditioned SD for different numbers of grid levels, using the C_{60} molecule as an example. The number of orbitals N was 180.

However, the use of one or two coarse grids significantly improves the convergence rate at a very low additional cost. The good efficiency of the preconditioner allows us to avoid the conjugate-gradient algorithm, which would largely increase the cost per iteration. Note that one self-consistent (SC) iteration consists of only one simultaneous PSD correction for all the orbitals, followed by an update of the electron density and the potential. At each iteration, the Poisson equation for the Hartree potential is iteratively solved by a standard multigrid algorithm.

III. LOCALIZATION AND PARTIAL LINEAR SCALING

In the method described above, the removal of the orthogonalization step has been replaced by a more complicated and expensive SD calculation, Eq. (9), so that the computational cost of the minimization scheme is roughly the same. Furthermore, the calculation of the electron density is much more expensive when using a basis different from Ritz functions, since in the latter case $\rho^{(\Phi)}$ becomes diagonal. However, when working with nonorthogonal functions Φ instead of the Ritz functions Ψ, one can easily impose localization constraints on Φ to reduce the cost of the calculation. Localization is imposed by forcing the functions to be zero outside of a spherical region centered on an ion. In particular, such truncation will linearize the computational cost of $D^{(\Phi)}$ in Eq. (9) and ρ in Eq. (13), the most expensive parts of the algorithm. In addition, the nonorthogonal basis allows for the use of techniques that accelerate convergence for systems with a small band gap, see, e.g., Ref. 52.

The application of the MFD Laplacian operator to a wave function localized in a sphere of radius R_c generates a function localized in a sphere of radius $R_c + h\sqrt{3}$, which is used as the localization radius for $H\phi_j$ and $B\phi_j$. This truncation suppresses some components of $H\phi_j$ that are generated by the nonlocal, short-range projectors of the pseudopotential operator. These components are not used to correct the wave functions since the latter should remain localized. However, they are included exactly in the computation of the matrix Θ and the total energy by writing $H^{(\Phi)}$ as the sum of two matrices

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

which isolates the nonlocal potential V_{nl} in the second term. This term can be easily computed in $O(N)$ operations using only the values of $(B\boldsymbol{\phi}_j)^T \boldsymbol{\varphi}_{KB}^{i,\alpha}$ and $(\boldsymbol{\varphi}_{KB}^{i,\alpha})^T \boldsymbol{\phi}_j$. Therefore, the only approximation in this approach is the use of a localization radius to limit the spatial extent of each nonorthogonal orbital.

If unoccupied states are used, both inverses and eigenvectors of $N \times N$ matrices are needed. Since the eigenfunctions are, in general, not localized, the matrix C that solves Eq. (6) is not sparse, and thus the required number of operations is proportional to N^3. However, when no partially occupied states are present and an estimate of the chemical potential μ is available, one can use the method of Refs. 28–30 to compute $\overline{\rho^{(\Phi)}}$. Specifically, $\hat{\rho}$ can be defined as the operator $\hat{X}$ that minimizes the functional

$$\Omega[\hat{X}] = \mathrm{Tr}\{[3(X^{(\Psi)})^2 - 2(X^{(\Psi)})^3][(B^{(\Psi)})^{-1}H^{(\Psi)} - \mu I]\}. \tag{18}$$

This minimum can be computed iteratively by a conjugate gradient algorithm using the *true* steepest descent directions.[53] In the nonorthogonal basis Φ, defining $\overline{X^{(\Phi)}} = S^{-1}X^{(\Phi)}S^{-1}$, one has to minimize (see Appendix)

$$\Omega[\overline{X^{(\Phi)}}] = \mathrm{Tr}[(3\overline{X^{(\Phi)}}S\overline{X^{(\Phi)}} - 2\overline{X^{(\Phi)}}S\overline{X^{(\Phi)}}S\overline{X^{(\Phi)}}) \times (H_B - \mu S)]. \tag{19}$$

The localization constraints, $\overline{X_{ij}^{(\Phi)}} = 0$ if the localization regions of orbitals i and j are separated by a distance larger than a truncation radius R_ρ, can be imposed at each step of the iterative minimization in order to achieve linear scaling. This procedure is justified by the exponential decay of $\rho(\mathbf{r},\mathbf{r}')$ as $|\mathbf{r}-\mathbf{r}'| \to \infty$ in insulators or metals at a finite temperature.[54] However, the matrix elements obtained from Eq. (19) are, in general, not exactly the same one would obtain in an N^3 calculation. Moreover, good accuracy can be obtained only by keeping the number of nonzero elements in $\overline{X}$ much larger than in S.[33,34]

When $M \gg N$, the full evaluation of S^{-1}, Θ, or $\overline{\rho^{(\Phi)}}$, even with an order N^3 algorithm, constitutes a small fraction of the calculations for a large range of system sizes. For example, in parallel multigrid calculations of Ref. 12, the Cholesky decomposition and the diagonalization of the $N \times N$ matrices were not even parallelized, since they accounted for a negligible part of the computational effort for practical system sizes. When using localized orbitals, the maximum practical system size increases substantially, but the cost of the $N \times N$ matrix operations remains a small fraction of the total for a large range of systems, if these operations are effectively parallelized using standard libraries, such as PBLAS (Parallel Basic Linear Algebra Subprograms) and ScaLapack.[55] Table I shows some timings for the Scalapack diagonalization routine used in the $O(N^3)$ part of our method. For systems with more than 1000 orbitals, diagonalization on a single processor becomes very expensive, if required at each self-consistent iteration. However, since the multigrid calculations are fully parallel, it is natural to also parallelize the diagonalization, which substantially re-

TABLE I. Diagonalization times of a typical Hamiltonian matrix, using the ScaLapack subroutine PSSYEV, as a function of the number of processing elements (PE's) on the Cray T3E.

(s)	1 PE	4 PE's	16 PE's	64 PE's
$N=560$	12.8	4.5	2.6	1.9
$N=1120$	98	26.5	11.2	7.2
$N=2240$		175	60	28.8

duces the elapsed time and thus greatly increases the maximum system size that can be studied at reasonable cost.

In the current calculations, S, $H^{(\Phi)}$, Θ, and $\overline{\rho}^{(\Phi)}$ are stored as full $N\times N$ matrices, distributed on the processors according to PBLAS and ScaLapack requirements. Although most of the operations on these matrices can be optimized using their sparsity—except the diagonalization in Eq. (7)—the full storage approach is adequate for a substantial range of calculations. It is also the easiest to implement given the numerical libraries and the parallel computers we have access to. The solution of the eigenvalue problem (7) is clearly the dominant part of these $O(N^3)$ operations.

Since the method described above determines the eigenfunctions of the Kohn-Sham equations in the basis of localized functions, it can be viewed as a generalization of *ab initio* methods that use a linear combination of atomic orbitals (LCAO) or Gaussians functions G to expand the eigenfunctions: $\Psi_j=\Sigma_i c_i G_i$. The main difference is that our local functions Φ_i are defined by their values on a grid and are variationally optimized. In particular, the functions Φ_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. The number of basis functions in high-precision calculations can thus be much smaller than in Gaussian or LCAO approaches, which minimizes the $O(N^3)$ part. Furthermore, since exact and explicit $O(N^3)$ diagonalization is performed, cf. Eq. (7), both 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. Having a small number of basis functions is also important in other contexts. For example, transfer-matrix-based calculations of quantum transport benefit substantially from the small size of the basis and the optimized functions Φ have already been used to evaluate the conductances of carbon nanotubes.[46] The above approach is also very useful for evaluating the effects of localization constraints alone, while maintaining high *local* accuracy. However, for very large future calculations the $O(N^3)$ terms will eventually dominate, and the formulation (19) may become advantageous.

IV. APPLICATIONS TO TEST SYSTEMS

The algorithm described above has been implemented on the massively parallel Cray T3E supercomputer and tested on C_{60} and carbon nanotubes of varying lengths. Our calculations used the pseudopotentials of Hamann[56,57] and the LDA for μ_{xc}. The pseudopotentials were filtered in order to remove high frequency components that cannot be represented on the grids employed.[12] Periodic boundary conditions were used in all cases. The starting trial basis functions were atom-centered random functions strongly localized around their respective atoms. In order to improve the conditioning of the overlap matrix, the functions localized on the same atom were kept orthonormal during the entire calculation. The initial trial density was a linear combination of atom-centered Gaussians, neutralized by the positive ions. The density also defined the initial potential. To improve stability, the initial potential was kept fixed during the first 2-3 SC iterations. Thereafter, the potential was updated at each step, mixing usually 20–50 % of the current potential with the KS potential associated with the previous density. In our test cases the condition number of the overlap matrix S increased during the iterations, but it remained small enough to ensure stable numerical results.

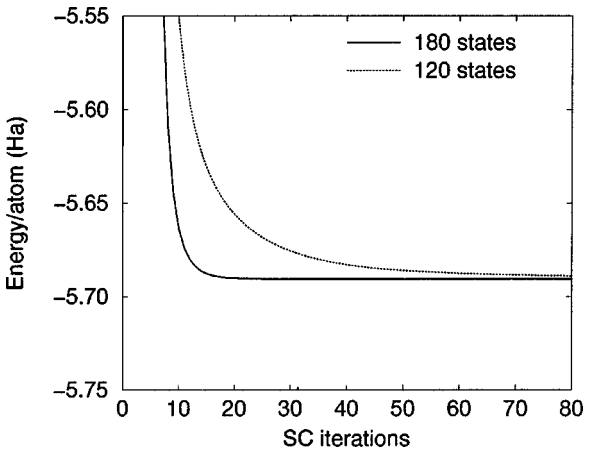

FIG. 2. The effect of unoccupied states on the convergence rate for the case of C_{60}.

To take advantage of the localization constraints, several nonoverlapping localized orbitals were stored in the same *global* array that represents a function extending throughout the whole discretization grid. These orbitals can thus be relaxed simultaneously at the same cost as the relaxation of one spatially extended orbital. In practice, when computing matrix elements between pairs of orbitals, the real-space grid was divided into regular blocks whose contributions were summed up. For this purpose, it is important that only one orbital of each global function is non-zero on each block, so that the contributions can be associated with one and only one orbital. On a massively parallel computer, a block can be a subdomain associated with a processor, or a fraction of this subdomain. The main advantage of the above approach is that it avoids indirect addressing and can efficiently use standard, optimized linear algebra libraries in each block.

As mentioned in Sec. II B, the introduction of unoccupied states in the calculations can substantially improve the convergence rate. This effect is illustrated in Fig. 2, which compares the convergence rates for C_{60} when either no or one unoccupied state per carbon atom is included in the calculations. When introducing localization constraints, the inclusion of unoccupied states has an even greater effect since these states introduce additional degrees of freedom that partially compensate for the localization-induced truncation. In the impractical limit of having as many localized functions as the number of grid points, the localized and nonlocalized

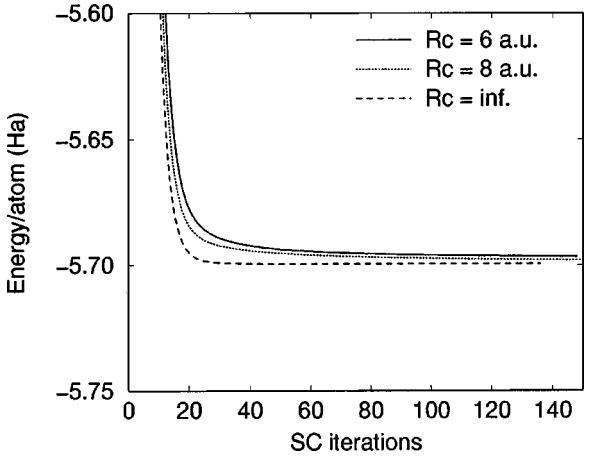

FIG. 3. Convergence rate for different localization radii for the (10,0) carbon nanotube. The supercell contained 160 atoms and the calculations used three orbitals per atom.

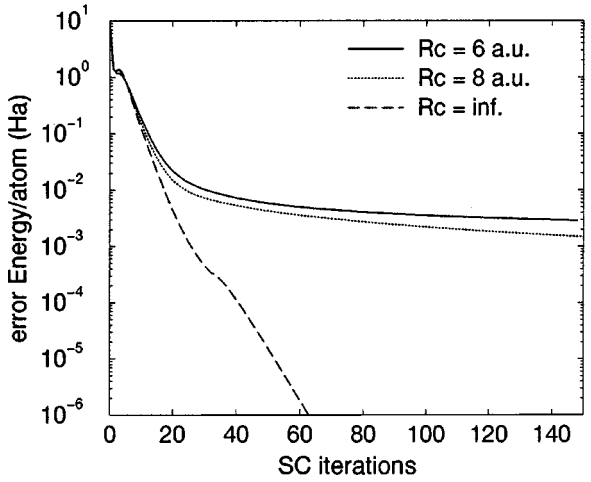

FIG. 4. The error, i.e., the difference between the converged DFT total energy without localization and the current energy as a function of the number of self-consistent iterations. Different localization radii are compared. The supercell for the (10,0) carbon nanotube contained 160 atoms.

results would be equal at convergence.

The tests with localization constraints were carried out for carbon nanotubes of various lengths. Three localized orbitals per atom were used, i.e., one unoccupied orbital per atom. The grid spacing was $h=0.34$ bohr, which was sufficiently small for the pseudopotentials employed.[56]

In the minimization of the energy functional with localization constraints, one cannot exclude the possibility of local minima.[25] However, if the number of localized functions is larger than the number of occupied orbitals, Kim *et al.*[27] observed that trapping does not occur. Moreover, if the localization regions are large enough, one can expect that the difference in energy between two local minima close to the ground state is lower than the useful accuracy—the error introduced by the localization constraint. In practice, we did not observe any apparent trapping in high-energy local minima.

At each step of the iteration process, the computed corrections were truncated in order to preserve the confinement of the orbitals in their assigned localization regions. At some point, the truncated part may then become more important than the part actually used. In order to evaluate the accuracy of the localization approximation, it is thus important to compare the Kohn-Sham energies obtained with and without localization for realistic test cases. Figures 3 and 4 show the convergence rates and the absolute error for a 160-atom nanotube. For a fixed number of orbitals, the rate of convergence decreases with increasing localization when the number of orbitals is kept fixed. Turning to accuracy, the logarithmic plot in Fig. 4 shows that an accuracy of 10^{-2} [Ha/atom] for a system of 160 carbon atoms is obtained in 20-35 iterations. Convergence towards more precise results then depends on the localization radii. Significantly more iterations are required to reach a precision of 3×10^{-3} [Ha/atom], which is sufficient to obtain a good estimate of eigenvalue differences, e.g., the highest occupied molecular orbital–lowest unoccupied molecular orbital (HOMO-LUMO) gap. The calculated value of this gap depends on the localization radius, see Table II. In particular, $R_c=8$ bohr is sufficient to reproduce the value obtained without the localization approximation.

The accuracy of 3×10^{-3} [Ha/atom] is not sufficient to reliably compute total energy differences between different atomic configurations, e.g., in order to extract defect formation energies. However, it is still adequate to compute forces and initiate atomic relaxations. Only a small number of SC iterations is required for each ionic configuration in order to have reliable forces that lead the atoms to their equilibrium positions. The wave-function components that require numerous iterations before being completely relaxed are thus improved step by step during the relaxations, and are sufficiently well converged by the time the atoms reach their equilibrium positions.

Our tests of relaxation and of total energy differences used the (10,0) carbon nanotube, either perfectly symmetric or with the (5-7-7-5) defect.[58] The supercell contained 160 atoms. All atoms were relaxed with and without localization constraints. An accurate relaxed atomic configuration could be obtained with a localization radius as small as 6 bohr. Moreover, with a localization radius of 8 bohr the absolute total energies of the relaxed supercells, with and without the defect, were shifted upwards by less than 0.02 [Ha]. The defect formation energy, defined as the difference between the two total energies, was 0.13 [Ha] both without localization and with $R_c=8$ bohr.

A contour plot of a typical localized orbital for the (10,0) nanotube is shown in Fig. 5. The orbital is strongly localized in the region of a carbon-carbon bond and smoothly decreases to zero when approaching the localization boundary. If the localization radius were reduced, one would mainly cut off the *tail* of the orbital. However, these tails are important in the evaluation of the total energy and calculations with

TABLE II. HOMO-LUMO gap for a (10,0) nanotube. The supercell contained 160 atoms.

(eV)	$R_c=5$ bohr	$R_c=6$ bohr	$R_c=8$ bohr	$R_c=\infty$
LUMO	1.53	1.27	1.17	1.17
HOMO	0.64	0.41	0.32	0.32
Gap	0.89	0.86	0.85	0.85

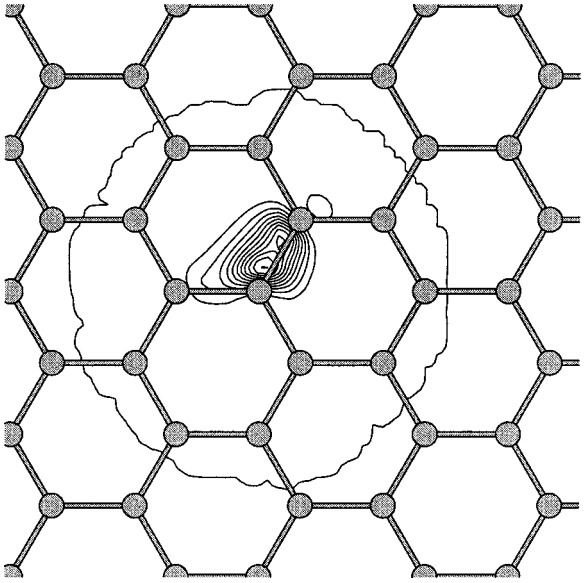

FIG. 5. Contour plot of the square of a typical localized orbital in the plane defined by the cylindrical surface of the (10,0) nanotube. The contour of lowest value (close to zero) shows the localization region with a radius of 6 bohr.

$R_c < 5$ bohr may lead to unphysical results.

The computer time for the relaxation of the nonorthogonal orbitals scales as $\sim N(R_c/h)^3$, where R_c is the localization radius and h the grid spacing. In self-consistent iterations, additional time is spent in $N \times N$ matrix operations, including the solution of Eq. (7). In Table III, we provide timings for the (5,5) nanotube with $R_c = 6.2$ bohr. Comparisons are made with a constant number of atoms per processor. With a perfect coding and linear scaling, the timings would be identical for all the examples. The solution of Eq. (7), the only strictly $O(N^3)$ part, takes less than 20% of the total time for systems with as many as 3360 orbitals. It is also clear that further restructuring and optimization of the code could result in a significant speedup.

One should stress that the localization of the nonorthogonal orbitals reduces the cost of the calculations only for systems significantly larger than the sizes of the individual localization regions. Furthermore, the minimal number of atoms for which nonorthogonal localized orbitals become advantageous depends on the number of nonzero overlaps between the orbitals, and thus on the atomic geometry.

TABLE III. Timing for 1 SC step for electronic structure calculations for (5,5) nanotubes of several lengths on the Cray T3E. For 140 atoms, the global grid is 96×56×56. The number of storage functions denotes how many *global* arrays—extended over the whole grid—were required to store all the localized orbitals.

No. of atoms	140	280	560	1120
No. of orbitals	420	840	1680	3360
No. of PEs	32	64	128	256
No. of storage func.	237	252	255	255
CPU time/PE (s)	69	82	104	173
Solution of Eq. (7) (s)	1.4	2.6	9	30

V. SUMMARY AND CONCLUSIONS

An efficient and accurate method for self-consistent *ab initio* electronic structure calculations was developed and implemented on the massively parallel Cray T3E supercomputer. The method is based on a description of the subspace spanned by the occupied orbitals in a nonorthogonal basis of variationally optimized localized functions. These functions are defined on a grid and are strictly zero outside of spheres centered on the atoms. The localization is essential to achieve a linear scaling of the computational effort for the most expensive part of the calculation—the relaxation of the basis functions and the evaluation of the electronic density. A *Mehrstellen* finite-difference scheme is used to discretize the Laplacian operator on the grid. A multigrid preconditioner has been developed in order to have a very efficient minimization scheme for basis-invariant steepest descent directions. Since it appears to be essential for a high convergence rate to include unoccupied orbitals in the calculations, a density-matrix formalism is used to define the occupation of the orbitals in the subspace defined by the nonorthogonal basis functions. This requires operations on $N \times N$ matrices that scale as $O(N^3)$, but when efficiently parallelized this part is relatively cheap when compared to the cost of the relaxation of the orbitals. Our tests show that the $O(N^3)$ part is less than 20% of the total time for carbon nanotubes with as many as 1120 atoms.

The accuracy of the calculations can be systematically improved by mesh refinement and/or by extending the localization regions of the grid-based orbitals. Numerical tests on carbon nanotubes show that accurate relaxed atomic configurations, band-gap, and total energy differences can be obtained for localization radii as small as 8 bohr.

ACKNOWLEDGMENTS

We wish to thank Dr. E. L. Briggs for discussions and providing his parallel orthogonal-orbital multigrid code for the Cray T3E. J.-L. F. gratefully acknowledges the financial support of the Fonds National Suisse de la Recherche Scientifique. This work was also supported in part by NSF and ONR. Supercomputer calculations were carried out at DoD and NC Supercomputing Centers.

APPENDIX: MATRIX EXPRESSIONS IN A NONORTHOGONAL BASIS

The relation (5) can be used to derive matrix expressions in the basis Φ. In the basis of the Ritz functions Ψ one has at convergence

$$(B^{(\Psi)})^{-1} H^{(\Psi)} = \Lambda,$$

where Λ is a real diagonal matrix. Therefore,

$$\begin{aligned}\Lambda &= (C^T B^{(\Phi)} C)^{-1} (C^T H^{(\Phi)} C) \\ &= C^{-1} (B^{(\Phi)})^{-1} H^{(\Phi)} C \\ &= C^T S [(B^{(\Phi)})^{-1} H^{(\Phi)}] C.\end{aligned}$$

The matrix C is then a solution of the generalized eigenvalue problem

$$H_B C = S C \Lambda$$

for $H_B = S(B^{(\Phi)})^{-1} H^{(\Phi)}$. Since $(B^{(\Psi)})^{-1} H^{(\Psi)}$ is symmetric at convergence and

$$\begin{aligned} H_B &= S(B^{(\Phi)})^{-1} H^{(\Phi)} \\ &= SC(B^{(\Psi)})^{-1} H^{(\Psi)} C^{-1} \\ &= C^{-T}(B^{(\Psi)})^{-1} H^{(\Psi)} C^{-1}, \end{aligned}$$

H_B is also symmetric at convergence. Furthermore,

$$(B^{(\Psi)})^{-1} H^{(\Psi)} - \mu I = C^{-1}[(B^{(\Phi)})^{-1} H^{(\Phi)} - \mu I] C.$$

Using the invariance of the trace of an operator when changing the basis,

$$\Omega[\overline{X}^{(\Phi)}] = \mathrm{Tr}[(3\overline{X}^{(\Phi)} S \overline{X}^{(\Phi)} - 2\overline{X}^{(\Phi)} S \overline{X}^{(\Phi)} S \overline{X}^{(\Phi)}) \times (H_B - \mu S)]$$

follows then directly from definition (18).

*Present address: CASC, Lawrence Livermore National Laboratory, L-551, Livermore, CA 94551.

[1] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).
[3] A. Brandt, Math. Comput. **31**, 333 (1977).
[4] B. Hermansson and D. Yevick, Phys. Rev. B **33**, 7241 (1986).
[5] S. R. White, J. W. Wilkins, and M. P. Teter, Phys. Rev. B **39**, 5819 (1989).
[6] H. Murakami, V. Sonnad, and E. Clementi, Int. J. Quantum Chem. **42**, 785 (1992).
[7] E. Tsuchida and M. Tsukada, Phys. Rev. B **52**, 5573 (1995).
[8] J. E. Pask, B. M. Klein, C. Y. Fong, and P. A. Sterne, Phys. Rev. B **59**, 12 352 (1999).
[9] J. Bernholc, J.-Y. Yi, and D. J. Sullivan, Faraday Discuss. Chem. Soc. **92**, 217 (1991).
[10] J. R. Chelikowsky, N. Trouiller, and Y. Saad, Phys. Rev. Lett. **72**, 1240 (1994); J. R. Chelikowsky, N. Trouiller, K. Wu, and Y. Saad, Phys. Rev. B **50**, 11 355 (1994).
[11] F. Gygi and G. Galli, Phys. Rev. B **52**, R2229 (1995).
[12] E. L. Briggs, D. J. Sullivan, and J. Bernholc, Phys. Rev. B **52**, R5471 (1995); **54**, 14 362 (1996).
[13] A. P. Seitsonen, M. J. Puska, and R. M. Nieminen, Phys. Rev. B **51**, 14 057 (1995).
[14] J.-L. Fattebert, BIT **36**, 509 (1996); J. Comput. Phys. **149**, 75 (1999).
[15] N. A. Modine, G. Zumbach, and E. Kaxiras, Phys. Rev. B **55**, 1337 (1997).
[16] T. L. Beck, K. A. Iver, and M. P. Merrick, Int. J. Quantum Chem. **61**, 341 (1997); T. L. Beck, *ibid.* **65**, 477 (1997).
[17] F. Ancilotto, P. Blandin, and F. Toigo, Phys. Rev. B **59**, 7868 (1999).
[18] K. Cho, T. Arias, J. Joannospoulos, and P. Lam, Phys. Rev. Lett. **71**, 1808 (1993).
[19] S. Wei and M. Y. Chou, Phys. Rev. Lett. **76**, 2650 (1996).
[20] C. J. Tymczak and X.-Q. Wang, Phys. Rev. Lett. **78**, 3654 (1997).
[21] G. Galli and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992).
[22] L.-W. Wang and M. P. Teter, Phys. Rev. B **46**, R12 798 (1992).
[23] W. Kohn, Chem. Phys. Lett. **208**, 167 (1993).
[24] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, R9973 (1993); F. Mauri and G. Galli, *ibid.* **50**, 4316 (1994).
[25] P. Ordejon, D. A. Drabold, M. P. Grumbach, and R. M. Martin, Phys. Rev. B **48**, 14 646 (1993); P. Ordejon, D. A. Drabold, R. M. Martin, and M. P. Grumbach, *ibid.* **51**, 1456 (1995).
[26] E. B. Stechel, A. R. Williams, and P. J. Feibelman, Phys. Rev. B **49**, 10 088 (1994).
[27] J. Kim, F. Mauri, and G. Galli, Phys. Rev. B **52**, 1640 (1995).
[28] X.-P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10 891 (1993).
[29] M. S. Daw, Phys. Rev. B **47**, 10 895 (1993).
[30] R. W. Nunes and D. Vanderbilt, Phys. Rev. B **50**, 17 611 (1994).
[31] S. Goedecker and L. Colombo, Phys. Rev. Lett. **73**, 122 (1994).
[32] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).
[33] J. M. Millam and G. E. Scuseria, J. Chem. Phys. **506**, 5569 (1997).
[34] E. Hernandez and M. J. Gillan, Phys. Rev. B **51**, 10 157 (1995); E. Hernandez, M. J. Gillan, and C. M. Goringe, *ibid.* **53**, 7147 (1996).
[35] T. Hoshi and T. Fujiwara, J. Phys. Soc. Jpn. **66**, 3710 (1997).
[36] W. Yang, Phys. Rev. B **56**, 9294 (1997).
[37] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).
[38] W. Yang and T.-S. Lee, J. Chem. Phys. **103**, 5674 (1995).
[39] R. Baer and M. Head-Gordon, J. Chem. Phys. **109**, 10 159 (1998).
[40] S. Baroni and P. Giannozzi, Europhys. Lett. **17**, 547 (1991).
[41] D. Sanchez-Portal, P. Ordejon, E. Artacho, and J. M. Soler, Int. J. Quantum Chem. **65**, 453 (1997).
[42] G. Galli, Curr. Opin. Solid State Mater. Sci. **1**, 864 (1996).
[43] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).
[44] L. Collatz, *The Numerical Treatment of Differential Equations* (Springer, Berlin, 1966).
[45] M. Buongiorno Nardelli, Phys. Rev. B **60**, 7828 (1999).
[46] M. Buongiorno Nardelli, J.-L. Fattebert, and J. Bernholc (unpublished).
[47] I. Stich, R. Car, M. Parrinello, and S. Baroni, Phys. Rev. B **39**, 4997 (1989).
[48] M. P. Teter, M. C. Payne, and D. C. Allan, Phys. Rev. B **40**, 12 255 (1989).
[49] S. Costiner and S. Taasan, Phys. Rev. E **52**, 1181 (1995).
[50] D. R. Bowler and M. J. Gillan, Comput. Phys. Commun. **112**, 103 (1998).
[51] J. H. Bramble, J. E. Pasciak, and J. Xu, Math. Comput. **55**, 1 (1990).
[52] N. Marzari, D. Vanderbilt, and M. C. Payne, Phys. Rev. Lett. **79**, 10 289 (1997).
[53] C. A. White, P. Maslen, M. S. Lee, and M. Head-Gordon, Chem. Phys. Lett. **276**, 133 (1997).
[54] S. Ismael-Beigi and T. A. Arias, Phys. Rev. Lett. **82**, 2127 (1999).
[55] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley, *SCALAPACK User's Guide* (SIAM, Philadelphia, 1997).
[56] D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).
[57] M. Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1989).
[58] M. Buongiorno Nardelli, B. I. Yakobson, and J. Bernholc, Phys. Rev. B **57**, R4277 (1998).