Electronic-structure calculations and molecular-dynamics simulations with linear system-size scaling

Francesco Mauri and Giulia Galli
Institut Romand de Recherche Numérique en Physique des Matériaux (IRRMA), PHB-Ecublens, 1015 Lausanne, Switzerland
(Received 22 December 1993; revised manuscript received 13 April 1994)

We present a method for total-energy minimizations and molecular-dynamics simulations based either on tight-binding or on Kohn-Sham Hamiltonians. The method leads to an algorithm whose computational cost scales linearly with the system size. The key features of our approach are (i) an orbital formulation with single-particle wave functions constrained to be localized in given regions of space, and (ii) an energy functional that does not require either explicit orthogonalization of the electronic orbitals, or inversion of an overlap matrix. The foundations and accuracy of the approach and the performances of the algorithm are discussed, and illustrated with several numerical examples including Kohn-Sham Hamiltonians. In particular, we present calculations with tight-binding Hamiltonians for diamond, graphite, a carbon linear chain, and liquid carbon at low pressure. Even for a complex case such as liquid carbon—a disordered metallic system with differently coordinated atoms–the agreement between standard diagonalization schemes and our approach is very good. Our results establish the accuracy and reliability of the method for a wide class of systems and show that tight-binding molecular-dynamics simulations with a few thousand atoms are feasible on small workstations.

I. INTRODUCTION

Many studies of materials carried out nowadays in condensed-matter physics are based on total-energy calculations and molecular-dynamics simulations with forces derived either from first principles (FP) or tight-binding (TB) Hamiltonians.[1] These computations rely on a single-particle orbital formulation of the electronic problem. Within such a framework, the calculation of the total energy amounts to the solution of a set of eigenvalue equations (e.g., the Kohn-Sham equations, in density-functional theory), which is obtained by diagonalizing the Hamiltonian matrix (H). H is usually set up according to a chosen basis set for the electronic orbitals. Both direct and iterative diagonalizations imply an overall scaling of the computational effort which grows as the third power of the number of electronic states, and thus as the cube of the number of atoms in the systems. This unfavorable scaling is a major limitation to the use of TB and FP Hamiltonians for systems containing more than a few hundred and a few thousand electrons, respectively.

Iterative diagonalizations have been utilized in the study of a variety of systems in recent years; indeed, when the number M of basis functions is much larger than the number N of electronic states these schemes are much more efficient than direct diagonalizations. There are two types of iterative approaches: constrained minimization (CM) methods[1] in which the single-particle wave functions are required to be orthonormal and unconstrained (UM) methods,[2,3] in which the orbitals are allowed to overlap. In computations with plane wave (PW) basis sets and pseudopotentials—which are the ones most widely used in, e.g., first principles molecular-dynamics simulations[4]—the evaluation of $\{H\phi\}$, i.e., of $H\phi$ for the N electronic states, requires $O(NM)$ operations (M is proportional to N). This is so if advantage is taken of fast Fourier transform techniques and of the localized nature of nonlocal pseudopotentials. The application of orthogonality constraints implies instead $O(N^2M)$ operations. When UM are used, the calculation of the overlap matrix ($\mathbf{S}$) and of its inverse are of $O(N^2M)$ and $O(N^3)$, respectively.

Recently several groups have proposed methods to overcome the problem of the so called N^3 scaling, and devised algorithms with linear system-size scaling.[5,6,2,7–14] These approaches are usually referred to as $O(N)$ methods. Some of them are based on an orbital formulation of the electronic problem,[2,7,8,12,13] whereas others rely on formulations without single-particle wave functions, but based on the direct calculation either of the one-electron Green function[6,14] or of the density matrix.[9,10]

A key idea[2] of $O(N)$ orbital schemes is to use wave functions forced to be localized in given regions of space. These regions are to be chosen appropriately, i.e., large enough so that the effect of localization constraints can be made negligible on the computed properties. The solution of the eigenvalue problem by searching the eigenstates directly is therefore abandoned in favor of a search for a linear combination of eigenstates which is localized in real space. In this way the total number of expansion coefficients used to represent the localized electronic orbitals depends linearly on the size of the system and the number of operations needed for the evaluation of $\{H\phi\}$ can be reduced[2] to $O(N)$. The idea of working with localized wave functions is directly related to that of taking advantage of the local nature of the density ma-

trix (ρ) in real space,[9,10] by considering the elements ρ_{ij} to be zero for distances larger than an appropriate cutoff (localization) radius.

In order to reduce to $O(N)$ operations not only the calculation of $\{H\phi\}$ but also iterative orthogonalization procedures or the $\mathbf{S}$ inversion, one should in principle resort to assumptions on the form of the overlap matrix. If the off-diagonal elements of $\mathbf{S}$ can be made appropriately small, with respect to its diagonal elements, then the matrix can be inverted with an iterative procedure whose number of iterations does not increase with the size of the system, and which therefore implies a number of operations scaling linearly with system size. However, the problem of imposing *explicit* orthogonalization constraints or of inverting $\mathbf{S}$ can be solved without any assumption or approximation. One can define a functional with *implicit* orthogonalization constraints, containing only the $\mathbf{S}$ matrix but not its inverse, in such a way that it has exactly the same minimum as the Kohn-Sham density functional. One can therefore use a functional which is "easier" both to evaluate and to minimize than those used in standard CM and UM methods, which nevertheless has the "correct" ground state energy and charge density. This is another key idea of the $O(N)$ orbital scheme which was introduced in Ref. 8.

Such an approach will be presented in detail in Sec. II of this paper. In Sec. III we discuss numerical results obtained for first principles calculations within density-functional theory, in the local density approximation. In Sec. IV we demonstrate that an algorithm with linear system-size scaling can be obtained when the functional with implicit orthogonalization constraints is minimized with respect to localized orbitals. Sections V and VI contain our results for the minimization of tight-binding Hamiltonians, and for molecular-dynamics simulations, respectively. Summary and conclusions are given in Sec. VII.

II. AN ENERGY FUNCTIONAL WITH IMPLICIT ORTHOGONALIZATION CONSTRAINTS

A. Definition and characterization of the energy functional

The key points of the unconstrained minimization method introduced in Ref. 8 are (i) the replacement of the inverse of the overlap matrix, entering the energy functional used in standard UM methods, with its series expansion in $(\mathbf{I}-\mathbf{S})$ up to an odd order $\mathcal{N}$, where $\mathbf{I}$ is the identity matrix; (ii) the implicit inclusion of orthonormality constraints in the energy functional, at variance with standard CM methods, where orthonormality constraints are treated explicitly, i.e., as Lagrange multipliers. After defining the energy functional which satisfies properties (i) and (ii), we will prove that (1) this energy functional has the Kohn-Sham ground state energy (E_0) as its absolute minimum and (2) its minimization yields orthonormal orbitals.

We consider an energy functional of $N/2$ overlapping orbitals $\{\phi\}$ expanded in a finite basis set, and of the $(N/2\times N/2)$ matrix $\mathbf{A}$:

$$E[\mathbf{A},\{\phi\}] = 2\left(\sum_{ij}^{N/2} A_{ij}\langle\phi_i|-\tfrac{1}{2}\nabla^2|\phi_j\rangle + F[\tilde{\rho}]\right) + \eta\left(N-\int d\mathbf{r}\tilde{\rho}(\mathbf{r})\right), \quad (1)$$

where $\tilde{\rho}(\mathbf{r}) = \tilde{\rho}[\mathbf{A},\{\phi\}](\mathbf{r}) = 2\sum_{ij}^{N/2} A_{ij}\phi_j(\mathbf{r})\phi_i(\mathbf{r})$, $F[\tilde{\rho}]$ is the sum of the Hartree, exchange-correlation, and external potential energy functionals, and η is a constant to be specified. The factor 2 accounts for the electronic occupation numbers, which are assumed to be all equal. For simplicity we consider real orbitals. According to the choice of the matrix $\mathbf{A}$, one can obtain either the functional used in standard UM methods or the energy functional which we introduced in Ref. 8. If $\mathbf{A}=\mathbf{S}^{-1}$, where $S_{ij}=\langle\phi_i|\phi_j\rangle$, then $\tilde{\rho}[\mathbf{S}^{-1}]$ is the single-particle charge density $\rho(\mathbf{r})$ and the term multiplying η is zero; in this case the functional of Eq. (1) is the total energy of interacting electrons in an external field according to density-functional theory (DFT), written for overlapping orbitals.[2,3] In particular, if the wave functions are orthonormal then $\mathbf{A}=\mathbf{S}^{-1}=\mathbf{I}$, and Eq. (1) gives the total-energy functional of DFT used in CM and *ab initio* molecular-dynamics (MD) simulations.[1] We indicate with $\{\psi\}$ and $\{\phi\}$ sets of orthonormal and overlapping orbitals, respectively, and with $E^{\perp}[\{\psi\}]$ the energy functional of CM procedures. The sets $\{\psi\}$ and $\{\phi\}$ are related by the Löwdin transformation[15] $\psi_i=\sum_j S_{ij}^{-1/2}\phi_j$ and then $E^{\perp}[\mathbf{S}^{-1/2}\phi]=E[\mathbf{S}^{-1},\{\phi\}]$. Therefore

$$\min_{\{\psi\}}E^{\perp}[\{\psi\}] = \min_{\{\phi\}}E[\mathbf{S}^{-1},\{\phi\}] = E_0. \quad (2)$$

The energy functional of $\{\phi\}$, $E[\mathbf{Q}[\{\phi\}],\{\phi\}]$, which we introduced in Ref. 8 is obtained by taking $\mathbf{A}=\mathbf{Q}$ where

$$\mathbf{Q}=\sum_{n=0}^{\mathcal{N}}(\mathbf{I}-\mathbf{S})^n \quad (3)$$

and $\mathcal{N}$ is odd. $\mathbf{Q}$ is the truncated series expansion of $\mathbf{S}^{-1}$. We note that, similarly to $E^{\perp}$ and $E[\mathbf{S}^{-1}]$, $E[\mathbf{Q}]$ is invariant under unitary transformations in the subspace of occupied states, i.e., under the transformation $\phi_i'=\sum_j^{N/2}U_{ij}\phi_j$, where $\mathbf{U}$ is an $(N/2\times N/2)$ unitary matrix.

We now prove that the absolute minimum of $E[\mathbf{Q},\{\phi\}]$ is E_0. If the orbitals are orthonormal, i.e., $S_{ij}=\delta_{ij}$, then $Q_{ij}=\delta_{ij}$ and $E[\mathbf{Q},\{\psi\}]$ coincides with $E^{\perp}[\{\psi\}]$: $\min_{\{\psi\}}E^{\perp}[\{\psi\}]=\min_{\{\psi\}}E[\mathbf{Q},\{\psi\}]$. Furthermore, since $\{\psi\}$ is a subset of $\{\phi\}$, $\min_{\{\psi\}}E[\mathbf{Q},\ \{\psi\}]\geq\min_{\{\phi\}}E[\mathbf{Q},\ \{\phi\}]$. As a consequence

$$\min_{\{\psi\}}E^{\perp}[\{\psi\}]\geq\min_{\{\phi\}}E[\mathbf{Q},\ \{\phi\}]. \quad (4)$$

This shows that E_0 is an upper bound to $\min_{\{\phi\}}E[\mathbf{Q},\ \{\phi\}]$.

We consider the difference between the functionals $E[\mathbf{Q},\ \{\phi\}]$ and $E[\mathbf{S}^{-1},\ \{\phi\}]$, i.e.,

$$\Delta E = E[\mathbf{Q},\{\phi\}] - E[\mathbf{S}^{-1},\{\phi\}] = \int_0^1 \frac{\partial E[\mathbf{A}(\lambda),\{\phi\}]}{\partial\lambda} d\lambda \ , \tag{5}$$

where $\mathbf{A}(\lambda) = \lambda(\mathbf{Q}-\mathbf{S}^{-1}) + \mathbf{S}^{-1}$. Using Eq. (1), Eq. (5) becomes

$$\Delta E = 2\sum_{ij}^{N/2} \langle\phi_j|\overline{H}_{\rm KS} - \eta|\phi_i\rangle (Q_{ij} - S_{ij}^{-1}) \ , \tag{6}$$

where $\overline{H}_{\rm KS} = -\frac{1}{2}\nabla^2 + \overline{V}_{\rm KS}$, with $\overline{V}_{\rm KS} = \int_0^1 d\lambda V_{\rm KS}[\tilde{\rho}(\lambda)]$ and $V_{\rm KS}[\tilde{\rho}] = \frac{\delta F}{\delta\tilde{\rho}}$. $\overline{H}_{\rm KS}$ is a Kohn-Sham Hamiltonian, where the self-consistent potential is averaged over the integration path (λ) of Eq. (5). Given a finite basis set for the orbitals $\{\phi\}$, one can choose η large enough so that the operator $(\overline{H}_{\rm KS} - \eta)$ is negative definite; then also the $(N/2 \times N/2)$ matrix $\langle\phi_j|\overline{H}_{\rm KS} - \eta|\phi_i\rangle$ is negative definite. Using the expression of the sum of a geometric series for $\mathbf{Q}$, we can express the difference between $\mathbf{Q}$ and $\mathbf{S}^{-1}$ as $(\mathbf{Q}-\mathbf{S}^{-1}) = -\mathbf{S}^{-1}(\mathbf{I}-\mathbf{S})^{\mathcal{N}+1} = -(\mathbf{I}-\mathbf{S})^{\mathcal{N}+1}\mathbf{S}^{-1}$. If $\mathcal{N}$ is odd, the difference between $\mathbf{Q}$ and $\mathbf{S}^{-1}$ is a nonpositive definite matrix since $\mathbf{S}$, $\mathbf{S}^{-1}$, and $(\mathbf{I}-\mathbf{S})^{\mathcal{N}+1}$ are commuting non-negative definite matrices. Therefore if η and $\mathcal{N}$ fulfill the above requirements, ΔE is non-negative since it is equal to the trace of the product of a negative and of a nonpositive definite matrix. As a consequence, for each set of $\{\phi\}$,

$$E[\mathbf{Q},\{\phi\}] \geq E[\mathbf{S}^{-1},\{\phi\}]. \tag{7}$$

The equality holds only if $(\mathbf{Q}-\mathbf{S}^{-1})$ is equal to zero and therefore only if $\mathbf{S} = \mathbf{I}$. Equation (7) shows that E_0 is a lower bound to $\min_{\{\phi\}} E[\mathbf{Q},\{\phi\}]$. From Eqs. (2), (4), and (7) we have

$$\begin{aligned} \min_{\{\psi\}} E^{\perp}[\{\psi\}] &= \min_{\{\phi\}} E[\mathbf{Q},\{\phi\}] \\ &= \min_{\{\phi\}} E[\mathbf{S}^{-1},\{\phi\}] = E_0. \end{aligned} \tag{8}$$

This proves that *the energy functional $E[\mathbf{Q}]$ has the Kohn-Sham ground state energy (E_0) as its absolute minimum*, if η and $\mathcal{N}$ fulfill the requirements discussed above.

We showed that $E[\mathbf{Q}]$ and $E[\mathbf{S}^{-1}]$ are equal only if the orbitals are orthonormal [Eq. (7)] and that at the minimum the two functionals are equal [Eq. (8)]. It then follows that *the minimization of $E[\mathbf{Q}]$ yields orthonormal orbitals.*

The choice of η which makes $(\overline{H}_{\rm KS} - \eta)$ negative definite deserves some comments. If the Hamiltonian of the system does not depend on ρ, an η larger than the Hamiltonian maximum eigenvalue $\epsilon_{\max}$ insures that $\Delta E \geq 0$. Within the local density approximation (LDA), one can prove that $H_{\rm KS}[\tilde{\rho}[\mathbf{Q}]] \leq H_H[\rho]$, where $H_H[\rho] = [-\frac{1}{2}\nabla^2 + V_H[\rho] + V_{\rm ext}]$, V_H and $V_{\rm ext}$ are the Hartree and external potentials, respectively, and $\rho = \tilde{\rho}[\mathbf{S}^{-1}]$. This follows from the property $\tilde{\rho}[\mathbf{Q}](\mathbf{r}) \leq \rho(\mathbf{r})$, valid for each point $\mathbf{r}$, and from the explicit LDA expression of the exchange and correlation energy as a function of $\rho(\mathbf{r})$. Within, e.g., a plane wave implementation with a finite cutoff, H_H has an upper bound. This insures the existence of η such that $\Delta E \geq 0$. However, in practical implementations one can choose η smaller than the upper bound of H_H. Indeed for practical purposes it is not necessary to require E_0 to be the absolute minimum of $E[\mathbf{Q}]$, but it is sufficient to require it to be a local minimum of $E[\mathbf{Q}]$; the constant η which fulfills this weaker condition is in general much smaller than the upper bound of H_H, as we will discuss in the next section.

B. Iterative minimization of the energy functional

In this section we discuss the choices of η appropriate in practical applications and the convergence rate of iterative minimizations of $E[\mathbf{Q}]$ with $\mathcal{N} = 1$, compared to that of $E^{\perp}$. For non-self-consistent Hamiltonians, we will show that, if η is larger than the Fermi energy, then E_0 is a local minimum of $E[\mathbf{Q}]$; furthermore, if a value of η is chosen which is close to the Fermi energy, the minimum of $E[\mathbf{Q}]$ and that of $E^{\perp}$ can be obtained with the same computational efficiency.

The asymptotic convergence rate of iterative minimizations of a functional $E[\{\phi\}]$ can be estimated by expanding it around its minimum E_0, up to second order in the variation of the wave functions $\{\phi\}$. As discussed, e.g., in Ref. 16, in the minimization asymptotic regime the number of integration steps to reach convergence is directly related to the ratio between the maximum and the minimum eigenvalues of the quadratic form which results from the second order expansion of $(E - E_0)$.

We consider a non-self-consistent Hamiltonian (H) and we relate its eigenvalues ($\{\epsilon\}$) to those of the quadratic expansion of $(E[\mathbf{Q}] - E_0)$. Since $E[\mathbf{Q}]$ is invariant under unitary transformations in the subspace of occupied states, a generic variation of the wave function with respect to the ground state can be written as

$$|\phi_i\rangle = |\chi_i^0\rangle + |\Delta_i\rangle, \tag{9}$$

with

$$|\Delta_i\rangle = \sum_{l\in[\mathcal{TOT}]} c_l^i |\chi_l^0\rangle. \tag{10}$$

Here $|\chi_l^0\rangle$ are the eigenstates of H, and the indices i and l belong to the set of occupied states and to the set of occupied plus empty states, respectively. We denote with $[\mathcal{OCC}]$ and $[\mathcal{EMP}]$ the sets of occupied and empty states, and with $[\mathcal{TOT}]$ the union of the two sets. c_l^i are expansion coefficients of $|\Delta_i\rangle$ over the eigenstates of H. If Eq. (9) is substituted into the expression of $\Delta E = E[\mathbf{Q}] - E_0$, the first order term vanishes, showing that for each value of η the orbitals $|\chi_i^0\rangle$ make $E[\mathbf{Q}]$ stationary. The stationary point is, in particular, an absolute minimum if $\eta \geq \epsilon_{\max}$, as shown in the previous section. One is then left with a second order term, which can be recast as follows:

$$\begin{aligned} \Delta E = &\sum_{m\in[\mathcal{EMP}]} \sum_{i\in[\mathcal{OCC}]} 2[\epsilon_m - \epsilon_i](c_m^i)^2 \\ &+ \sum_{ij\in[\mathcal{OCC}]} 8\left[\eta - \frac{(\epsilon_i+\epsilon_j)}{2}\right]\left[\frac{1}{\sqrt{2}}(c_j^i + c_i^j)\right]^2 . \end{aligned} \tag{11}$$

From Eq. (11) it is seen that the quadratic form ΔE has two sets of normal modes. The first set has eigenvalues $k_{(mi)} = 2[\epsilon_m - \epsilon_i]$, which are always positive and independent of η; they correspond to the coordinates c_m^i. These modes are associated with an increase of the total energy when the orbitals acquire nonzero components on empty eigenstates of H. They are the same as the normal modes of $(E^{\perp} - E_0)$, calculated in Ref. 17. The second set of normal modes of ΔE has eigenvalues $\bar{k}_{(ij)} = 8[\eta - \frac{(\epsilon_i+\epsilon_j)}{2}]$; they correspond to the coordinates $[\frac{1}{\sqrt{2}}(c_j^i + c_i^j)]$. These modes are associated with a change of $E[\mathbf{Q}]$ due to the overlap of the electronic wave functions; they are indeed associated with the orthogonality constraints implicitly included in the definition of $E[\mathbf{Q}]$ and they are not present in $(E^{\perp} - E_0)$.

For η larger than the highest occupied eigenvalue of H, $\epsilon_{N/2}$ (i.e., the Fermi energy), the $\bar{k}_{(ij)}$ are positive and thus E_0 is a local minimum of $E[\mathbf{Q}]$. $\eta \geq \epsilon_{N/2}$ is a weaker condition than the one required to prove Eq. (8); it is, however, a sufficient condition to ensure that the minimization of $E[\mathbf{Q}]$ leads to the correct ground state energy, provided a reasonable starting point for the minimization is chosen. This will be shown also with numerical examples in the next section.

The minimizations of $E[\mathbf{Q}]$ and $E^{\perp}$ can be obtained with the same efficiency provided the weaker condition on η is adopted. For example, one can choose $\eta \simeq \epsilon_{N/2+1}$. Under such a condition the ratio between the maximum and the minimum eigenvalues of the expansion of $(E[\mathbf{Q}] - E_0)$ and of $(E^{\perp} - E_0)$ is the same in most systems. Indeed, the eigenvalues $\bar{k}$ lie in the interval defined by the eigenvalues k, if the spread in energy of the excited states of H is four times smaller than the valence band width. This condition is satisfied in most systems of interest. This means that in practice iterative minimizations of $E[\mathbf{Q}]$ and MD simulations with $E[\mathbf{Q}]$ can be performed with the same efficiency as the corresponding calculations with $E^{\perp}$. However, if η is chosen so that E_0 is an absolute minimum of $E[\mathbf{Q}]$, the time step used in MD simulations, which is proportional[17,16] to the square root of the maximum eigenvalue of ΔE [Eq. (11)], is reduced by a factor of 2 with respect to that used in standard calculations.

The functional introduced in Sec. II A has clear advantages over standard energy functionals when conjugate and preconditioned conjugate gradient minimization procedures are used: the complication of imposing orthonormality constraints is avoided, and unlike ordinary unconstrained methods an automatic control of the $\mathbf{S}$ matrix is provided, since at the minimum $\mathbf{S} = \mathbf{I}$. Furthermore, when preconditioning of the high frequency components of the single-particle wave functions is introduced, e.g., in Car-Parrinello molecular-dynamics simulations, the integration of the electronic equation of motion does not imply any extra work, at variance with integration schemes with explicit orthogonalization constraints.[16]

C. Relationship with other functionals

The total-energy minimization scheme which we introduced in Ref. 8 is related to other approaches recently proposed in the literature for electronic-structure calculations with linear system-size scaling. In particular, Ordejón *et al.*[13] derived the same functional as that of Eqs. (1) and (3) for $\mathcal{N} = 1$ for non-self-consistent Hamiltonians. Their derivation is based on a Lagrangian formulation with explicit orthogonalization constraints, where the Lagrange multipliers (λ_{ij}) are approximated by an expression which is exact only at the minimum, i.e., $\lambda_{ij} = \langle\phi_i|H|\phi_j\rangle$. The approach presented by Ordejón *et al.*[13] is similar to that of Wang and Teter,[7] although in Ref. 7 constraints are introduced by means of a penalty function. However, the minimum of the Wang and Teter functional is E_0 only if the weight of the penalty function goes to infinity, at variance with our and Ordejón *et al.*'s functionals which at the minimum are always equal to E_0.

Instead of using an orbital formulation, Li, Numes, and Vanderbilt[9] and Daw[10] proposed a functional for total-energy minimizations within a density matrix formulation. In this case one minimizes the energy functional with respect to the density matrix, which must fulfill the idempotency condition. This condition is enforced by minimizing the total energy with respect to a *purified* version of the density matrix[9,18] $[\tilde{\rho}(r,r')]$, constructed from a trial density matrix $\rho(r,r')$ in such a way that its eigenvalues lie on the interval [0,1]. The energy functional $E[\mathbf{Q}]$ [Eqs. (1) and (3)] for non-self-consistent Hamiltonians can be rederived within the formulation of Ref. 9 if $\rho(r,r')$ is expressed in terms of the occupied single-particle wave functions, i.e., $\rho(r,r') = \sum_{i\in[occ]} \phi_i(r)\phi_i(r')$, and a purification transformation is chosen such that $\tilde{\rho} = I - (I-\rho)^{\mathcal{N}+1}$. This transformation forces the eigenvalues of $\tilde{\rho}$ to be less than 1 only if $\mathcal{N}$ is odd; one does not need to force the eigenvalues to be positive, as is done in Ref. 9, since by construction $\rho(r,r') = \sum_{i\in[occ]} \phi_i(r)\phi_i(r')$ has a number of nonzero eigenvalues equal to the number of occupied states.

III. NUMERICAL RESULTS OF FIRST PRINCIPLES CALCULATIONS

The validity of the minimization scheme presented in Sec. II was tested numerically for KS Hamiltonians within the LDA, by computing the ground state energy of Si in the diamond structure. We used an expansion coefficient $\mathcal{N} = 1$ to define the $\mathbf{Q}$ matrix entering the energy functional [see Eq. (3)]. We chose η smaller than the maximum eigenvalue of H_{KS}; this choice ensures that the iterative minimization properly converges to the ground state energy E_0, unless a pathological starting point for the electronic orbitals is chosen.

$E[\mathbf{Q}]$ was minimized by steepest descent; the derivative of the functional with respect to the single-particle orbitals is given by

$$\frac{\partial E[\mathbf{Q}]}{\partial \phi_i} = 4\sum_j^{N/2}[(H_{\mathrm{KS}} - \eta)|\phi_j\rangle(2\delta_{ji} - S_{ji}) - |\phi_j\rangle\langle\phi_j|(H_{\mathrm{KS}} - \eta)|\phi_i\rangle] \ . \quad (12)$$

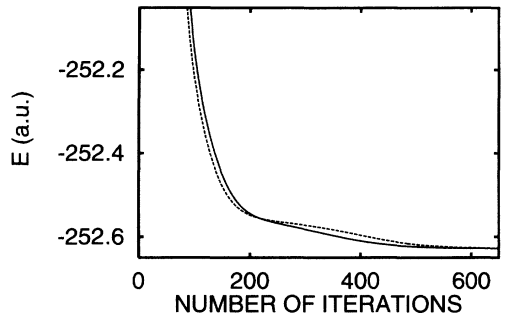

FIG. 1. Total energy (E) as a function of the number of iterations for a steepest descent minimization of 64 Si atoms in the diamond structure, described within the LDA with a PW basis set. The solid and dotted lines correspond to the minimization of $E[\mathbf{Q}]$ and $E^{\perp}$ (see text), respectively. $\mathbf{Q}$ was defined with $\mathcal{N} = 1$. We used kinetic energy cutoffs of 12 and 36 Ry for the wave functions and charge density, respectively, and we set η at 3 Ry above the top of the valence band. Each run was started from the same set of random Fourier coefficients.

The orbitals were expanded in PW's with a kinetic energy cutoff ($E_{\rm cut}$) of 12 Ry and the interaction between ionic cores and valence electrons was described by a norm-conserving pseudopotential[19] expressed in a separable form.[20] The calculation was started from orbitals set up from random numbers, with η set at 3.0 Ry above the top of the valence band. In Fig. 1 we report $E^{\perp}$ and $E[\mathbf{Q}]$ as a function of the number of iterations; it is seen that the minimizations of the two functionals require the same number of iterations and lead to the same energy. Figure 2 shows the integral of the charge density during the minimization procedure. For $\mathcal{N} = 1$, $\Delta N = N - \int d\mathbf{r}\tilde{\rho}(\mathbf{r}) = N - \mathrm{Tr}(\mathbf{QS})$ is given by

$$\Delta N = \mathrm{Tr}[(\mathbf{I} - \mathbf{S})^2]. \tag{13}$$

This is a positive quantity which goes to zero as the orbitals become orthonormal. In our calculation the difference ΔN between the total number of electrons and the integrated charge reaches a value very close to zero ($\simeq 10^{-6}$) after ten iterations, showing that the single-particle wave functions are orthonormal already well before reaching the minimum.

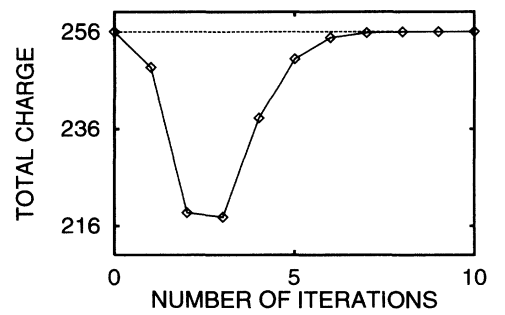

FIG. 2. Total electronic charge as a function of the number of iterations for the energy minimizations reported in Fig. 1. The total number of electrons in the system is 256.

IV. LOCALIZED ORBITALS AND AN ALGORITHM WITH LINEAR SYSTEM-SIZE SCALING

We now turn to the discussion of the approach introduced in Sec. II within a localized orbital (LO) formulation.[2] Within such a formulation, each single-particle wave function is constrained to be localized in an appropriate region of space, which we call the localization region (LR): the electronic orbitals are free to vary inside and are zero outside the LR. Different single-particle orbitals can be associated with the same LR, e.g., two doubly occupied orbitals per LR for C and Si, which have four valence electrons. The extension of a LR is determined by the bonding properties of the atomic species composing the system, and it is the same irrespective of the size of the system which is simulated. The choice of the centers of the LR's is arbitrary. In all of our calculations (see the next section) we centered the LR's on atomic sites; this choice is physically unbiased, i.e., it can be adopted for a generic system whose bonding properties are totally unknown. If one wants to take advantage of known properties of the system, LR's can, for example, be centered on atomic bonds or on positions compatible with the symmetry of the Wannier functions, if these can be defined. This is, however, difficult to do, e.g., at each step of a MD simulation, where the evolution of the bonding properties as a function of time is not known. One could also treat the centers of the LR's as variational parameters and optimize their locations during the calculation.

We now consider the minimization of $E[\mathbf{Q}]$ with respect to LO ($\{\phi^L\}$). When the orbitals are localized, S_{ij} and $\langle\phi_i^L|H_{\rm KS}|\phi_j^L\rangle$ are sparse matrices which have nonzero elements only if i and j belong to overlapping LR's. The evaluation of $E[\mathbf{Q}]$ [Eqs. (1) and (3)] as well as of $\frac{\partial E[\mathbf{Q}]}{\partial \phi_i^L}$ [Eq. (12)] implies only the calculation of matrix products containing S_{ij} and $\langle\phi_i^L|H_{\rm KS}|\phi_j^L\rangle$. No orthogonalization or $\mathbf{S}$ inversion is needed. Thus, at each step, the minimization of $E[\mathbf{Q}]$ can be performed with a number of operations which is proportional to the system size.

When localization constraints are imposed, the variational freedom of the minimization procedure is reduced. The energy obtained by minimizing a functional with respect to LO's is then larger than the absolute minimum (E_0) obtained with no constraints on the single-particle wave functions. In particular, the minimum of $E[\mathbf{Q}]$ with respect to LO $\{\phi^L\}$ does not coincide with that of $E[\mathbf{S}^{-1}]$, and the LO's which minimize $E[\mathbf{Q}]$ are in general not orthonormal. This is easily seen as follows. Whereas Eqs. (4) and (7) hold also for LO's, Eq. (2) is no longer valid when localization constraints are imposed. Indeed the transformation from $\{\psi\}$ to $\{\phi\}$ with $\mathbf{S}^{-1/2}$ does not preserve the size of the LR, i.e., it does not map functions localized in a given region onto functions localized in the

same region of space. Therefore Eq. (8) does not hold but is replaced by

$$\min_{\{\psi^L\}} E^{\perp} \geq \min_{\{\phi^L\}} E[\mathbf{Q}] \geq \min_{\{\phi^L\}} E[\mathbf{S}^{-1}] \geq E_0, \tag{14}$$

where the LR's for the $\{\psi^L\}$ and $\{\phi^L\}$ are the same. Since in Eq. (14) the equality is in general not satisfied, at the minimum $\mathbf{S}$ is different from $\mathbf{I}$, contrary to the case of extended orbitals.

The variational quality of the results obtained by minimizing $E[\mathbf{Q}]$, i.e., the difference $(\min_{\{\phi^L\}} E[\mathbf{Q}] - E_0)$, depends upon (i) the order $\mathcal{N}$ chosen for the definition of the $\mathbf{Q}$ matrix and (ii) the size of the LR. For $\mathbf{S} \leq 2\mathbf{I}$, it is easy to see that $E[\mathbf{Q}(\mathcal{N}-2)] \geq E[\mathbf{Q}(\mathcal{N})]$. Therefore by increasing $\mathcal{N}$ in the definition of $\mathbf{Q}$, one obtains an improvement of the total energy. This leads as well to an increase of the number of operations needed in the computation of $\mathbf{Q}$ [see Eq. (3)]. Most importantly, in order to improve the quality of the results one can choose to increase the size of the localization region. We note that the number of nonzero elements of $\mathbf{S}$ is proportional to $n_{\mathrm{LR}}N$, where n_{LR} is the average number of regions overlapping with a given one. Instead, the number of degrees of freedom needed to define the $N/2$ single-particle orbitals is proportional to mN, where m is the number of points belonging to a LR, e.g., the number of points where the wave function is nonzero. The ratio n_{LR}/m strongly depends on the basis set chosen to set up the Hamiltonian. The optimal choice of $\mathcal{N}$ and of size of the LR's, i.e., of the parameters determining the efficiency and accuracy of the computation, crucially depends upon the chosen basis set.

In calculations where $m \gg n_{\mathrm{LR}}$, the computer time for the $\mathbf{S}$ inversion amounts to a small fraction of the total time also for relatively large systems (e.g., systems with up to a few thousand electrons in LDA calculations with a PW basis). On the other hand, for computations with small basis sets, such as those with TB Hamiltonians, the computer time for the $\mathbf{S}$ inversion constitutes a considerable part of the total time already for small systems (i.e., containing a few tens of atoms).

V. MINIMIZATION OF TB HAMILTONIANS

The LO formulation was tested numerically using TB Hamiltonians[21,22] with the convention $\varepsilon_s + \varepsilon_p = 0$. We performed calculations for Si and C in different aggregation states. In calculations for crystalline structures, we considered nonzero hopping terms only between first neighbors. We chose a number of LR's equal to the number of atoms and we centered each LR at an atomic site (I). In a TB picture a LR can be identified with the set of atoms belonging to it. For each site I, we label the set of atoms which belong to a LR with LR_I. C and Si atoms have four valence electrons and there are two doubly occupied states for each atom in the system. We then associated two states to each LR: The two wave functions of the LR centered in I have nonzero components on the atoms belonging to the set LR_I and zero components (expansion coefficients) on the atoms which do not belong to LR_I. The expansion coefficients of the single-particle orbitals are treated as variational parameters in our calculations. The total number of expansion coefficients grows linearly with the size of the system.

We tested two different shapes of the LR. In one case an atom is defined as belonging to LR_I if its distance to the site I is less than or equal to a given radius r_c (in other words, a Euclidean metric is used to define the shape of the LR). In the second case, we took advantage of the form of the TB Hamiltonian and we considered an atom as belonging to LR_I if it is connected to the site I by a number of nonzero hopping terms less than or equal to a given number of shells N_h.

In all calculations $E[\mathbf{Q}]$ was minimized with respect to ϕ_L by a conjugate gradient (CG) procedure. The gradients $\frac{\partial E[\mathbf{Q}]}{\partial \phi_i^L}$ are simply obtained by projecting Eq. (12) onto the LR where ϕ_i^L is defined. For non-self-consistent Hamiltonians, the line minimization required in a CG procedure reduces to the minimization of a quartic polynomial in the variation of the wave function, along the conjugate direction. In our calculations the line minimization is performed exactly by evaluating the coefficients of the quartic polynomial.

We found that when localization constraints are imposed, $E[\mathbf{Q}]$ can have local minima and metastable states, where the system may be trapped for a long time during the minimization procedure, before reaching a minimum. This problem can be overcome if an appropriate choice of the initial guess for the iterative diagonalization is made. In all of our calculations we used starting wave functions with nonzero components only on the site I where they were centered; furthermore, orbital components were the same for each I. This choice

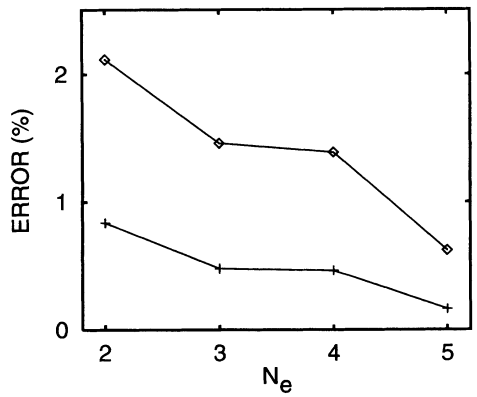

FIG. 3. Percentage error on the cohesive energy of Si (diamond structure) as a function of the number of shells (N_e) in the localization region, computed with a TB Hamiltonian (see text). Diamonds and crosses refer to minimizations of $E[\mathbf{Q}]$ with $\mathbf{Q}[\mathcal{N}=1]$ and $\mathbf{Q}[\mathcal{N}=3]$, respectively. The LR's were defined using a Euclidean metric (see text). The errors were evaluated with respect to a computation with extended orbitals. Calculations were performed at the same fixed volume.

allowed us to avoid local minima and metastable state traps for a wide class of ionic configurations. The problem of being trapped in metastable states or local minima involves only electronic minimizations; it does not concern MD simulations, where the ground state orbitals of a given step can be used as guess wave functions for the following step.

Figure 3 shows the percentage error on the cohesive energy E_c of Si in the diamond structure, as a function of the size of the LR, computed with respect to a calculation performed with extended orbitals. All computations were carried out with 216 atom supercells, simple cubic periodic boundary conditions, and the Γ point only for the supercell Brillouin zone (BZ) sampling. E_c was evaluated with $\mathbf{Q}[\mathcal{N}=1]$ and $\mathbf{Q}[\mathcal{N}=3]$ and with $\eta = 3$ eV. The shape of the LR was first chosen using a Euclidean metric. We denote with N_e the number of shells included in a LR, defined according to such a metric. It is seen that E_c converges rapidly as a function of N_e, with both $\mathcal{N}=1$ and 3. Already with $N_e = 2$ (17 atoms belong to a LR) the results are very good, i.e., E_c is higher than the result obtained with extended orbitals by only 2.1% and 0.8% for $\mathcal{N}=1$ and 3, respectively. For $\mathcal{N}=1$, the error on the total charge ΔN [see Eq. (13)] which gives the deviation from orthonormality due to localization constraints is in general very small; already for $N_e = 2$ we find it to be 0.2%. We note that when going from $N_e = 3$ (29 atoms in a LR) to $N_e = 4$ (35 atoms in a LR), we obtain the smallest variation of E_c. Indeed the atoms added to a LR when including also the fourth neighbor shell are not connected by hopping terms to those defining a LR when $N_e = 3$. This suggests that a definition of LR based on hopping terms is more physical than one based on the Euclidean metric. We repeated the calculations with $\mathcal{N}=1$ by choosing the LR's according to the hopping parameters and by setting the number of hopping shells N_h at 3. (For the diamond lattice, the definition of LR's using the two metrics is different for N_h and N_e larger than 2.) The choice $N_h = 3$ amounts to considering 41 atoms in a LR. The percentage error (0.7%) on E_c is very close to that obtained with $N_e = 5$ (0.6%), although the number of atoms in a given LR is bigger (47). The choice of the shape of the LR's according to the hopping parameters is superior to that of the Euclidean metric and it is especially so when energy differences between different structures are to be computed. This is the definition which was adopted in all subsequent calculations for C.

TABLE I. Cohesive energy E_c (eV) of different forms of solid carbon computed at a given lattice constant r_0 (Å) as a function of the number of shells (N_h) included in the LR. The calculations were performed with a TB Hamiltonian, with supercells containing 216, 128, and 100 atoms for diamond, two-dimensional graphite, and the linear chain, respectively.

Crystal structure	r_0	$\mathbf{E}_c(N_h=2)$	$\mathbf{E}_c(N_h=3)$	$\mathbf{E}_c$ $(N_h=\infty)$
Diamond	1.54	7.16	7.23	7.26
2D graphite	1.42	7.09	7.19	7.28
1D chain	1.25	5.62	5.75	5.93

TABLE II. Percentage errors on the equilibrium lattice parameters (δr_0), cohesive energy (δE_c) and bulk modulus (δB) of diamond, graphite, and a carbon linear chain, as obtained by minimizing E[Q] with $\mathbf{Q}[\mathcal{N}=1]$ and $N_h = 2$, described within a TB framework. The errors were evaluated with respect to a computation with extended orbitals.

Crystal structure	δr_0 (%)	δE_c (%)	δB (%)
Diamond	0.2	1.4	1.0
2D graphite	0.4	2.5	1.4
1D chain	0.5	4.7	2.7

Results for carbon in different crystal structures are presented in Tables I and II and in Fig. 4. We chose systems with different bonding and electronic properties: an sp^3 bonded insulator, diamond, an sp^2 bonded semimetal, planar graphite, and an sp bonded metal, a nondimerized C chain. Table I shows the binding energy of the three structures as a function of the size of the LR. The calculations were performed with $E[\mathbf{Q}(\mathcal{N}=1)]$. The errors for $N_h = 2$ and $N_h = 3$ are of the same order as those found in the case of silicon, and in particular we find that already for $N_h = 2$ the LO formulation and a direct diagonalization scheme are in good accord. In Fig. 4 we compare the total energy of the three C systems as a function of the lattice parameter, as obtained by direct diagonalization of the Hamiltonian and by minimizing $E[\mathbf{Q}(\mathcal{N}=1)]$ with respect to LO, with $N_h = 2$. The agreement between the two calculations is again very good for the three systems, in spite of their dif-

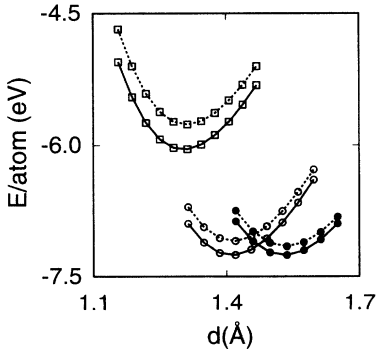

FIG. 4. Total energy (E) of diamond (dots), bidimensional graphite (open circles), and a carbon linear chain (squares) as a function of interatomic distance (d), computed with a TB Hamiltonian and supercells containing 216, 128, and 100 atoms, respectively. The dotted lines were obtained by minimizing at each volume $E[\mathbf{Q}]$ with $\mathbf{Q}[\mathcal{N}=1]$, and by using localization regions defined with $N_h = 2$. The solid lines were instead obtained by diagonalizing the Hamiltonian, with no constraints on the wave functions.

ferent bonding and electronic properties. The percentage differences between the computed equilibrium properties (lattice constant, cohesive energy, and bulk modulus) are given in Table II.

VI. MOLECULAR DYNAMICS WITH TB HAMILTONIANS

By using the functional $E[\mathbf{Q}]$ and localized orbitals one can set up a MD scheme in which the computational cost of each step scales linearly with the system size. According to the Helmann-Feynman theorem, one can obtain the forces acting on a given atom I by computing $\mathbf{F}_I = -\nabla_I E[\mathbf{Q};\{\phi_L\},\{\mathbf{R}_I\}]$; here $(\mathbf{R}_I)$ denotes ionic positions and $\{\phi_L\}$ are the localized orbitals which minimize $E[\mathbf{Q}]$. The general expression of the ionic forces is given by $\mathbf{F}_I = -2\sum_{i,j}^{N/2} Q_{ij}\langle\phi_i|\frac{\partial V}{\partial \mathbf{R}_I}|\phi_j\rangle$, where V indicates the external potential in a LDA calculation and the Hamiltonian in a TB calculation. In practical computations it is convenient to first calculate the auxiliary wave function

$$|\overline{\phi_i}\rangle = \sum_j^{N/2} Q_{ij}|\phi_j\rangle \tag{15}$$

and then to evaluate the expression of $\mathbf{F}_I$ as follows:

$$\mathbf{F}_I = -2\sum_i^{N/2}\langle\phi_i|\frac{\partial V}{\partial \mathbf{R}_I}|\overline{\phi_i}\rangle. \tag{16}$$

The ground state wave functions $\{\phi_L\}$ can be obtained either by evolving the electronic states according to Car-Parrinello[4] dynamics (see, e.g., Ref. 8), or by minimizing the energy functional $E[\mathbf{Q}]$ at each ionic move. In our simulations, we determine the sets LR_I at each ionic step; consequently the sites belonging to a set vary as a function of time, when, e.g., the atoms are diffusing or changing their local coordination. This implies an abrupt modification of the basis functions used for the expansion of $\{\phi_L\}$ and therefore a discontinuity of $\{\phi_L\}$ as a function of the ionic positions. In correspondence to any change of the sets LR_I, $E[\mathbf{Q}]$ must be minimized with respect to the electronic degrees of freedom; we therefore chose to minimize the energy functional at each ionic step, irrespective of whether the LR changes at a given step. The minimizations were performed with a conjugate gradient procedure where we used as initial guess for the orbitals the linear extrapolation of the minimized wave functions of the two previous ionic steps, as suggested in Ref. 3.

In order to test the accuracy and efficiency of the LO scheme for different classes of systems, we performed MD simulations for a crystalline insulator, i.e., diamond at low temperature, and for a liquid metal, i.e., liquid carbon at $T \simeq 5000$ K. As for the calculations for C presented in the previous section, we adopted LR's centered on atoms, which include up to the second shell of neighbors and whose shape is determined by the hopping parameters.

We first discuss the case of crystalline diamond, when the sets LR_I do not vary in time. We find that for diamond our MD scheme allows for a correct description of the total-energy oscillations, around equilibrium, consistently with what was previously obtained[8] for Si. We performed two simulations, one with a 64 atom and the other with a 1000 atom supercell. In both cases we started from an ionic configuration with zero velocities, generated by giving a random displacement to the atoms up to 0.03 Å with respect to their equilibrium positions. The integration time step (Δt) used in the simulations was 30 a.u. and the number of CG iterations per ionic move was 10. In Fig. 5 we show the potential energy (E) and the sum of the kinetic (E_{kin}) and potential energy of the system as a function of the simulation time. It is seen that the same energy drift $\Delta(E + E_{\mathrm{kin}})/E_{\mathrm{kin}}$ (0.1 in 0.5 ps) was found for the two simulations. This shows that the number of CG iterations to obtain a given accuracy in the energy conservation does not depend on the size of the system and that the overall scaling of the computational scheme is therefore linear. Finally we evaluated the relative error on the ionic forces $\mathbf{F}_I$ introduced by localization contraints as $\frac{\Delta F}{F} = \frac{\overline{\sum_I |\mathbf{F}_I^{\mathrm{loc}} - \mathbf{F}_I^{\mathrm{ext}}|}}{\sum_I |\mathbf{F}_I^{\mathrm{ext}}|}$, where the overbar indicates time averages, and the superscripts "loc" and "ext" refer to calculations performed with localized and extended states, respectively. This error was found to be $\simeq 6\%$ in crystalline diamond at room temperature.

We note that, if extended states are used, the number of iterations needed to have the same conservation of energy as the one reported in Fig. 5 is smaller than 10. Nevertheless our MD scheme applied to ordered systems becomes more efficient than direct diagonalization of the Hamiltonian already for small systems, i.e., for systems containing more than 40 atoms. This can be seen in Fig. 6 where we compare the efficiency of our approach to that of direct-diagonalization-based MD schemes.

We now analyze a MD simulation run during which the sets LR_I change as a function of time. In Fig. 7 we

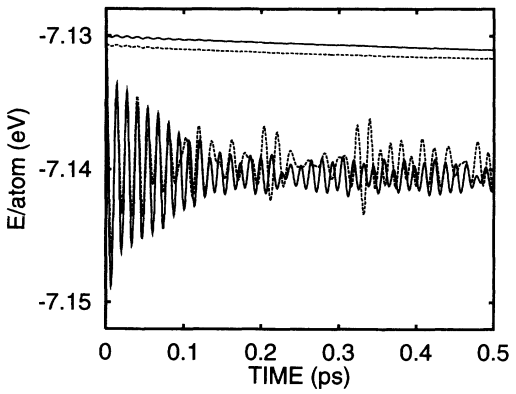

FIG. 5. Potential energy (lower part) and the sum of the potential and kinetic energy (upper part) as a function of simulation time for crystalline C in the diamond structure at 70 K. The dotted and solid lines refer to two calculations performed with a TB Hamiltonian, with 64 atom and 1000 atom supercells, respectively. In both cases we used $\mathbf{Q}[\mathcal{N} = 1]$ and $N_h = 2$; the LR's were computed for the configuration at 0 K and held fixed during the whole simulation.

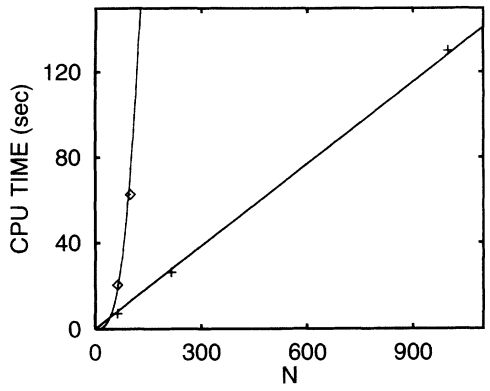

FIG. 6. CPU time per ionic step (30 a.u.) as a function of the number N of atoms in the system for a TB-MD simulation of C diamond at low temperature (see text). Squares and crosses indicate the CPU time in a direct-diagonalization-based scheme and in our MD approach (with 10 CG iterations per ionic move), respectively. Calculations were carried out on a Silicon Graphics Iris Indigo 4000.

show the potential energy for an oscillation of crystalline diamond around equilibrium, computed with extended ($E^{\rm ext}$, dotted line) and with localized ($E^{\rm loc}$, solid line) orbitals as a function of simulation time (t). The two energies have been computed for the same ionic trajectories, generated by a simulation with localized orbitals. The MD run shown in Fig. 7 is the same as the one reported in Fig. 5 but now the LR's are allowed to vary in time. At $t = t_1$, the evolution of the ionic positions makes the number of atoms belonging to given localization regions increase. At $t = t_2$, the ionic configuration is such as to restore the localization regions as they were at $t \leq t_1$. Since at $t = t_1, t_2$ an abrupt modification of the

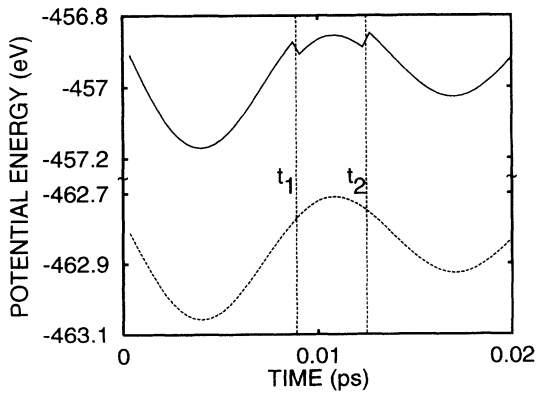

FIG. 7. Potential energy for an oscillation of crystalline diamond around equilibrium, computed with extended (dotted line) and with localized (solid line) orbitals as a function of simulation time. The two energy curves have been computed for the same ionic trajectories, generated by a simulation with localized orbitals. The LO calculation is the same as the one carried out in Fig. 5, but here the LR's are allowed to vary during the simulation. t_1 and t_2 denote times at which the LR's change.

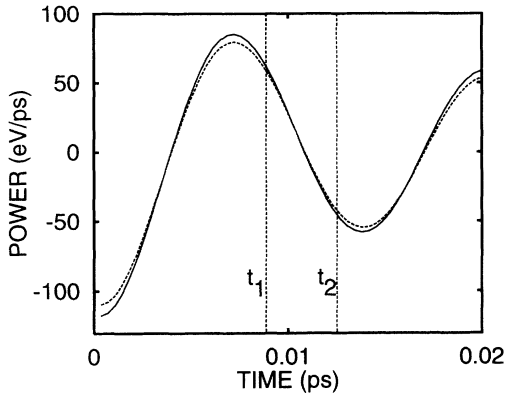

FIG. 8. Time derivatives $dE^{\rm ext}/dt = \sum_I \mathbf{F}_I^{\rm ext} \cdot \mathbf{v}_I$ (dotted line) and $dE^{\rm loc}/dt = \sum_I \mathbf{F}_I^{\rm loc} \cdot \mathbf{v}_I$ (solid line) of the potential energy curves reported in Fig. 7 (see text).

basis functions used for the expansion of $\{\phi_L\}$ occurs, the potential energy $E^{\rm loc}$ is discontinous and its derivative with respect to ionic positions is not well defined. However, ionic forces can still be defined by neglecting the discontinuity in E and by evaluating either the left or the right derivatives of the potential energy. The numerical values of the left and right derivatives are in fact the same within a very small error. This error is negligible, being much smaller than the one introduced by localization contraints. This can be seen in Fig. 8 where we compare forces obtained in calculations with extended and localized orbitals by plotting $dE^{\rm ext}/dt = \sum_I \mathbf{F}_I^{\rm ext} \cdot \mathbf{v}_I$ (dotted line) and $dE^{\rm loc}/dt = \sum_I \mathbf{F}_I^{\rm loc} \cdot \mathbf{v}_I$ (solid line). On the scale of the picture no discontinuity is observable in $dE^{\rm loc}/dt$ at $t = t_1, t_2$.

We now turn to the discussion of the simulation of liquid C, during which many changes of LR_I were observed. We generated a diffusive state at $T \simeq 5000$ K starting from a diamond network prepared at a macroscopic density of 2 gr cm^{-3}; we then heated the system by means of a Nosé-Hoover thermostat. We used a 64 atom cell with simple cubic periodic boundary conditions and only the Γ point to sample the BZ. We used a cutoff radius of 2.45 Å for the hopping parameters entering the TB Hamiltonian and for the two-body repulsive potential[22] (i.e., the cutoff distances r_m and d_m of Ref. 22 are set at 2.45 Å). Equilibration of the system was performed in the canonical ensemble and temporal averages were taken over 3.8 ps. The same simulation was repeated twice: once with our MD scheme and once by using direct diagonalization at each step. The radial distribution function $g(r)$ and the partial atomic coordinations obtained in the two cases are shown in Fig. 9 and Table III, respectively. The agreement between the two descriptions is excellent, showing that the LO scheme is accurate even for a difficult case such as a disordered system with differently coordinated atoms and metallic properties. The self-diffusion coefficients obtained in the two cases are 0.4×10^{-4} cm^2 s^{-1} and 0.6×10^{-4} cm^2 s^{-1}, respectively. The difference between the cohesive energies computed within the extended orbitals and the LO formulation for given ionic configurations is of the order of

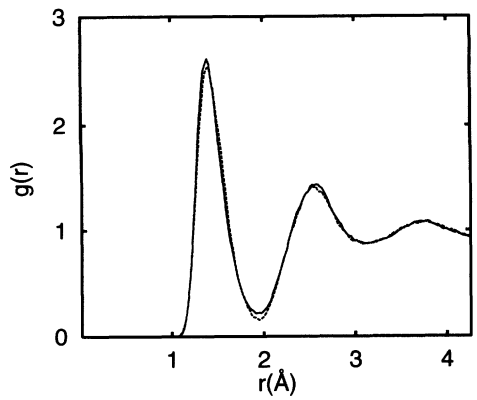

FIG. 9. Radial distribution function g(r) of liquid C (see text) computed as average over a TB-MD simulation of 3.8 ps. The results of the LO formulation with $N_h = 2$ and $\mathcal{N} = 1$ (dotted line) are compared to those of a direct diagonalization scheme (solid line). The average number of atoms in a LR is 18.

2%, similar to what we found for crystalline structures.

In the simulation for the liquid with LO's, we used $\Delta t = 5$ a.u. and we performed 50 iterations per ionic move, in order to minimize $E[\mathbf{Q}]$. This number is much larger than that needed for ordered systems, such as crystalline diamond. Consequently, in the case of liquid C our scheme becomes advantageous with respect to direct diagonalization when the number of atoms is larger than 200.

VII. CONCLUSIONS

We have presented an approach to total-energy minimizations and molecular-dynamics simulations whose computational workload is linear as a function of the system size. This favorable scaling is obtained by using an energy functional whose minimization does not imply either explicit orthogonalization of the electronic orbitals or inversion of an overlap matrix, together with a localized orbital formulation. The use of LO's reduces the evaluation of the energy functional and of its functional derivative to the calculation of products of sparse matrices.

TABLE III. Percentage number of differently coordinated sites (N_c) in liquid C computed as averages over a TB-MD simulation of 3.8 ps. The results of the LO formulation with $N_h = 2$ and $\mathcal{N} = 1$ are compared to those of a direct diagonalization scheme ($N_h = \infty$).

N_c	$N_h = 2$	$N_h = \infty$
one fold	5	4
two fold	38	42
three fold	53	50
four fold	4	4

The performances and efficiency of the method have been illustrated with several numerical examples for semiconducting and metallic systems. In particular, we have presented molecular-dynamics simulations for liquid carbon at 5000 K, showing that even for the case of a disordered metallic system the description provided by the LO formulation is reliable and very accurate. We have also shown that tight-binding molecular-dynamics simulations with 1000 atoms are easily feasible on small workstations, implying a one day run to obtain 0.5 ps. Molecular-dynamics simulations for very large C systems are under way.

ACKNOWLEDGMENTS

It is a pleasure to thank R. Car, R. M. Martin, and D. Vanderbilt for useful discussions and A. Possoz for her help in the optimization of the computer codes. We acknowledge support by the Swiss National Science Foundation under Grant No. 21-31144.91.

[1] For a review see, e.g., G. Galli and A. Pasquarello, in *Computer Simulation in Chemical Physics*, edited by M. P. Allen and D. J. Tildesley (Kluwer, Dordrecht, 1993), p. 261; M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1993).
[2] G. Galli and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992).
[3] T. A. Arias, M. C. Payne, and J. D. Joannopoulos, Phys. Rev. Lett. **69**, 1077 (1992).
[4] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[5] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).
[6] S. Baroni and P. Giannozzi, Europhys. Lett. **17**, 547 (1991).
[7] W.-L. Wang and M. Teter, Phys. Rev. B **46**, 12 798 (1992).
[8] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993).
[9] X.-P. Li, R. Numes, and D. Vanderbilt, Phys. Rev. B **47**, 10 891 (1993).
[10] M. S. Daw, Phys. Rev. B **47**, 10 895 (1993).
[11] A. D. Drabold and O. Sankey, Phys. Rev. Lett. **70**, 3631 (1993).
[12] W. Kohn, Chem. Phys. Lett. **208**, 167 (1993).
[13] P. Ordejón, D. Drabold, M. Grunbach, and R. Martin, Phys. Rev. B **48**, 14 646 (1993).
[14] M. Aoki, Phys. Rev. Lett. **71**, 3842 (1993).
[15] P. Löwdin, J. Chem. Phys. **18**, 365 (1950).
[16] F. Tassone, F. Mauri, and R. Car (unpublished).

[17] G. Pastore, E. Smargiassi, and F. Buda, Phys. Rev. A **44**, 6334 (1991).
[18] R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).
[19] G. Bachelet, D. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).
[20] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
[21] L. Goodwin, A. Skinner, and D. Pettifor, Europhys. Lett. **9**, 701 (1989).
[22] C. Xu, C. Wang, C. Chan, and K. Ho, J. Phys. Condens. Matter **4**, 6047 (1992).