# Order-$N$ projection method for first-principles computations of electronic quantities and Wannier functions

Uwe Stephan and David A. Drabold

*Department of Physics and Astronomy, Ohio University, Athens, Ohio 45701-2979*

(Received 10 March 1997; revised manuscript received 7 October 1997)

We present a generalized projection-based order-$N$ method which is applicable within nonorthogonal basis sets of spatially localized orbitals. The projection to the occupied subspace of a Hamiltonian, performed by means of a Chebyshev-polynomial representation of the density operator, allows the nonvariational computation of band-structure energies, density matrices, and forces for systems with nonvanishing gaps. Furthermore, the explicit application of the density operator to local basis functions gives a powerful method for the calculation of Wannier-like functions without using eigenstates. In this paper, we investigate such functions within models of diamond and fourfold-coordinated amorphous carbon starting from bonding pairs of hybrid orbitals. The resulting Wannier states are exponentially localized and show an ellipsoidal spatial dependence. These results are used to maximize the efficiency of a linear-scaling orthonormalization scheme for truncated Wannier functions. [S0163-1829(98)01611-7]

## I. INTRODUCTION

One of the most exciting developments in computational solid-state physics during this decade has been the creation of effective quantum-mechanical order-$N$ methods for the revelation of the electronic structure as well as the energetic relaxation of large model systems. With these techniques, both computational and memory efforts for computing band-structure energies, total energies, forces, and related quantities scale only *linearly* with the number $N$ of atoms in the system. As a consequence, this development has tremendously increased the applicability range of electronic-structure methods; in particular, *ab initio* procedures are now applicable to systems which a few years ago could only be investigated by means of empirical or semiempirical methods. Examples considered to date include giant single-shell fullerenes, multishell fullerenes, tubular systems, and large amorphous model structures[1-3] currently containing up to a few thousand atoms. On the other hand, the relative decrease in computational cost for medium-sized and large systems enables the reduction of time steps in molecular-dynamics (MD) simulations and therefore the investigation of short-time growth and relaxation processes.

The fundamental principle of all order-$N$ total-energy–force techniques is the utilization of the spatial locality or ''near-sightedness'' (Ref. 4) of electronic effects. One way of treating this phenomenon has been known for many years and is expressed by the possibility of performing unitary transformations of the occupied canonical eigenstates to obtain spatially localized wave functions. More precisely, these Wannier-like states are exponentially localized in insulators and decay algebraically in metals.[5,6] Based on this principle, methods have been devised which take only the local environment around a reference atom into account. As a further consequence, these techniques are particularly well suited to systems which have not too small gaps between occupied and unoccupied eigenstates.

Since the 70s and 80s, linear-scaling methods for the computation of local quantities such as local electronic and vibrational densities of states have widely been applied.

Above all, these are the recursion or Lanczos methods[7] and the maximum-entropy scheme.[8] These procedures already employ the sparse forms which the Hamiltonian and, if present, overlap matrices attain within a tight-binding (TB)-like description of the electronic states. To obtain a general $O(N)$ scheme for total densities of states, one has, in correspondence with the above principle, either to implement a local procedure independent of the system size[9] or to resort to the application of random vectors.[10-12] Though in principle possible, the computation of forces is much more expensive within these schemes and, when using random vectors, associated with the problem of slowly decreasing random noise.[13] These techniques have therefore not yet gained practical relevance for MD simulations.

Apart from the methods just described, there are, perhaps, three principal approaches to achieve an order-$N$ computational scheme for solving the electronic-structure problem. A common feature of all these methods is that they are not based on the computation of eigenvalues and eigenstates of the Hamiltonian; instead, physical quantities are expressed as traces of certain matrix expressions. These traces have to be computed either directly within the occupied subspace of the Hamiltonian spanned by a set of Wannier-like functions (WF's), or with use of the original basis orbitals and simultaneous inclusion of the density operator of the system.

A first group of methods uses a differential[14] or variational[15-17] solution for the density matrix of the system. Because the density operator can be constructed by means of localized, orthonormal WF's, the density matrix with respect to a spatially localized basis set is localized as well. Using some additional effort to ensure the idempotency of this matrix (for zero temperature), one obtains a linear-scaling procedure by performing traces within the full Hamiltonian space.

A second group of procedures is based on the variational computation of the WF's themselves. Representing these functions by local-orbital[18-22] or plane-wave expansions[23] and enforcing them to be localized in space, the final WF's minimize an appropriate energy functional. With use of this

Wannier basis, all traces are automatically performed within the occupied subspace.

Finally, a third, nonvariational scheme is based on a direct, polynomial representation of the density operator $\hat{\rho}$. This operator can be used with finite temperatures, but at zero temperature $\hat{\rho}$ is a projection operator which projects out the occupied fraction of any wave function it is applied to. A polynomial expansion of $\hat{\rho}$ was employed in Ref. 24 to derive a Lanczos method within the occupied subspace, whereas Goedecker and co-workers[25] used it to develop a projection method for the computation of total energies and forces within orthonormal, local basis sets. Similar to the first group of methods described above, traces are performed within the original basis of local orbitals. However, the number of polynomial terms needed to represent $\hat{\rho}$ at zero temperature increases with decreasing gap width between occupied and unoccupied eigenstates. Furthermore, this approach requires precomputed estimates for the Fermi energy and the band edges of the systems. These estimates, however, can be extracted in an effective order-$N$ way from the aforementioned recursion or maximum-entropy schemes.

One advantage of the third method is that no initial-guess functions and no initial density matrix are needed when computing electronic properties as traces. If such initial guesses are far from the final solutions, they may seriously restrict the effectiveness of variational schemes during the first MD cycles. The projection method is therefore a powerful technique particularly for systems with high degrees of disorder. Special initial functions are only necessary when the method is used for explicit computations of Wannier functions, because then one has to ensure to obtain the correct number of linearly independent projected functions which can be orthonormalized. But even in this case we could show by first investigations[26] that the direct projection to the occupied subspace is superior to the usual variational techniques at the beginning of a MD run. If this turns out to be generally true, the method will be capable of providing quite ideal initial-guess density matrices *and* WF's for variational MD schemes, and this should apply even for systems with small or vanishing gaps. This connection would allow very effective MD simulations for wide classes of materials. As another important application, the knowledge of the WF's opens a very attractive way for real-space computations of the electric polarization and related dielectric constants.[6,27,28]

So far, the projection method has essentially been developed and applied within orthogonal TB-like Hamiltonians only. One purpose of the present paper is to demonstrate a very effective generalization of the method to nonorthogonal basis sets using an approximation of the ''upper-lower indexed'' Hamiltonian matrix[29] as proposed in Ref. 30. With this generalization, the method can be used within local-basis *ab initio* Hamiltonians as well. Furthermore, we will examine the use of this method for computations of Wannier-like functions in crystalline and amorphous systems. In this paper, we will restrict these investigations to completely fourfold-coordinated systems. In this case, physically reasonable initial functions for the performance of the projection are given by the ($\sigma$-like) bonding combinations of adjacent hybrid orbitals. In a forthcoming second paper, we will then present a general method for the computation of hybrid-

related WF's in undercoordinated systems (with coordination numbers $\leqslant 4$).

The paper is organized as follows. In Sec. II we present a short review of the projection method for orthogonal basis systems as recently published.[24,25] Section III is devoted to our generalization of this method to nonorthogonal local-orbital systems. We will give the basic expressions for a direct application of the method within MD simulations. This section also contains a discussion of the main sources which can lead to errors in the projection method. To illustrate our results, we will present test calculations using a simple first-principles non-self-consistent one-particle Hamiltonian[31] which is based on the local-density approximation (LDA). In Sec. IV we then turn to the computation of WF's and show results for models of crystalline and amorphous carbon. In particular, we investigate the radial dependence of these functions obtained with and without localization constraints and explore the degree of orthogonality which truncated projected functions can achieve within a linear-scaling orthonormalization. Finally, we conclude in Sec. V.

## II. PROJECTION METHOD FOR ORTHONORMAL BASIS STATES

The projection method for the nonvariational computation of electronic properties of model systems is based on a polynomial representation of the density operator

$$\hat{\rho} = 2\hat{F} = 2F(\hat{H}), \tag{1}$$

where $F(E) = [e^{\beta(E-\mu)} + 1]^{-1}$ is the Fermi distribution function at some temperature $T$ ($\beta = 1/kT$, and $\mu$ is the chemical potential), and $\hat{H}$ is the Hamiltonian operator of the system.[24,25] The factor 2 accounts for spin degeneracy. In this paper, we use a direct Chebyshev approximation of $F(E)$ to derive a numerically stable representation of $\hat{\rho}$. At zero temperature, $F(E)$ is a step function, but can be approximated for systems with an energy gap between occupied and unoccupied eigenstates by using a sufficiently large but finite value of $\beta$. In this case, the Fermi operator $\hat{F}$ is an idempotent projection operator having eigenvalues 1 and 0 only. An alternative approach has been proposed by Silver *et al.*[32] These authors employ the kernel polynomial method to find the smoothest Chebyshev approximation of the Heaviside step function when using a finite number of expansion terms (moments). This may prove to be a valuable scheme especially in the case of small energy gaps.

For a set of orthonormal electronic basis states $\{|\varphi_\alpha\rangle\}$, Eq. (1) can immediately be rewritten in terms of the usual Hamiltonian matrix $\mathbf{H}$ with elements $H_{\alpha\beta} = \langle\varphi_\alpha|\hat{H}|\varphi_\beta\rangle$,

$$\boldsymbol{\rho} = 2F(\mathbf{H}),$$

where $\boldsymbol{\rho}$ is the density matrix with elements

$$\rho_{\alpha\beta} = \langle\varphi_\alpha|\hat{\rho}|\varphi_\beta\rangle.$$

.

The band-structure energy $E_{\text{bs}}$, for instance, can then be computed as the trace

$$E_{\mathrm{bs}} = \mathrm{Tr}[\hat{\rho}\hat{H}] = \sum_{\alpha} \langle \varphi_{\alpha} | \hat{\rho}\hat{H} | \varphi_{\alpha} \rangle = 2\,\mathrm{Tr}\,[F(\mathbf{H})\mathbf{H}]. \quad (2)$$

The force on atom $j$ which originates from the band-structure energy is given by the derivative

$$\mathbf{F}_j = -\frac{\partial}{\partial \mathbf{R}_j} \mathrm{Tr}[\hat{\rho}\hat{H}].$$

For insulators at zero temperature, this formula can be simplified to[25]

$$\mathbf{F}_j = -\sum_{\alpha\beta} \rho_{\alpha\beta} \frac{\partial H_{\beta\alpha}}{\partial \mathbf{R}_j} = -2\,\mathrm{Tr}\left[F(\mathbf{H})\frac{\partial}{\partial \mathbf{R}_j}\mathbf{H}\right], \quad (3)$$

including the Pulay corrections with respect to the Hamiltonian operator $\hat{H}$. In Sec. III we will derive these equations and similar expressions for other electronic properties for the general case of nonorthogonal basis states.

A linearly scaling behavior results from the fact that $\mathbf{H}$ becomes a sparse matrix when using spatially localized TB-like functions $|\varphi_{\alpha}\rangle$. In the same representation, the density matrix $\boldsymbol{\rho}$ for insulators can likewise be approximated by a sparse matrix setting all matrix elements to zero beyond some critical localization (LOC) radius. This is possible because orthonormal WF's which diagonalize $\hat{\rho}$ decay exponentially in insulators.[5,6] Thus, e.g., in Eq. (2) the computations $\hat{\rho}\hat{H}|\varphi_{\alpha}\rangle$ are independent of the system size, and the computational effort scales linearly with the number of atoms in the system.

Another generalization of these expressions is related to systems with periodic boundary conditions where the localized basis states $|\varphi_{\alpha}\rangle$ are used to construct Bloch functions $|\chi_{\alpha k}\rangle$. The spatially localized WF's are then expanded in terms of these Bloch states. As a consequence, the matrices $\mathbf{H}$ and $\boldsymbol{\rho}$ with respect to these Bloch functions are now $\boldsymbol{k}$ dependent but nevertheless sparse. The complete method for this case can therefore be derived in a straightforward manner. However, because linearly scaling procedures are mainly designed for applications to large systems where a $\Gamma$-point approximation is often sufficiently accurate, it is customary to restrict these schemes to $\boldsymbol{k}=0$.

## III. PROJECTION METHOD FOR NONORTHOGONAL BASIS STATES

### A. Method

One way of generalizing the projection method to nonorthogonal basis orbitals was outlined by Goedecker,[33] and consists of representing the density operator by a contour integration in the complex energy plane over the Green's operator. Within this approach, one has to invert the matrix $\langle \varphi_{\alpha} | E - \hat{H} | \varphi_{\beta} \rangle$ for every energy value $E$ used in the integration. In this paper, we will use a different scheme which is based on an approximation for the upper-lower indexed (or "contracovariant") Hamiltonian matrix

$$H^{\alpha}{}_{\beta} = \sum_{\gamma} (S^{-1})^{\alpha\gamma} H_{\gamma\beta}. \quad (4)$$

Here $(S^{-1})^{\alpha\beta}$ is the inverse of the overlap matrix $S_{\alpha\beta} = \langle \varphi_{\alpha} | \varphi_{\beta} \rangle$. In Eq. (4) we use the very convenient tensorlike notation proposed by Ballentine and Kolář.[29] In this notation lower indices are related to the original basis functions $|\varphi_{\alpha}\rangle$ whereas upper indices refer to the so-called dual basis $\sum_{\beta}(S^{-1})^{\alpha\beta}|\varphi_{\beta}\rangle$. This convention greatly simplifies the correct formulation of all equations in which the matrix elements of arbitrary quantities with respect to nonorthogonal basis states occur. However, to rewrite equations such as Eq. (4) in a compact matrix form, we also introduce the shorthand notation

$$\bar{\mathbf{H}} = \mathbf{S}^{-1}\mathbf{H},$$

where the bar over the matrix symbol indicates the raise of the first index. Hence $\bar{\mathbf{H}}$ and $\mathbf{H}$ are the matrices with elements $H^{\alpha}{}_{\beta}$ and $H_{\alpha\beta}$, respectively.

The importance of the matrix $\bar{\mathbf{H}}$ has been known for many years.[7,34] It is the proper matrix representant of the operator $\hat{H}$ in the sense that the action of $\hat{H}$ on a basis function $|\varphi_{\beta}\rangle$ is represented by $\bar{\mathbf{H}}$:

$$\hat{H}|\varphi_{\beta}\rangle = \sum_{\alpha} |\varphi_{\alpha}\rangle H^{\alpha}{}_{\beta}. \quad (5)$$

This implies that $\hat{H}$ and $\bar{\mathbf{H}}$ have the same eigenvalues. One has to note, however, that the matrix $\bar{\mathbf{H}}$ in general is not Hermitian, as $\hat{H}$ and the matrices $\mathbf{H}$ and $\mathbf{S}$ are.

The important point here as shown in Ref. 30 is that even though $\bar{\mathbf{H}}$ would extend beyond any tight-binding-like cutoff in $\mathbf{H}$, the matrix elements $H^{\alpha}{}_{\beta}$ decay faster with increasing distance between the centers of the localized basis functions $|\varphi_{\alpha}\rangle$ and $|\varphi_{\beta}\rangle$ than the corresponding elements $H_{\alpha\beta}$. It is therefore justified to approximate $\bar{\mathbf{H}}$ in a tight-binding-like picture by a similar sparse matrix form as used for $\mathbf{H}$, i.e., to set all elements of $\bar{\mathbf{H}}$ to zero beyond some finite cutoff radius $R_{\bar{\mathbf{H}}}$. The matrix elements $H^{\alpha}{}_{\beta}$ can now be obtained by solving $NL$ ($N$ is the number of atoms and $L$ is the mean number of orbitals per atom) linear systems of equations[30]

$$\sum_{\gamma} S_{\alpha\gamma} H^{\gamma}{}_{\beta} = H_{\alpha\beta}, \quad (6)$$

where $\beta$ runs through all $NL$ basis orbitals in the system while $\alpha$ and $\gamma$ in Eq. (6) only index the orbitals within the environment around $|\varphi_{\beta}\rangle$ defined by $R_{\bar{\mathbf{H}}}$. Note that due to the restricted extent of the submatrices in Eq. (6) and the positive definiteness of the overlap matrices, these equations can, in general, easily be computed by a Cholesky procedure [cf. Sec. III B 3].

Having computed the matrix $\bar{\mathbf{H}}$, it is now easy to evaluate certain operator equations in matrix form. For instance, due to Eq. (5) any power of $\hat{H}$ is given by

$$\hat{H}^k|\varphi_{\beta}\rangle = \sum_{\alpha} |\varphi_{\alpha}\rangle (H^k)^{\alpha}{}_{\beta}, \quad (7)$$

where $(H^k)^{\alpha}{}_{\beta}$ are the matrix elements of $\bar{\mathbf{H}}^k$. Therefore, the band-structure energy now reads

$$E_{bs} = \text{Tr}[\hat{\rho}\hat{H}] = \sum_{\alpha\beta} (S^{-1})^{\alpha\beta}\langle\varphi_\beta|\hat{\rho}\hat{H}|\varphi_\alpha\rangle = 2\ \text{Tr}[F(\bar{\mathbf{H}})\bar{\mathbf{H}}].$$

(8)

Moreover, the moments of the total density of states (DOS) $\rho(E)$ which can be used to obtain the DOS by means of, e.g., maximum-entropy schemes, are expressed as

$$\mu^{(k)} = \int_{-\infty}^{\infty} E^k\rho(E)dE = \frac{1}{N}\text{Tr}[\hat{H}^k] = \frac{1}{N}\text{Tr}[\bar{\mathbf{H}}^k].$$

(9)

The normalization here is chosen in such a way that the integral $\int_{-\infty}^{\infty}\rho(E)dE$ equals the mean number of basis orbitals per atom.

Analogously to $\bar{\mathbf{H}}$, one can define a ''contra-covariant'' density matrix $\bar{\boldsymbol{\rho}} = 2F(\bar{\mathbf{H}})$ with matrix elements

$$\rho^\alpha{}_\beta = 2F(H)^\alpha{}_\beta = \sum_\gamma (S^{-1})^{\alpha\gamma}\rho_{\gamma\beta},$$

(10)

and $\rho_{\alpha\beta} = \langle\varphi_\alpha|\hat{\rho}|\varphi_\beta\rangle$ as in Sec. II. The trace of this matrix gives the total number of electrons:

$$N_{el} = \text{Tr}[\hat{\rho}] = \text{Tr}[\bar{\boldsymbol{\rho}}] = \sum_\alpha \rho^\alpha{}_\alpha.$$

Of course, this expression can also be used to define the Fermi energy in the system.

It is interesting to note that the density matrix used by Nunes and Vanderbilt[16] is expressed in the present picture as an upper-indexed matrix $\rho^{\alpha\beta}$ with

$$\rho^\alpha{}_\beta = \sum_\gamma \rho^{\alpha\gamma}S_{\gamma\beta}.$$

(11)

If the eigenfunctions of the system are given by $|\psi_i\rangle = \sum_\alpha c_i^\alpha|\varphi_\alpha\rangle$, this matrix is simply related to the coefficients $c_i^\alpha$ by

$$\rho^{\alpha\beta} = \sum_i n_i c_i^\alpha c_i^{\beta*}$$

$$= \sum_i n_i \sum_{\gamma\delta} (S^{-1})^{\alpha\gamma}\langle\varphi_\gamma|\psi_i\rangle\langle\psi_i|\varphi_\delta\rangle(S^{-1})^{\delta\beta},$$

(12)

where $n_i$ denote the occupation numbers of the eigenstates $|\psi_i\rangle$. As can be seen from Eq. (11), the matrix $\rho^{\alpha\beta}$ can be obtained from the matrix $\rho^\alpha{}_\beta$ by solving linear systems of equations similar to those as used in Eq. (6). We remark that it may be convenient to introduce a compact notation for such upper-indexed matrices as well. In analogy to using $\bar{\boldsymbol{\rho}}$ for the matrix $\rho^\alpha{}_\beta$, we will also adopt the notation $\bar{\bar{\boldsymbol{\rho}}}$ for $\rho^{\alpha\beta}$, thus indicating the raise of both indices.

As we will see, the density matrix $\bar{\bar{\boldsymbol{\rho}}}$ is the important quantity when calculating electronic charge densities and forces. The charge density is given in the usual way:

$$\rho(\mathbf{r}) = \langle\mathbf{r}|\hat{\rho}|\mathbf{r}\rangle = \sum_{\alpha\beta} \langle\mathbf{r}|\varphi_\alpha\rangle\rho^{\alpha\beta}\langle\varphi_\beta|\mathbf{r}\rangle.$$

We now generalize Eq. (3) for the force on atom $j$ which originates from the band-structure energy term. Again, we simplify the equations by restricting ourselves to the case of insulators at zero temperature. In this case the occupation numbers $n_i$ can be considered constant. Let us first recall the corresponding equation for this force in terms of eigenstate coefficients $c_i^\alpha$ and eigenvalues $E_i$ with respect to some spatially localized basis set (such that $\sum_\beta H_{\alpha\beta}c_i^\beta = E_i\sum_\beta S_{\alpha\beta}c_i^\beta$):[35,36]

$$\mathbf{F}_j = -\frac{\partial E_{bs}}{\partial\mathbf{R}_j} = -\sum_i n_i\frac{\partial}{\partial\mathbf{R}_j}\sum_{\alpha\beta} c_i^\alpha c_i^{\beta*}\langle\varphi_\beta|\hat{H}|\varphi_\alpha\rangle$$

$$= \sum_i n_i\sum_{\alpha\beta} c_i^\alpha c_i^{\beta*}\left(-\frac{\partial H_{\beta\alpha}}{\partial\mathbf{R}_j} + E_i\frac{\partial S_{\beta\alpha}}{\partial\mathbf{R}_j}\right).$$

(13)

This expression can easily be verified by taking the secular equation and $\sum_{\alpha\beta} c_i^\alpha c_i^{\beta*}S_{\beta\alpha} = 1$ into account. Note that Eq. (13) is *exact* within the (e.g., TB-like) approximation used for the eigenstates $|\psi_i\rangle$, i.e., it contains all Pulay corrections with respect to the ''exact'' Hamiltonian operator $\hat{H}$. To rewrite the second term in Eq. (13), we need an equation similar to Eq. (12):

$$\sum_i n_i c_i^\alpha c_i^{\beta*}E_i$$

$$= \sum_i n_i\sum_{\gamma\delta} (S^{-1})^{\alpha\gamma}\langle\varphi_\gamma|\psi_i\rangle\langle\psi_i|\hat{H}|\psi_i\rangle\langle\psi_i|\varphi_\delta\rangle(S^{-1})^{\delta\beta}$$

$$= (\hat{H}\hat{\rho})^{\alpha\beta}.$$

(14)

Inserting Eqs. (12) and (14) into Eq. (13), we obtain the desired equation

$$\mathbf{F}_j = \sum_{\alpha\beta}\left(-\rho^{\alpha\beta}\frac{\partial H_{\beta\alpha}}{\partial\mathbf{R}_j} + \sum_\gamma H^\alpha{}_\gamma\rho^{\gamma\beta}\frac{\partial S_{\beta\alpha}}{\partial\mathbf{R}_j}\right).$$

(15)

This equation contains the matrices $\bar{\mathbf{H}}$ and $\bar{\bar{\boldsymbol{\rho}}}$ from Eqs. (6) and (11) as well as the derivatives of the usual Hamiltonian and overlap matrices $\mathbf{H}$ and $\mathbf{S}$ with respect to the atomic coordinates. We now show that Eq. (15) leads directly to an expression for the forces derived in a different way by Ordejón *et al.*[20] To this end, let us introduce a set of *nonorthonormal* but linearly independent Wannier-like functions $\{|\phi_\mu\rangle\}$ which span the occupied subspace of $\hat{H}$ (at zero temperature). The density operator can now be written as $\hat{\rho} = \sum_{\mu\nu}|\phi_\mu\rangle\rho^{\mu\nu}\langle\phi_\nu|$ with $\rho^{\mu\nu} = 2(S^{-1})^{\mu\nu}$ and $S_{\mu\nu} = \langle\phi_\mu|\phi_\nu\rangle$. If we use the expansion $|\phi_\mu\rangle = \sum_\alpha c_\mu^\alpha|\varphi_\alpha\rangle$, we find, in this representation,

$$\rho^{\alpha\beta} = \sum_{\gamma\delta} (S^{-1})^{\alpha\gamma}\langle\varphi_\gamma|\hat{\rho}|\varphi_\delta\rangle(S^{-1})^{\delta\beta}$$

$$= 2\sum_{\mu\nu} c_\mu^\alpha(S^{-1})^{\mu\nu}c_\nu^{\beta*},$$

$$(\hat{H}\hat{\rho})^{\alpha\beta} = \sum_{\gamma\delta} (S^{-1})^{\alpha\gamma}\langle\varphi_\gamma|\hat{H}\hat{\rho}|\varphi_\delta\rangle(S^{-1})^{\delta\beta}$$
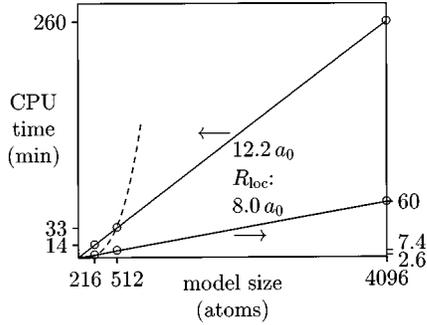
FIG. 1. Linear CPU time scaling (Ref. 38) obtained with the projection method for computing band-structure energies using a non-self-consistent two-center LDA Hamiltonian (Ref. 31) and applied to models for diamond (216 atoms) and tetrahedrally coordinated amorphous carbon (Ref. 39) (512 and 4096 atoms). The results are given for two different localization radii $R_{loc}$ as indicated (solid lines). For comparison, the $O(N^3)$ scaling of the CPU time needed for direct diagonalizations of the corresponding secular equations is also presented (dashed line).

$$= 2 \sum_{\kappa \lambda \mu \nu} c_\kappa^\alpha (S^{-1})^{\kappa \mu} H_{\mu \nu} (S^{-1})^{\nu \lambda} c_\lambda^{\beta *} ,$$

with $H_{\mu \nu} = \langle \phi_\mu | \hat{H} | \phi_\nu \rangle$. Inserting these formulas into Eq. (15), we obtain an expression derived and already applied within MD simulations by Ordejón *et al.*[20] [Eq. (37) in their paper].

A detailed investigation of the forces obtained with Eq. (15) will be presented elsewhere. In particular, we will investigate the numerical properties of this expression when computed within the projection method. Furthermore, it may be interesting to examine the results of Eq. (15) in the presence of localization constraints for the density matrix. The influence of this spatial confinement on the forces should be related to a recent investigation by Voter, Kress, and Silver[37] using a truncated-moment approach.

As in Sec. II, to achieve a linear-scaling behavior of the entire method, we have to utilize the sparseness of the density matrix. This means that the matrix-vector multiplications for performing the projections in Eqs. (8) and (10) can be accomplished within certain localization regions which can be defined by some radius $R_{loc}$. To minimize the error in these computations, $R_{loc}$ should be, in general, somewhat larger than the radius $R_{\bar{\mathbf{H}}}$ used above for defining the sparse matrix $\bar{\mathbf{H}}$. We will present a detailed discussion of these errors in Sec. III B. Here note that this localization argument cannot be applied to Eq. (9), which does not contain the projection operator. The linear-scaling evaluation of the traces for computing moments of the total DOS is therefore usually done by means of global random vectors.[10–12]

To illustrate now the $O(N)$ scaling of the method, we present in Fig. 1 the CPU time[38] needed for computations of the band-structure energy in three models for diamond and fourfold-coordinated amorphous carbon containing up to 4096 atoms.[39] For these calculations, we employed a simple first-principle non-self-consistent one-particle Hamiltonian[31] in which the localized wave functions and atomic potentials are taken from modified self-consistent atomic LDA calculations. The matrix elements in extended systems are obtained

by using a minimal set of valence orbitals, but the core corrections are omitted together with the three-center integrals and part of the two-center integrals. To obtain comparable results, in all calculations we used the same values for the cutoff radii of the Hamiltonian [6.1 Bohr radii ($a_0$) for $\mathbf{H}$ and $\bar{\mathbf{H}}$ corresponding to three nearest-neighbor spheres in diamond], the parameter $\beta$ ($=50$ with respect to the scaled Fermi distribution on $[-1,1]$, this corresponds to $\beta \approx 50$ hartree$^{-1}$) and the number of Chebyshev polynomials (80) for the expansion of $F(E)$. The lower curve was obtained with a relatively small LOC radius for performing the projection (8.0 $a_0$); the upper curve was computed to achieve more accurate band-structure energies as compared to results from direct diagonalization (computed for the two smaller models). For these calculations we set the LOC radius to twice the cutoff radius of $\hat{H}$.

The relative errors in $E_{bs}$ for the diamond model and the amorphous model with 512 atoms ($a$-C512) are $1.1 \times 10^{-4}$ and $7.2 \times 10^{-4}$, respectively, for the lower curve and decreased to $1.5 \times 10^{-5}$ and $2.2 \times 10^{-4}$ in the case of the upper curve. Note that the amorphous model has a relatively small HOMO-LUMO gap (between highest occupied and lowest unoccupied molecular orbitals) of 0.9 eV (for the Hamiltonian used) due to some defect states in the band gap. For comparison, the CPU time necessary for computing all eigenvalues has been included in Fig. 1. Of course, the crossing point of the $O(N)$ and $O(N^3)$ curves moves to smaller models if one can afford to use a somewhat lower accuracy in the projection method. In Sec. III B we will discuss in more detail how this accuracy depends on the parameters inherent in the present method.

### B. Error discussion

We now discuss the main sources which can lead to errors in the projection method. Again, we will consider these errors with respect to the band-structure energy. It seems to be clear that inaccuracies in $E_{bs}$ will also affect more subtle quantities like forces and the charge density. For the following discussion we will presume to have a given Hamiltonian with a certain TB-like cutoff radius $R_{\mathbf{H}}$ in $\mathbf{H}$. Furthermore, we consider systems with a finite HOMO-LUMO gap.

### 1. Use of a finite value for $\beta$

Although the method described can be regarded as a finite-temperature scheme, the temperature used for getting a Fermi distribution which can be represented by a tractable number of Chebyshev polynomials is often much higher than the physical temperature one is interested in. Furthermore, if one wants to use the projection property of $\hat{\rho}$, e.g., for computing WF's, one has to approximate the Fermi distribution at $T=0$. We therefore estimate the deviation of $E_{bs}$ from its value at $T=0$ when using a finite value of $\beta$ and setting $\mu$ equal to the Fermi energy $E_F$. This error, say $\epsilon_1$, is given by the sum

$$\epsilon_1 = 2 \left[ \sum_{E_i > E_F} \frac{E_i}{e^{\beta(E_i - E_F)} + 1} - \sum_{E_i < E_F} \frac{E_i}{e^{\beta(E_F - E_i)} + 1} \right] ,$$

(16)

where $E_i$ are the eigenvalues of $\hat{H}$. For our $\beta$ values (usually $\beta = 50, \ldots, 100$ for $E$ on $[-1,1]$) the exponents in the denominators rapidly attain values $\beta|E_i - E_F| \gg 1$. This means that the weights of the eigenvalues in sum (16) decay exponentially when going away from the Fermi energy in either direction. The order of magnitude of $\epsilon_1$ can therefore be estimated by taking only the HOMO and LUMO states in Eq. (16) into account, thus yielding

$$\epsilon_1 \approx 2\{F(E_{LUMO})E_{LUMO} + [F(E_{HOMO}) - 1]E_{HOMO}\}$$
$$= 2F(E_{LUMO})\Delta E \qquad (17)$$

with $E_{LUMO} = E_F + \frac{1}{2}\Delta E$ and $E_{HOMO} = E_F - \frac{1}{2}\Delta E$. To consider as an example the $a$-C512 model presented in Fig. 1 and the $\beta$ value used there, this gives a relative error $\epsilon_1/|E_{bs}| \approx 1.6 \times 10^{-5}$. The exact error as computed from all eigenvalues is $5.2 \times 10^{-5}$. For the calculation using the smaller LOC sphere in Fig. 1, this error is more than one order of magnitude smaller than the total deviation in $E_{bs}$ given above. This means that even for $\Delta E \approx 0.9$ eV ($\approx 0.03$ hartree) the value $\beta = 50$ hartree$^{-1}$ is sufficient. For the larger LOC radius, $\epsilon_1$ may still not be the decisive error. However, if more accurate computations are desired, the value of $\beta$ should also be increased in this case.

### 2. Expanding the Fermi distribution into a finite sum of Chebyshev polynomials

The error related to this expansion, say $\epsilon_2$, can be estimated by replacing the density of states with its mean value $\bar{\rho}$ between some band edges $E_{min}$ and $E_{max}$ and performing explicitly the integrations in the following expression:

$$\epsilon_2 \approx 2N\bar{\rho}\left[ \int_{E_{min}}^{E_{max}} F_\beta^{Ch}(E)E \; dE - \int_{E_{min}}^{E_{max}} F_\beta(E)E \; dE \right]. \qquad (18)$$

Here, $F_\beta^{Ch}(E)$ is the Chebyshev approximation of the Fermi distribution $F_\beta(E)$ for some given value $\beta$. The second term within the square brackets in Eq. (18) can be evaluated in a similar way as used for the degenerate electron gas in statistical physics:

$$\int_{E_{min}}^{E_{max}} \frac{E \; dE}{e^{\beta(E - E_F)} + 1}$$

$$= \int_{E_{min}}^{E_F} E \; dE - \int_{E_{min}}^{E_F} \frac{E \; dE}{e^{\beta(E_F - E)} + 1}$$

$$+ \int_{E_F}^{E_{max}} \frac{E \; dE}{e^{\beta(E - E_F)} + 1}$$

$$= \frac{1}{2}(E_F^2 - E_{min}^2) + \int_0^{E_F - E_{min}} \frac{(y - E_F)dy}{e^{\beta y} + 1}$$

$$+ \int_0^{E_{max} - E_F} \frac{(y + E_F)dy}{e^{\beta y} + 1}. \qquad (19)$$

At the band edges $y = E_F - E_{min}$ and $y = E_{max} - E_F$, we obviously have $e^{\beta y} \gg 1$, one can therefore replace the upper integration boundaries in Eq. (19) with $\infty$. This results in

$$\int_{E_{min}}^{E_{max}} F_\beta(E)E \; dE \approx \frac{1}{2}(E_F^2 - E_{min}^2) + \frac{2}{\beta^2}\int_0^\infty \frac{x \; dx}{e^x + 1}$$

$$= \frac{1}{2}(E_F^2 - E_{min}^2) + \frac{\pi^2}{6\beta^2}. \qquad (20)$$

The first term here corresponds to the Fermi distribution at $T = 0$, hence the second term again describes the deviation which was discussed in Sec. III B 1.[44] Turning now to the first term in Eq. (18), the integration can easily be performed with use of the Chebyshev coefficients of $F_\beta^{Ch}$. This can be seen by employing the formulas

$$\int T_n(x)dx = \frac{1}{2}\left[ \frac{T_{n+1}}{n+1} - \frac{T_{n-1}}{n-1} \right] \quad (n \geq 2),$$

$$\int xT_n(x)dx = \frac{1}{4}\left[ \frac{T_{n+2}}{n+2} - \frac{T_{n-2}}{n-2} \right] \quad (n \geq 3),$$

which can be derived by differentiating[45] the defining equations for the Chebyshev polynomials $T_n(x) = \cos(n \arccos x)$. By means of the expansion

$$F_\beta^{Ch}\left( \frac{E_{max} - E_{min}}{2}x + \frac{E_{max} + E_{min}}{2} \right)$$

$$= \sum_{n=1}^{N_{Ch}} c_n T_{n-1}(x), \quad (-1 \leq x \leq 1)$$

we then obtain

$$\int_{E_{min}}^{E_{max}} F_\beta^{Ch}(E)E \; dE$$

$$= \frac{1}{2}(E_{max} - E_{min}) \times \left[ (E_{max} + E_{min}) \right.$$

$$\times \left( \frac{c_1}{2} - \sum_{k=1} \frac{c_{2k+1}}{(2k-1)(2k+1)} \right)$$

$$\left. - (E_{max} - E_{min})\sum_{k=1} \frac{c_{2k}}{(2k-3)(2k+1)} \right].$$

Introducing the shorthand notation $I_\beta/2$ for this integral and setting $\bar{\rho} = N_{el}/[2N(E_F - E_{min})]$ where $N_{el}$ is the number of electrons in the system, we arrive at the final expression for the error $\epsilon_2$:

$$\epsilon_2 \approx \frac{1}{2}N_{el}\left[ \frac{I_\beta - \pi^2/(3\beta^2)}{E_F - E_{min}} - E_F - E_{min} \right]. \qquad (21)$$

Let us again consider as an example the $a$-C512 model with $\beta \approx 50$ hartree$^{-1}$ and 80 Chebyshev polynomials as used in Fig. 1. In this case, the relative error $\epsilon_2/|E_{bs}|$ obtained from Eq. (21) turns out to be $-1.6 \times 10^{-5}$, compared to its exact value of $-1.9 \times 10^{-5}$ as computed from all eigenvalues. This is again more than one order of magnitude less than the total errors, showing that (at least for the band-structure energy) the Chebyshev expansion used is sufficiently accurate for the Fermi distribution function applied. One should also note

TABLE I. Relative errors $\Delta E_{\text{bs}}$ in the band-structure energy for the $a$-C512 model computed with the projection method in dependence on the cutoff radii $R_{\text{loc}}$ for performing the matrix multiplications [Eq. (8)], and $R_{\bar{\mathbf{H}}}$ for computing the matrix $\bar{\mathbf{H}}$ [Eq. (6)]. $N_{\text{loc}}$ and $N_{\bar{\mathbf{H}}}$ are the maximum number of atoms within the spheres with radii $R_{\text{loc}}$ and $R_{\bar{\mathbf{H}}}$, respectively. The last column contains the CPU time (Ref. 38) needed for computing the matrices $\bar{\mathbf{H}}$ and performing the projections [Eqs. (6) and (8)].

| $R_{\text{loc}}$ ($a_0$) | $N_{\text{loc}}$ | $R_{\bar{\mathbf{H}}}$ ($a_0$) | $N_{\bar{\mathbf{H}}}$ | $\Delta E_{\text{bs}}$ | CPU (min) |
|---|---|---|---|---|---|
| 8.00 | 62 | 6.1 | 30 | $7.2 \cdot 10^{-4}$ | 7 |
|  |  | 7.0 | 45 | $5.9 \cdot 10^{-4}$ | 11 |
| 10.37 | 133 | 6.1 | 30 | $5.2 \cdot 10^{-4}$ | 18 |
|  |  | 7.0 | 45 | $3.5 \cdot 10^{-4}$ | 28 |
| 12.20 | 210 | 6.1 | 30 | $2.2 \cdot 10^{-4}$ | 33 |
|  |  | 7.0 | 45 | $4.9 \cdot 10^{-5}$ | 50 |

that the error $\epsilon_2$ can be positive or negative dependent on the special oscillating behavior of the Chebyshev approximation chosen.

Let us finally remark here that in the case of larger gaps one should estimate the error $\epsilon_2$ by explicitly taking the gap width into account setting the DOS to zero between $E_{\text{HOMO}}$ and $E_{\text{LUMO}}$. The formulas for this case can be derived in a similar manner to those presented above.

### 3. Use of a confinement radius for the matrix $\bar{\mathbf{H}}$

This problem was already addressed in Ref. 30. As these authors showed, when using the same confinement radii for $\mathbf{H}$ and $\bar{\mathbf{H}}$ the maximum matrix element of the product $\mathbf{S}\bar{\mathbf{H}}$ [cf. Eq. (6)] beyond this radius is at most of the same order of magnitude as the matrix elements of $\mathbf{H}$ at the cutoff radius. This means that the error in computed eigenvalues or band-structure energies due to the confinement of $\bar{\mathbf{H}}$ is also of this order of magnitude (or smaller).[30] For our Hamiltonian,[31] the matrix elements at the cutoff radius of 6.1 $a_0$ are $\approx 1 \times 10^{-3}$ hartree. This gives a relative error of about $< 10^{-3}$ for our band-structure energies which is approximately of the same order of magnitude than the total errors for the $a$-C512 model considered above. [For diamond, this error is about one order of magnitude smaller because the next (fourth) nearest-neighbor shell is placed at $6.74a_0$.]

To check the influence of the cutoff radius $R_{\bar{\mathbf{H}}}$ on the band-structure energy, we computed the relative error $\Delta E_{\text{bs}}$ for the radius $R_{\bar{\mathbf{H}}} = 6.1a_0$ used above as well as for the radius $R_{\bar{\mathbf{H}}} = 7.0a_0$ at which all our matrix elements in $\mathbf{H}$ naturally vanish. (Actually, we kept $R_{\mathbf{H}} = R_{\bar{\mathbf{H}}}$ in these computations because the increase in $R_{\mathbf{H}}$ as shown by direct diagonalization only leads to a relative change in the total band-structure energy of $2 \times 10^{-5}$, and can therefore be neglected here.) The results for three different LOC radii $R_{\text{loc}}$ are presented in Table I. As can be seen, the use of the smaller radius $R_{\bar{\mathbf{H}}}$ results in an error of about $2 \times 10^{-4}$. This error is comparatively small for small LOC radii but becomes essential for larger ones. In that case, the increase of $R_{\bar{\mathbf{H}}}$ relative to $R_{\mathbf{H}}$ or the increase of both cutoff radii may be useful. The reader may also notice from Table I that the computational cost of

the present implementation scales approximately linearly with the number of atoms $N_{\bar{\mathbf{H}}}$ within $R_{\bar{\mathbf{H}}}$ and $N_{\text{loc}}$ within $R_{\text{loc}}$. The numerical effort for computing the matrix $\bar{\mathbf{H}}$ which scales quadratically with $N_{\bar{\mathbf{H}}}$ is in general small compared to the effort for performing all the projections [cf. Eq. (8)]. The total numerical effort of the present method, therefore, scales approximately as $N N_{\bar{\mathbf{H}}} N_{\text{loc}} N_{\text{Ch}}$, where $N_{\text{Ch}}$ is the number of Chebyshev polynomials used as discussed above.

Finally in this paragraph, we should consider the numerical accuracy in the solution of the linear system of equations (6) for computing the matrix $\bar{\mathbf{H}}$. This accuracy is controlled by the condition number $z$ of the overlap matrices occurring in Eq. (6). The numerical uncertainty in the solution for $\bar{\mathbf{H}}$ is of the order of $z$ units of the last figure in the absolutely largest matrix element of $\mathbf{H}$. The condition number is given as the ratio of the absolutely largest to the absolutely smallest eigenvalue of a matrix. Since the overlap matrices are positive definite, these eigenvalues are in fact the extreme ones and can easily be computed by a Lanczos (recursion) procedure.

To estimate this influence, we computed the condition numbers for the orbitals used in the simpler Hamiltonian of Ref. 31 as well as for the Sankey orbitals used in Ref. 35. In the latter case, the relatively large extent and overlap of the pseudoatomic orbitals in the case of carbon resulted in the largest condition numbers we found which turned out to be $\approx 36$ (from extreme eigenvalues 0.093 and 3.37 in the $a$-C512 model mentioned above). For comparison, the number $z$ for the more localized orbitals of Ref. 31 was only 4.7. Keeping the other influences discussed in this section in mind, these results imply that an 8-byte-precision arithmetic as used by us in solving Eq. (6) allows overlap matrices which have by several orders of magnitude larger condition numbers before any significant influence of this problem may be expected.

### 4. Use of a finite localization radius within the projection

According to the discussion carried out so far, the LOC radius $R_{\text{loc}}$ applied within the matrix multiplications for performing the projection [cf. Eq. (8)] appears to be the most crucial parameter in the present scheme. For reasonable values of $\beta$ and $N_{\text{Ch}}$, the choice of the LOC radius determines the accuracy of the projection method for LOC spheres containing up to about 100 atoms. This result can already be inferred from Table I. In Fig. 2 we investigated in more detail the influence of $R_{\text{loc}}$ on the band-structure energy $E_{\text{bs}}$. We present the result for diamond and amorphous carbon; to reduce the effect of the finite cell size, we considered models with 512 atoms in both cases. To eliminate the influence of the confinement of $\bar{\mathbf{H}}$, the value $R_{\bar{\mathbf{H}}} = 7.0a_0$ was used for both models. Furthermore, we applied the values $\beta = 50$ hartree$^{-1}$ and $N_{\text{Ch}} = 100$ in these computations.

Figure 2 demonstrates the global decrease in the error $\Delta E_{\text{bs}}$ with increasing $R_{\text{loc}}$. But it shows another important and unexpected result. The decrease in $\Delta E_{\text{bs}}$ is not always monotonic, as it would be in an analogous variational procedure. Increasing $R_{\text{loc}}$ in diamond from $7.4a_0$ (including five nearest-neighbor spheres) to $8.8a_0$ (seven nearest-neighbor spheres) results in an *increase* in $\Delta E_{\text{bs}}$ by about one order of magnitude. Outside this region there is still no significant
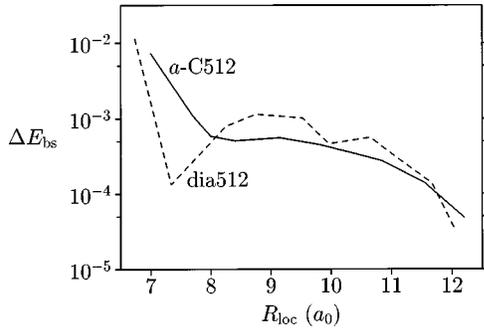
FIG. 2. Relative deviation of the band-structure energy computed with the projection method from the value obtained by direct diagonalization of the secular equation, plotted in dependence on the localization radius $R_{\text{loc}}$ for two models of diamond and fourfold-coordinated amorphous carbon containing 512 atoms.

reduction in $\Delta E_{\text{bs}}$ up to $R_{\text{loc}} = 10.7a_0$. In the amorphous model this behavior is to a large extent smeared out, leading to an almost monotonic decrease. This fact suggests that the stated effect is related to the symmetry of the crystal structure and consequently to the symmetry of the Wannier-like functions obtained within the projection. However, within the medium-range environment seen by the WF's, the amorphous structure is not as different from the crystal as needed to entirely blur the special radial dependence found in $\Delta E_{\text{bs}}$.

To check this behavior, we recalculated the radial dependence of $\Delta E_{\text{bs}}$ using the more elaborate Hamiltonian by Sankey and Niklewski,[35] which is based on the Harris functional[46] and the evaluation of all three-center integrals (more detailed results and further applications of these investigations will be presented in a subsequent paper). With these computations, we could exactly reproduce the qualitative shape of the curves in Fig. 2 showing the deep minimum at 47 atoms and the second flat minimum at 123 atoms. The error $\Delta E_{\text{bs}}$ at the first minimum even reached negative values. This again indicates that the effect seems to be the result of the superposition of two errors of opposite sign, one representing the general increase in $\Delta E_{\text{bs}}$ with decreasing $R_{\text{loc}}$, which, however, is modified by the way the projected functions spread through the crystalline environment during the matrix multiplications.

For comparison, we also investigated LOC regions which are determined by the number of bonds steps necessary to reach an atom. In this way we found that the region with the largest drop in energy (negative $\Delta E_{\text{bs}}$ for both Hamiltonians applied) is composed of all atoms within three bond steps starting from one central atom. This structure contains 41 atoms and is obtained from the minimum structure in Fig. 2 by breaking off eight-membered rings emanating from the central atom. Any further reduction in the number of atoms leads to an abrupt rise in the band-structure energy, whereas the enlargement of the LOC region after an increase in $\Delta E_{\text{bs}}$ results in a convergence to the exact energy from higher-energy values within numerical accuracy.

The observed behavior in $\Delta E_{\text{bs}}$ can be traced back to the amount of ''charge'' maintained at the initial basis orbitals when performing the projections. We estimated this charge by computing Mulliken's atomic gross populations $N_g$ for the Wannier-like functions resulting from the projection. The minimal-energy LOC regions discussed above are character-

ized by an increase in $N_g$ at the central atom and a corresponding increase in the norm of these functions by about $10^{-3}$. This is just the magnitude of the effect found in Fig. 2. Apart from this fact, the radial dependence of the gross populations shows the expected exponential behavior but is again nonmonotonic with respect to the radius of the LOC region. We will investigate this radial dependence more thoroughly in Sec. IV, where we will start from more appropriate initial functions constructed from hybrid orbitals.

## IV. COMPUTATION OF GENERALIZED WANNIER FUNCTIONS

The projection method as described so far works without explicitly calculating Wannier-like functions. In this section, however, we want to show how the method can effectively be used to compute such functions which are linearly independent and span the occupied subspace of the Hamiltonian. Furthermore, as has been known for a long time,[22,40] orthogonal WF's in general have worse localization properties compared with nonorthogonal ones. We therefore investigate the localization behavior of our functions before and after orthonormalization. Furthermore, we will investigate how well approximate WF's which are truncated to certain LOC regions can be orthonormalized by a linear-scaling procedure. We want to emphasize here that it is not immediately clear how well an order-$N$ orthonormalization should work at all. The reason is that in every orthonormalization (ON) step, localized functions will acquire some weight outside their previous LOC regions, but one has to truncate these functions to ensure the linear scaling of the method. Such a scheme, therefore, must not significantly increase the number of atoms in these regions.

The technique of applying (band) projection operators for computations of WF's within ordinary and one-dimensional perturbed crystals was already used several years ago in Refs. 41–43. The projection method presented here is essentially similar in spirit to these former computations. However, the main differences, as described above, consist of the actual construction of our projection operators by a Chebyshev representation of the Fermi-Dirac operator instead of using any eigenfunctions, and in the formulation of the method as a real-space approach, thus allowing an effective linear-scaling implementation of this technique.

To obtain linearly independent WF's, a generally applicable procedure consists in selecting the correct number of initial functions at random. In systems with four valence orbitals per atom and as many electrons as the total number of valence orbitals, one could use two atom-centered functions per atom formed by random combinations of the atomic orbitals at one atom. However, such functions result in relatively large overlap values between the projected functions which in turn make their orthonormalization more difficult (especially in the presence of localization constraints). This also leads to inferior localization properties of the approximately orthonormalized Wannier-like functions. For locally well-relaxed structures we therefore prefer an approach which takes local physics in terms of bonding information into account. Such functions are especially easy to obtain in completely fourfold-coordinated covalent materials which are the systems of interest in this section. In these systems, a

TABLE II. Norms for various atomic states before and after their projection onto the occupied subspace. The first three rows contain orbitals at one atom, whereas the last two rows contain the bonding and antibonding combinations of $sp^3$ hybrid orbitals at neighboring atoms pointing in the direction of the bond between these atoms.

| initial state | norm | |
|---|---|---|
| | initially | after proj |
| $\lvert s\rangle$ | 1.0 | 0.778 |
| $\lvert p\rangle$ | 1.0 | 0.654 |
| $\lvert sp^3\rangle$ | 1.0 | 0.685 |
| $\lvert sp^3\rangle_A+\lvert sp^3\rangle_B$ | 2.737 | 2.718 |
| $\lvert sp^3\rangle_A-\lvert sp^3\rangle_B$ | 1.263 | 0.023 |



FIG. 3. Mulliken's atomic gross populations $N_g$ for orthonormalized bond-centered Wannier functions in diamond in dependence on the distance from the bond center; dots, functions obtained after projection; circles, Löwdin-orthonormalized functions. No localization constraints were applied.

natural choice for the initial functions needed to compute WF's is provided by the bonding combinations of ($\sigma$-like) hybrid orbitals which point in the direction of the bond between two neighboring atoms.[2,20,21] As shown by Des Cloizeaux[47] and Kohn,[48] orthonormal occupied WF's in the diamond lattice can be chosen to belong to the identity representation of the point group associated with the bond center between the neighboring atoms. Bonding pairs of hybrid orbitals already have all these symmetry elements, but their excess symmetry is broken during the multiplications with the Hamiltonian matrix. The suitability of hybrid orbitals as initial guesses for WF's can also be seen by considering the change of these functions after projection onto the occupied subspace of $\hat{H}$. Table II contains the norms of the bonding and antibonding combinations of $sp^3$ hybrid orbitals in diamond before and after projection. The bonding hybrid pair changes only slightly and is to a large extent situated in the occupied subspace. Correspondingly, the antibonding pair disappears almost entirely during the projection. For comparison, we also included in Table II the norms for the functions generated by orbitals at one atom. These orbitals possess large contributions in both subspaces.

In fourfold-coordinated amorphous systems, bond-directed hybrid orbitals can be used in a similar manner as initial orbitals for WF's. They immediately lead to the correct number of Wannier states and again reflect the local symmetry of the structure. However, in amorphous systems these bond-directed orbitals are, in general, nonorthogonal also to the orbitals attached at the same atom.

In the following, we first investigate in more detail the properties of orthonormalized WF's in diamond as obtained by the projection method. Let us then discuss the generalizations needed and the results obtained for a model of fourfold-coordinated amorphous carbon. To reduce the influence of the unit cell size, we use in all cases models containing 512 atoms.

### A. Diamond

Let us first discuss the properties of our WF's obtained without any localization restrictions. In Fig. 3 we present Mulliken's atomic gross populations $N_g$ for the normalized functions immediately after projection (dots) and those functions obtained by a subsequent Löwdin orthonormalization (circles). One advantage of using hybrid orbitals as initial
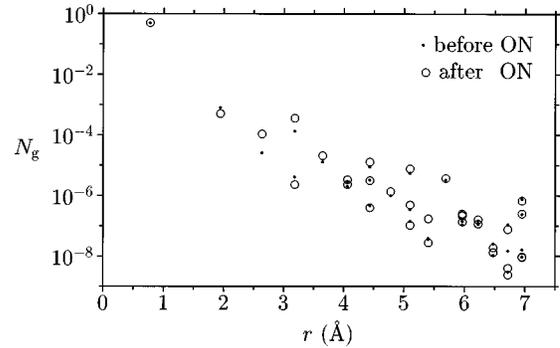
functions is that the functions resulting from the projection already have relatively small overlap values (0.031 in our case[31] for functions centered at neighboring bonds). This means that during a subsequent orthonormalization, these functions will not change very much. In fact, as can be seen, the largest changes occur at atoms within about 3 Å or two bond steps around the initial bond. The atoms forming the central bond and their nearest-neighbor atoms lose some weight, whereas the ''charge'' at the atoms one bond step farther away increases by about or less than one order of magnitude. The change in the weight at most of all other atoms is very small. In particular, the exponential decay does not noticeably change during orthonormalization when starting from hybrid orbitals. However, as we will show below, one must not infer from these results that an easy orthonormalization of these functions would be possible by simply truncating the tails of the projected functions when LOC constraints are used.

The global exponential decay of our WF's in diamond is clearly visible in Fig. 3. This fact suggests that these bond-related functions are close to the optimally localized orthonormal occupied Wannier functions in this system. Note that for overlapping energy bands exponentially localized Wannier functions are not uniquely determined;[48] instead different sets of such functions can be transformed into each other by unitary transformations. Apart from the exponential decay, however, the radial dependence of these functions is by no means monotonic in the (real-space) distance from the bond center. (We had already referred to this fact in Sec. III B.) One immediate suggestion is that the weight of a WF at a certain atom should also significantly depend on the number of bond steps needed to reach this atom when starting from the central bond. This dependence is demonstrated in Fig. 4. In this figure, different symbols at the same real-space distance $r$ describe different distances from the central bond in terms of bond steps. This ''bond-step'' distance is indicated in the lower panel of Fig. 4. At the same distance $r$, atoms marked with diamonds are separated by more bond steps from the bond center than atoms marked with circles. The expected result here is that the diamonds in general also have lower gross populations than the circles (see the upper panel in Fig. 4). However, both distances (the real-space distance $r$ and the bond-step distance) are not yet sufficient to
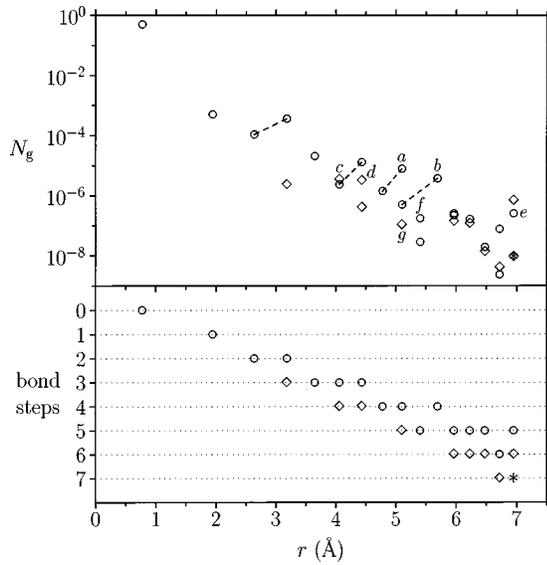
FIG. 4. Mulliken's atomic gross populations $N_g$ for orthonormalized bond-centered Wannier functions in diamond as given in Fig. 3. For each distance $r$ from the bond center, the symbols used describe the number of bond steps needed to reach a certain atom starting from the central bond. This number of bond steps is indicated at the same distance in the lower panel. The dashed lines and letter symbols are explained in the text.



FIG. 5. Mulliken's atomic gross populations $N_g$ for orthonormalized bond-centered Wannier functions in diamond as given in Fig. 3. For each distance $r$ from the bond center, the symbols used describe the distance $d_{bd}$ of a certain atom from the axis specifying the direction of the central bond. This distance is indicated at the same $r$ value in the lower panel. The dashed lines connect the same atoms as in Fig. 4.

describe the spatial dependence of the WF's. In particular, there are frequent cases (compare the atom groups connected by dashed lines in Fig. 4) in which of two atoms with the same bond-step distance the more distant atom in terms of $r$ acquires the *larger* weight within the WF. Note that these differences in $N_g$ can reach about one order of magnitude.

To gain a better understanding of this effect, we replot Fig. 4 (see Fig. 5), but now different symbols at the same "spherical" distance $r$ indicate different distances of these atoms from an axis along the direction of the central bond. This figure therefore describes the nonspherical spatial behavior of the WF's beyond the bond-step picture used above. Indeed, in all the cases just described (which again have been marked with dashed lines), the more distant atoms with the larger gross populations are situated closer to the bond direction than the corresponding atoms at smaller $r$. Furthermore, the diamonds which now describe larger distances or angles from the bond direction, are again mostly situated at smaller gross populations in the upper panel of Fig. 5. This behavior is only weakened for atoms far away from the originating bond. We therefore conclude that the WF's in diamond are more ellipsoidal than one would expect from a picture where the decay of the WF's is determined by the spherical and bond-step distances alone.

To illustrate these quantitative results, in Figs. 6 and 7 we present the charge densities of the orthonormalized untruncated Wannier function in diamond as obtained with our Hamiltonian.[31] Figure 6 shows the charge density within a (110) plane which contains the nearest-neighbor bonds. The plane in Fig. 7, on the other hand, is tilted around the central bond in the former plot to include the "perpendicularly located" second-nearest-neighbor atoms of the bond [(211) plane]. Note that the second-nearest neighbors in Fig. 6 and the second-nearest neighbors in Fig. 7 are just the atoms
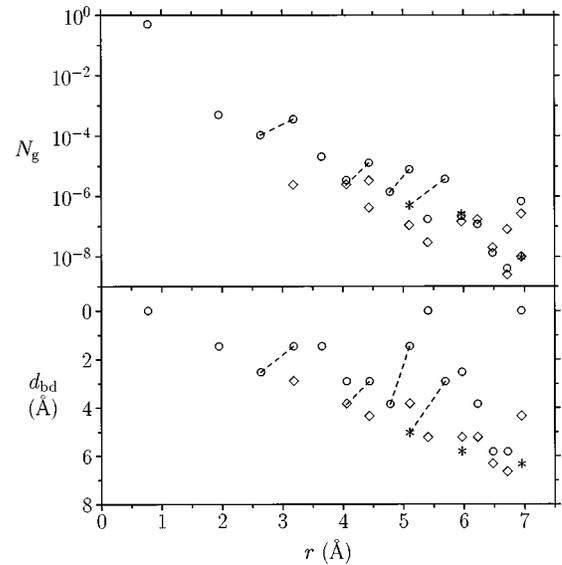
connected by the dashed line near 3 Å in Figs. 4 and 5. Furthermore, the corresponding fourth-nearest-neighbor atoms in both charge-density plots are connected by the dashed line at 5.5 Å in the former figures. In accord with these examples, Fig. 6 shows a preferred "flow of charge" in the bond directions in the (110) plane. As we found, this spatial direction leads to larger weights of the WF's compared to the corresponding neighbors at the same bond-step distances but located perpendicular to the central bond in Fig. 7. Independent of this discussion, both figures clearly show the accumulation of bond charge in the central region of the bond, the minor accumulation of charge at neighboring atoms, and the "impressive" nodal structure of the WF's created by the superposition of the $2s$ and $2p$ atomic valence orbitals.

The nonspherical spatial behavior of the WF's has important consequences for the optimal shape of the LOC regions when one has to truncate these functions to obtain an order-$N$ computational scheme. Unfortunately, we have not found a simple parametrization of such a LOC region, say, in form of an ellipsoid, where one could have fixed one parameter and vary the other one with respect to the accuracy desired. During the projection, when starting with hybrid orbitals, we therefore employ LOC regions determined by a critical number of bond steps, as was done in previous work using variational methods.[2,20] After every ON step, however, we redefine each LOC region in such a way that it contains those atoms which have accumulated the largest weights within a WF. This dynamical reshaping of the LOC region takes the observed directional dependence of the WF's into account, and simultaneously allows these functions to expand to those atoms which have acquired significant weight during the ON process. As we found, this method results in a decrease in the remaining overlap values during an order-$N$ orthonormalization by about one order of magnitude. Also note, as we will
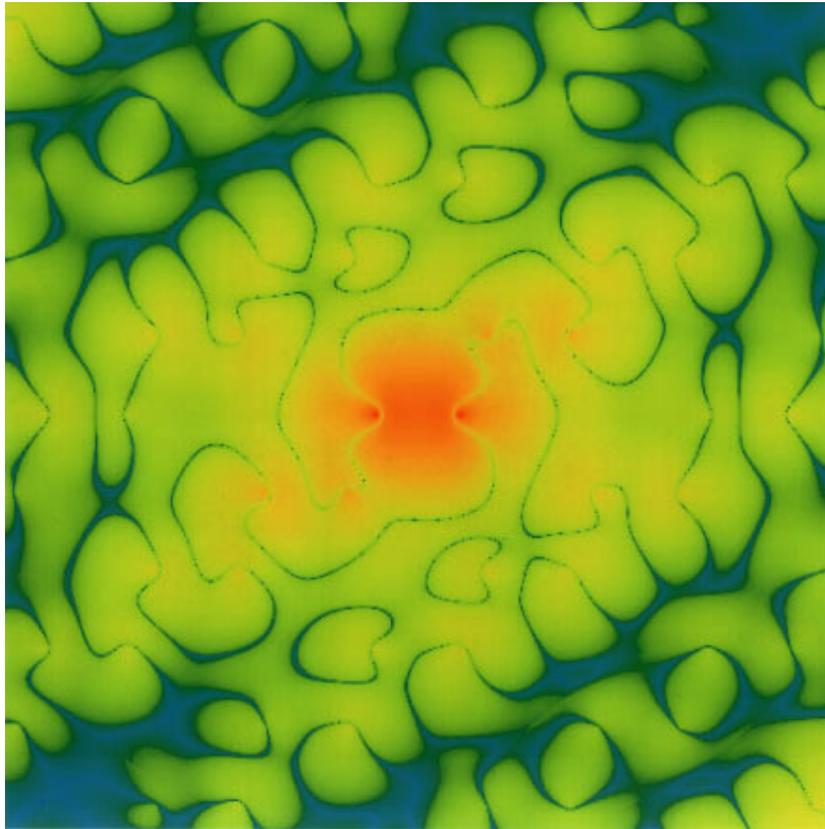
FIG. 6. (Color) charge-density plot of the untruncated and orthonormalized projected Wannier function within the (110) plane in diamond. The colors have been mapped to the logarithms of the charge density between its maximum value 7.8 (red) and minimum value $6.4 \times 10^{-21}$ (blue, values in electrons/$a_0^3$) found for the resolution of the figure.
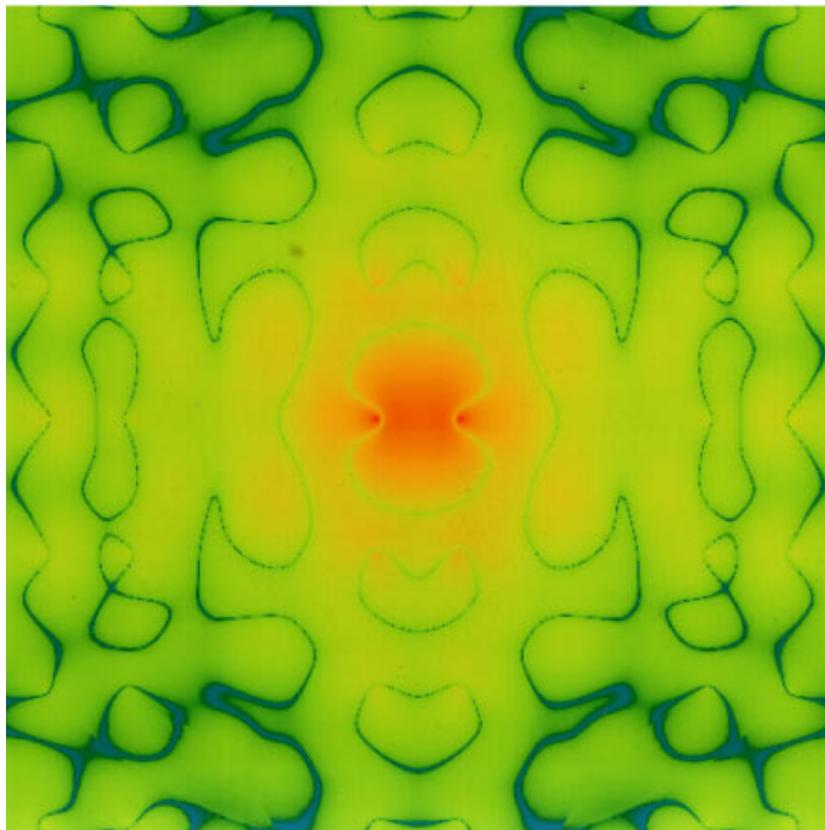


FIG. 7. (Color) charge-density plot of the untruncated and orthonormalized projected Wannier function within the (211) plane in diamond. The mapping of the colors is the same as in Fig. 6.

show in Sec. IV B, that this flexible scheme of defining LOC regions is not only useful in crystalline systems, but improves the computation of WF's in amorphous solids as well.

We have also compared different versions of the general Löwdin-related orthonormalization scheme with respect to using full or truncated overlap matrices. These results are reviewed in the Appendix. In general one has to decide whether to use one given overlap matrix for multiple ON steps, or to recompute this matrix for each new set of iterated functions. In our implementation the CPU time needed for computing the overlap matrix is small (of the order of one-tenth) compared to the time for performing the projections. The most effective scheme, therefore, is to use repeated first-order Löwdin iterations in conjunction with the dynamical determination of the LOC regions as described above.

When the atoms within a LOC region can change, one has to decide how the *number* of atoms may change there during an ON run. We do not allow the LOC regions to shrink; to prevent a distortion of the symmetry of the WF's in crystals, however, one then has to permit at least a slight increase in the number of atoms in these regions. One can minimize the growth of the LOC regions by, e.g., introducing a certain critical weight difference for additional atoms at the LOC boundaries. This leads primarily to an exchange of atoms altering the LOC region into a more ellipsoidal shape. To avoid such an additional parameter, one would have to leave all atoms within a LOC region which once have been there. Such an approach increases the accuracy of the orthonormalization, but is only reasonable as long as these regions do not expand up to the complete next "bond shell" of atoms. This does not happen in diamond, for which we will show the results of this latter version. Results with an almost constant number of atoms in the LOC regions will be presented for amorphous carbon in Sec. IV B.

The three columns of Fig. 8 demonstrate the results for the order-$N$ orthonormalization of WF's in diamond starting from different LOC conditions.[49] In the first two columns, the initial LOC regions for performing the projection contain all the atoms within three bond steps starting from the originating bond. This number is 4 in the last column. Another distinction is the number of overlap values taken into account during the ON process. As it turns out, it is not necessary to compute the overlap between all overlapping WF's. The first and last column in Fig. 8 are characterized by using only the overlap values between functions for which the originating bond of one function is attached at an atom within the LOC region of the other Wannier state. In the middle column, we use the overlaps to functions which are centered up to two bonds apart from the atoms in one LOC region. According to the figure, let us refer to these three situations as (a), (b), and (c), respectively.

In the top panels of Fig. 8, the full circles indicate the maximum of the (absolute) overlap values which are taken into account during an ON run. To check the maximum overlap between all WF's, the open circles in the same panels show the maximum of the (absolute) overlap values which are not included in an ON run. As another measure of the efficiency of an ON procedure, we have also computed the quantity

$$\Delta^2 = \frac{1}{M}\left[\sum_{ij}^{M} (\delta_{ij} - S_{ij})^2\right] \quad (22)$$

used in Ref. 50 (and, in a similar form, in Ref. 28) where here $S_{ij}$ are the overlap matrix elements between the Wannier states, and $M$ is the total number of WF's. In our case, we normalize all WF's after every ON step; hence $\Delta^2$ is just the square of the euclidic norm of the off-diagonal Wannier overlap matrix, divided by $M$ to make this quantity independent of the system size. To reduce CPU time in an actual computation, of course, one would not want to calculate overlap values which are not used in the ON run. In this respect, the question arises, how well the efficiency of the ON procedure can be inferred from the included overlap values only.

Considering Fig. 8, we first recognize that, when starting with hybrid orbitals, a small number of ON steps is sufficient to maximally reduce the largest overlap values. In our case, we reached a minimum of about $1.6 \times 10^{-3}$ in cases (a) and (b) and about $5 \times 10^{-4}$ in case (c). However, in cases (a) and (c) the ON process becomes inefficient when the overlap values included in an ON run reach the magnitude of the omitted values. In case (b), the LOC regions are much smaller than the regions where the overlap values are calculated. Therefore, the omitted overlap values are now negligible, but the WF's cannot form the tails necessary to reduce all the overlaps within the overlap regions. The process terminates "by itself" and is therefore a purely linear-scaling scheme with inclusion of an efficiency criterion.

However, the increase in the number of overlap values to be calculated when going from column (a) to column (b) does not significantly improve the accuracy of the ON scheme. Cases (a) and (c) are therefore more efficient approaches. Furthermore, when considering $\Delta^2$, we found that the procedure in general terminates after about four ON steps (when using hybrid orbitals as initial functions). Using this, it may be sufficient to extend the LOC regions by one bond to define an appropriate overlap region thereby greatly reducing the number of overlap values to be computed. Note that the final values attained for $\Delta^2$ after four ON steps are about $1.7 \times 10^{-5}$ in cases (a) and (b), and about $3.7 \times 10^{-6}$ in case (c).

The full circles in the bottom panels of Fig. 8 show the increase in the number of atoms in the LOC regions. Remember that we did not remove any atoms from this region. In diamond, an increase occurs only during the first ON run; this extends the LOC regions from 56 to 80 atoms when starting with three bond steps, and from 110 to 130 atoms when starting with four bond steps. In the first case, the new atoms are exclusively four bond steps away from the central bond and marked (in the order of decreasing weight) with $a$, $b$, $c$, and $d$ in Fig. 4. In the second case, all new atoms are separated by five bonds from the original bond and marked with $e$, $f$, and $g$ in the same figure.

As an example, in Fig. 9 we present the approximately orthonormalized and truncated WF's corresponding to column (a) in Fig. 8. For comparison, we again indicated the gross populations for the unconstrained functions. We found that the truncation of the tail of the WF raises the weights at all atoms except the two atoms at the originating bond. As can be seen in Fig. 9, the largest relative differences occur at
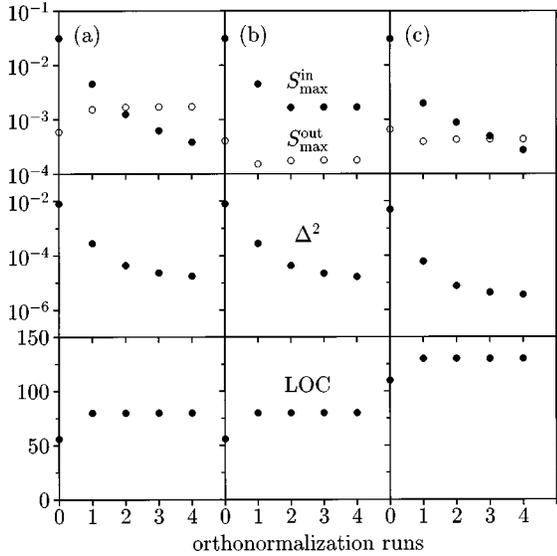
FIG. 8. Efficiency of an order-$N$ orthonormalization of truncated Wannier functions in diamond. Columns: (a) localization of WF's within 3/overlap computation within four bond steps from central bond of a WF; (b) 3/5; (c) 4/5. Rows: $S_{max}^{in}$, maximum overlap between WF's within region of overlap computation; $S_{max}^{out}$, maximum of remaining overlaps not used in orthonormalization; $\Delta^2$, orthonormality measure (22); LOC, number of atoms within localization regions of WF's; ON run 0 means after projection.

the boundary of the LOC region. It is just the WF's centered beyond the boundaries of the LOC regions which produce the maximum remaining overlap values presented in the top panels of Fig. 8. The truncation prevents a proper adjustment of the WF's to reduce these overlap values. On the other hand, the truncation reduces the weights at the two central atoms of a WF (by about $10^{-3}$, not visible in Fig. 9) which in general raises the mean energy of this WF.

## B. Fourfold-coordinated amorphous carbon

As already mentioned, the technique of using bonding combinations of neighboring hybrid orbitals as the initial functions for computing Wannier states can be applied in completely fourfold-coordinated amorphous systems as well. However, there is the following difficulty related to this approach.

When using a minimal valence orbital set ($s^1p^3$), it is in general not possible to compute orthogonal hybrid orbitals at fourfold-coordinated atoms, which point in the bond directions. One could solve a minimum problem to find those orthonormal hybrids which deviate least from these bond directions. It is, however, much easier to keep the hybrid orbitals in the directions of the bonds and to allow finite overlap values even for orbitals at one atom. As an easy way to find physically reasonable hybrids in this case, we compute the uniquely determined hybrid orbitals for all possibilities of selecting three of the four neighboring atoms, and then average over the resulting $s$ and $p$ characters. For the majority of atoms in a locally relaxed structure, the resulting overlap values are still very small. But even at locally distorted atoms, the initial overlap values are still significantly smaller compared, e.g., with the case of using random atom-centered functions.
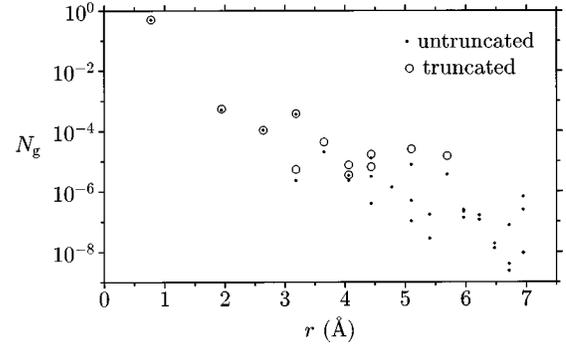


FIG. 9. Mulliken's atomic gross populations $N_g$ for approximately orthonormalized truncated WF's in diamond corresponding to column (a) in Fig. 8 (circles) in comparison with orthonormalized unconstrained WF's (dots).

We have computed the WF's in a model of fourfold-coordinated amorphous carbon containing 512 atoms in the unit cell. This model is a relaxed version of the original system generated by Djordjevic, Thorpe, and Wooten,[39] which we used in Sec. III. This original model, however, contains a few atoms which are only threefold-coordinated within any reasonable nearest-neighbor distance. The subsequent relaxation[51] of this model based on the Harris *ab initio* Hamiltonian[35] improved the fourfold coordination of this system and, as a consequence, produced a structure with a large HOMO-LUMO gap of width (4.3 eV) slightly smaller than that computed for diamond (5.9 eV).

In Fig. 10, we present the radial dependence of the gross populations for orthonormalized WF's without using localization constraints. Given are the averaged and maximum populations within distance intervals of 0.25 Å. For comparison, the dots indicate the corresponding WF's found in diamond. Figure 10 shows the clear exponential decay of our WF's even in the amorphous structure. This result is in accord with theoretical investigations[52] which verify the existence of exponentially localized WF's in nonperiodic systems obtained in some way from periodic ones. However, the amorphous structure shows a much greater variation of the populations within a certain distance interval compared with diamond. This is obviously the result of the local distortions of the atomic environments. Unlike diamond, the minimum gross populations even become negative in the amorphous structure. Although this is considered a deficiency of Mulliken's population concept, it again emphasizes the greater variations in the WF's in amorphous systems.

Considering the mean populations, it is remarkable that the exponential decay of the WF's has not significantly changed compared to the crystalline structure. It has been known for a long time[5,6] that the exponential decay of the WF's stemming from nondegenerate bands depends on the position of the branch points which connect these bands in the complex $k$ plane. The Wannier functions decay faster with increasing distance of these points from the real $k$ axis. To estimate this decay, one usually relates this distance to the width of the smallest gap between the two connected bands. In our case of covalently bonded insulators, the falloff of the WF's would be determined by the gap between the valence and conduction bands.
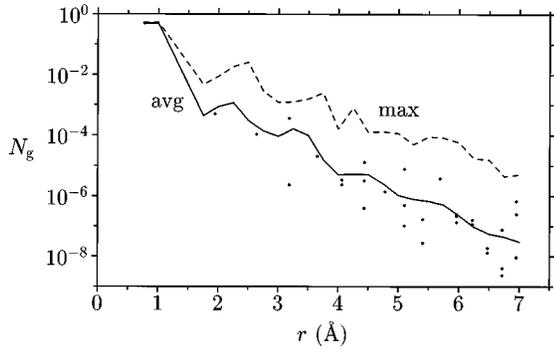
FIG. 10. Mulliken's atomic gross populations $N_g$ for orthonormalized bond-centered Wannier functions in a model of fourfold-coordinated amorphous carbon in dependence on the distance from the bond center; solid line, population averaged over all atoms within a distance interval of 0.25 Å; dashes line, maximum population in the same interval; dots, results for diamond as given in Fig. 3. No localization constraints were applied.

However, the LDA-based TB-like Hamiltonian[31] used in Fig. 10 shifts the conduction bands higher in energy which gives rise to an overestimation of the HOMO-LUMO gaps by a factor of about 1.5, and a slight increase in the ratio of the gaps in diamond and the amorphous system to 1.7. This displacement is, in a sense, similar to that created by a scissors operator, which leaves the valence bands and, correspondingly, the occupied Wannier functions unchanged.

Interestingly, our results agree with an early estimation given by Kohn[5] for the exponential decay of the WF's in one-dimensional TB-like systems. He found that a branch point between two bands in such systems may be situated close to the lower one of these bands, and that the decay of the WF's increases with the binding energy of these states. Indeed, the HOMO states responsible for the smallest falloff are reproduced with quite similar energies in our models for diamond ($-8.1$ eV) and amorphous carbon ($-8.6$ eV). This could therefore be the reason for the similar decay lengths found in Fig. 10.

Finally, let us again investigate the properties of truncated WF's within an order-$N$ orthonormalization. Unlike diamond, we kept the number of atoms in the LOC regions nearly constant. (This was done by choosing a relative critical weight difference for additional atoms at the boundary of the LOC regions of 0.1.) In the amorphous system, when allowing the LOC regions to expand, we found that they grow during several ON steps up to a saturated state which comprises approximately all atoms within the next bond shell of atoms. It is then better, however, to start immediately with larger LOC regions and to allow essentially only an exchange of atoms in these regions. This is done in Fig. 11.

First note that the maximum initial overlap values between WF's are about one order of magnitude larger than in the case of diamond. This is a consequence of using nonorthogonal hybrids, especially at locally distorted atoms. However, as asserted above, the majority of the overlap values is still relatively small. This can be seen at the initial values for $\Delta^2$ which are only slightly larger than the corresponding values in diamond. Furthermore, the fact that we prohibited a significant enlargement of the LOC regions leads in any case to a saturation of the maximum overlap values included in
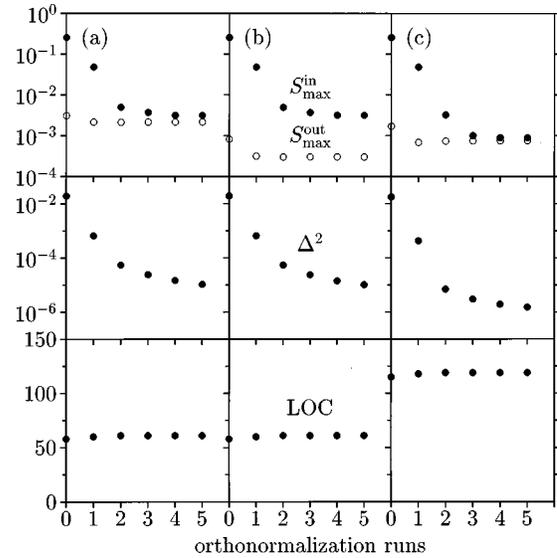


FIG. 11. Efficiency of an order-$N$ orthonormalization of truncated Wannier functions in a model of fourfold-coordinated amorphous carbon. The meaning of the symbols is the same as in Fig. 8, except that LOC is the mean number of atoms within LOC regions. The number of atoms in the LOC regions was kept nearly constant.

the orthonormalization. This saturation value is close to the magnitude of the neglected overlap values in cases (a) and (c) in Fig. 11. Nonetheless, the maximum number of atoms included in one LOC region was 20 in cases (a) and (b) and 30 in case (c). As in diamond, this reordering of atoms was now primarily performed during the first ON run. Bearing these facts in mind, it is very encouraging that the final values achieved for $\Delta^2$ are comparable to or even slightly better than those found in diamond [about $1.0 \times 10^{-5}$ in cases (a) and (b) and $1.5 \times 10^{-6}$ in case (c) after five ON steps]. This seems to be an effect of the amorphous structure itself and will be the subject of further investigations. However, we again found the expected result that the use of expanding LOC regions even more reduces the final deviations from orthonormality—but this is associated here with almost a doubling of the number of atoms in the LOC regions and a corresponding enlargement of the localization range of these functions.

## V. CONCLUSIONS

In this paper, we presented a generalized version of the order-$N$ projection method developed by Goedecker and co-workers,[25] in which electronic properties of an insulating system are computed by explicitly performing a projection to the occupied subspace of the Hamiltonian. The method can now be used with nonorthogonal sets of basis orbitals thus allowing an effective application of this approach within local-basis *ab initio* Hamiltonians. The overlap between the basis states is taken into account by solving linear systems of equations $\mathbf{S}\bar{\mathbf{H}} = \mathbf{H}$ within the cutoff radii of $\mathbf{H}$ as proposed by Gibson, Haydock, and LaFemina.[30] However, to increase the accuracy of the method it may be necessary to use somewhat larger cutoff radii for the computation of the matrix $\bar{\mathbf{H}}$.

We explored the projection method within two different computations using a simple non-self-consistent LDA-based

Hamiltonian.[31] First, we investigated the accuracy of the scheme as a means of performing traces within the occupied subspace. This was done by computing the band-structure energy for models of crystalline and amorphous carbon containing up to 4096 atoms. A detailed discussion of the possible errors in the method showed how the relative deviation from corresponding results based on eigenvalue calculations can be reduced below $10^{-5}$. The most crucial parameter in the scheme is the extent of the localization regions for performing the traces. We found the interesting effect that, especially in crystals, the error in the band-structure energy upon increasing the localization radius decreases in a nonmonotonic way in the projection method.

Additionally to the band-structure energy, we give all necessary formulas for computing other electronic quantities including density matrices, the moments of the density of states, forces, and the electron density. In principle, the method can therefore be applied directly within molecular-dynamics simulations as well as within self-consistent computational schemes.

As a second main application of the projection method, we investigated its use for computing first-principles Wannier-like functions in covalent insulators at zero temperature. In this paper, we restricted ourselves to completely fourfold-coordinated systems in which very efficient initial functions needed for performing the projections are provided by the bonding combinations of neighboring ($\sigma$-like) hybrid orbitals which point in the bond directions. This choice results in the correct number of exponentially localized projected functions which span the occupied subspace of the Hamiltonian. Furthermore, the use of hybrid orbitals results in relatively small overlap values between the projected functions. A subsequent orthonormalization can therefore be done with a small number of iterations (about 3–5).

We computed orthonormal Wannier functions without localization constraints as well as approximately orthonormalized functions with localization restrictions for models of diamond and fourfold-coordinated amorphous carbon containing 512 atoms. The radial dependence of the unconstrained functions was investigated in detail. At least within the first four ''bond shells'' which are determined by a constant number of bonds starting from the original bond, these functions turned out to be more ellipsoidal than the corresponding bond shells of atoms. Based on this result, we implemented an approximate linear-scaling orthonormalization scheme for truncated Wannier functions which employs a dynamical reshaping of the localization regions during the orthonormalization runs. With this technique, we could reduce the quantity $\Delta^2$ [used for estimating the deviation from orthonormality, cf. Eq. (22)] to about $10^{-5}$ or $10^{-6}$ for localization regions determined by three or four bond steps from the original bond, respectively. This was achieved with almost no or only moderate increase in the number of atoms in these regions.

In order to use the projection method for the computation of Wannier functions in general systems with coordination numbers less than 4, we need a scheme for the selection of the correct number of linearly independent initial functions in these systems. According to our experience with fourfold-coordinated systems, it would also be highly desirable to retain the use of hybrid orbitals in such structures. This approach can still straightforwardly be applied if the undercoordinated atoms form conjugated $\pi$ bonds as in a number of molecules and fullerene systems.[2,3] In more general structures, however, this pairwise selection of hybrid orbitals can no longer be used exclusively. In a subsequent paper, we will therefore present a quite general method, based on the evaluation of local Gram determinants, which nevertheless allows a selection of linearly independent hybrid functions.

Finally, we are currently exploring the use of our Wannier functions in computations of the electric polarization and the dielectric function as proposed recently.[27,28] These ideas seem to open a field of very important applications of Wannier functions in the theoretical investigation of materials. We will present our results related to this approach in a future paper.

## APPENDIX: ORTHONORMALIZATION SCHEMES

In this appendix, we will compare two different schemes for the orthonormalization of an extended set of linearly independent nonorthogonal functions $\{|\varphi_i\rangle\}$. In particular, these functions may be confined to certain LOC regions if the orthonormalization is to be linearly scaling. The first approach consists in applying repeated first-order Löwdin iterations[20,42,53]

$$|\varphi_j'\rangle = |\varphi_j\rangle - \tfrac{1}{2}\sum_{i(\neq j)} |\varphi_i\rangle S_{ij} \qquad \text{(A1)}$$

($S_{ij}=\langle\varphi_i|\varphi_j\rangle$), where the functions after every ON cycle have to be renormalized. This scheme has the following features.

(1) In the sense of a perturbation theory,[42] the iterations (A1) reduce the overlap between the functions $\{|\varphi_i\rangle\}$ in every step according to a power law. In fact, we found safe convergence even in cases where the largest off-diagonal matrix elements reached values close to 1.0 (as observed for Wannier functions created by random atom-centered orbitals). However, the price to be paid is that one has to recalculate the overlap matrix for every new set $\{|\varphi_i'\rangle\}$. This may be an essential point in cases where the computation of the overlap matrix determines the CPU time of the procedure.

(2) It is obvious from the previous point that due to the repeated updates of the $S_{ij}$ matrix only relatively few iterations are necessary to achieve a desired accuracy in the orthonormalization. A maybe important technical consequence for large systems is that the same memory space can be used for storing the overlap matrix and the new WF's resulting from the ON steps.

(3) When using truncated functions $\{|\varphi_i\rangle\}$, the new functions $\{|\varphi_i'\rangle\}$ extend beyond these localization boundaries. This means that in each ON step, in principle, one has to calculate an increasing number of overlap values. Of course,

if one keeps the LOC regions fixed, the number of overlap values will remain invariant as well. However, as we pointed out in Sec. IV, the accuracy of such a linear-scaling orthonormalization scheme can be significantly increased by redefining the LOC regions. This in general increases the number of overlap values to be computed.

In order to avoid the recomputation of the overlap matrix, we have also examined an orthonormalization scheme which is based on an extension of series (A1):

$$|\varphi'_j\rangle = |\varphi_j\rangle - \frac{1}{2}\sum_{i(\neq j)}|\varphi_i\rangle S_{ij} + \frac{3}{8}\sum_{\substack{i(\neq j)\\k(\neq i,j)}}|\varphi_i\rangle S_{ik}S_{kj} + \cdots.$$ 
(A2)

Of course, series (A2), using a fixed overlap matrix, has worse convergence properties than the previous scheme. One can, however, improve the convergence behavior of Eq. (A2) by a technique which is similar to the overrelaxation method used for solving linear systems of equations.[12] For the original Löwdin equation $\mathbf{x} = \mathbf{S}^{-1/2}\mathbf{v}$, this method leads to the iterative process (given here for $S_{ii} = 1$)

$$\mathbf{x}^{(k+1)} = -f_k(t\mathbf{S}_{\text{off}} + t - 1)\mathbf{x}^{(k)} + \sqrt{t}\mathbf{v}, \qquad (A3)$$

where $f_k = (I - k + 1/2)/(I - k + 1)$, $I$ denotes a prechosen total number of iterations $(0 \leq k \leq I)$, $\mathbf{S}_{\text{off}}$ is the matrix $S_{ij}$ with the diagonal entries replaced with zeros, and $t$ is an overrelaxation (or underrelaxation) factor (here often $t \approx 0.75$–0.95). The advantage of Eq. (A3) is that one (full or truncated) overlap matrix can be used for multiple ON steps. In fact, we found remarkable convergence improvements in cases with not too large euclidic norms of $\mathbf{S}_{\text{off}}$ [cf. Eq. (22)]. Unfortunately, the convergence of Eq. (A3) can still happen to fail or becomes rather slow for larger initial deviations from orthogonality (as observed for random atom-centered WF's). Moreover, as already pointed out, the CPU time needed for computing the overlap matrix turned out to be small compared to the time for performing the projections. The best results have therefore been found with use of the first-order Löwdin iterations (A1) in conjunction with the dynamical determination of the LOC regions described in Sec. IV.

[1] J. P. Lu and W. Yang, Phys. Rev. B **49**, 11 421 (1994).
[2] P. Ordejón, D. A. Drabold, R. M. Martin, and S. Itoh, Phys. Rev. Lett. **75**, 1324 (1995).
[3] S. Itoh, P. Ordejón, D. A. Drabold, and R. M. Martin, Phys. Rev. B **53**, 2132 (1996); S. Itoh, P. Ordejón, D. A. Drabold, R. M. Martin, and S. Ihara, Sci. Rep. Res. Inst. Tohoku Univ. A **41**, 163 (1996).
[4] W. Kohn, Int. J. Quantum Chem. **56**, 229 (1995).
[5] W. Kohn, Phys. Rev. **115**, 809 (1959).
[6] E. I. Blount, in *Solid State Physics*, edited by F. Seitz and C. Turnbull (Academic, New York, 1962), Vol. 13, p. 305.
[7] V. Heine, D. W. Bullett, R. Haydock, and M. J. Kelly, in *Solid State Physics*, edited by F. Seitz, C. Turnbull, and H. Ehrenreich (Academic, New York, 1980), Vol. 35.
[8] L. R. Mead and N. Papanicolaou, J. Math. Phys. **25**, 2404 (1984); R. H. Brown and A. E. Carlsson, Phys. Rev. B **32**, 6125 (1985).
[9] S. Baroni and P. Giannozzi, Europhys. Lett. **17**, 547 (1992).
[10] M. Krajči, J. Phys. F **17**, 2217 (1987); Š. Varga, J. Phys. Condens. Matter **2**, 8303 (1990).
[11] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer, Dordrecht, 1989), p. 455; D. A. Drabold and O. F. Sankey, Phys. Rev. Lett. **70**, 3631 (1993).
[12] U. Stephan, Ph.D. thesis, Technische Universität Chemnitz-Zwickau, 1995, available under http://archiv.tu-chemnitz.de/pub/1996/0007/.
[13] O. F. Sankey and R. E. Allen, Phys. Rev. B **33**, 7164 (1986); A. Demkov, D. A. Drabold, and O. F. Sankey (unpublished).
[14] M. S. Daw, Phys. Rev. B **47**, 10 895 (1993).
[15] X.-P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10 891 (1993).
[16] R. W. Nunes and D. Vanderbilt, Phys. Rev. B **50**, 17 611 (1994).
[17] A. E. Carlsson, Phys. Rev. B **51**, 13 935 (1995).
[18] L. W. Wang and M. P. Teter, Phys. Rev. B **46**, 12 798 (1992).
[19] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993); F. Mauri and G. Galli, *ibid.* **50**, 4316 (1994).
[20] P. Ordejón, D. A. Drabold, M. P. Grumbach, and R. M. Martin, Phys. Rev. B **48**, 14 646 (1993); P. Ordejón, D. A. Drabold, R. M. Martin, and M. P. Grumbach, *ibid.* **51**, 1456 (1995).
[21] E. B. Stechel, A. R. Williams, and P. J. Feibelman, Phys. Rev. B **49**, 10 088 (1994).
[22] W. Hierse and E. B. Stechel, Phys. Rev. B **50**, 17 811 (1994).
[23] G. Galli and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992).
[24] O. F. Sankey, D. A. Drabold, and A. Gibson, Phys. Rev. B **50**, 1376 (1994).
[25] S. Goedecker and L. Colombo, Phys. Rev. Lett. **73**, 122 (1994); S. Goedecker and M. Teter, Phys. Rev. B **51**, 9455 (1995).
[26] U. Stephan, D. A. Drabold, and P. Ordejón (unpublished).
[27] R. D. King-Smith and D. Vanderbilt Phys. Rev. B **47**, 1651 (1993); D. Vanderbilt and R. D. King-Smith, *ibid.* **48**, 4442 (1993); R. W. Nunes and D. Vanderbilt, Phys. Rev. Lett. **73**, 712 (1994).
[28] P. Fernández, A. Dal Corso, A. Baldereschi, and F. Mauri, Phys. Rev. B **55**, R1909 (1997).
[29] L. E. Ballentine, and M. Kolář J. Phys. C **19**, 981 (1986); see also A. P. Sutton, M. W. Finnis, D. G. Pettifor, and Y. Ohta, *ibid.* **21**, 35 (1988).
[30] A. Gibson, R. Haydock, and J. P. LaFemina, Phys. Rev. B **47**, 9229 (1993).
[31] P. Blaudeck, T. Frauenheim, D. Porezag, G. Seifert, and E. Fromm, J. Phys. Condens. Matter **4**, 6389 (1992); U. Stephan, T. Frauenheim, P. Blaudeck, and G. Jungnickel, Phys. Rev. B **50**, 1489 (1994).
[32] R. N. Silver, H. Röder, A. F. Voter, and J. D. Kress, J. Comput. Phys. **124**, 115 (1996).
[33] S. Goedecker, J. Comput. Phys. **118**, 261 (1995).
[34] J. D. Weeks, P. W. Anderson, and A. G. H. Davidson, J. Chem. Phys. **58**, 1388 (1973).
[35] O. F. Sankey and D. J. Niklewski, Phys. Rev. B **40**, 3979 (1989).
[36] G. Seifert, L. Bryja, and P. Ziesche, *Proceedings of the 19th International Conference on Electronic Structure of Solids*, edited by P. Ziesche (Technical University, Dresden, 1989), p. 220; G. Seifert and R. O. Jones, Z. Phys. D **20**, 77 (1991).

[37] A. F. Voter, J. D. Kress, and R. N. Silver, Phys. Rev. B **53**, 12 733 (1996).

[38] All computations were done on a workstation DEC Alpha 2000 4/275. The maximum memory space needed for computing all untruncated Wannier functions in the 512-atom models (Sec. IV) was 23 Mb.

[39] B. R. Djordjevic, M. F. Thorpe, and F. Wooten, Phys. Rev. B **52**, 5685 (1995). This reference presents and discusses the large 4096-atom model. Results for the smaller 512-atom model are unpublished.

[40] P. W. Anderson, Phys. Rev. Lett. **21**, 13 (1968).

[41] J. Des Cloizeaux, Phys. Rev. **135**, A685 (1964); **135**, A698 (1964).

[42] W. Kohn and J. R. Onffroy, Phys. Rev. B **8**, 2485 (1973).

[43] J. J. Rehr and W. Kohn, Phys. Rev. B **10**, 448 (1974).

[44] This second term alone, however, in general gives a poorer estimation of $\epsilon_1$ compared with Eq. (17) because it does not take into account the position of the significant eigenstates close to $E_F$. In Eq. (18), on the other hand, we utilize the fact that differences with respect to the same approximation are often much more accurate than the single terms alone.

[45] See, e.g., R. W. Hamming, *Numerical Methods for Scientists and Engineers*, 2nd. ed. (Dover, New York, 1986), p. 479.

[46] J. Harris, Phys. Rev. B **31**, 1770 (1985); W. M. C. Foulkes and R. Haydock, *ibid.* **39**, 12 520 (1989).

[47] J. Des Cloizeaux, Phys. Rev. **129**, 554 (1963).

[48] W. Kohn, Phys. Rev. B **7**, 4388 (1973).

[49] For the parameters discussed in Sec. III we chose the values $\beta = 50$ hartree$^{-1}$, $N_{Ch} = 100$, and $R_{\bar{H}} = 6.1a_0$.

[50] J. Kim, F. Mauri, and G. Galli, Phys. Rev. B **52**, 1640 (1995).

[51] D. A. Drabold and P. Ordejón (unpublished).

[52] A. Nenciu and G. Nenciu, Phys. Rev. B **47**, 10 112 (1993).

[53] W. Kohn Chem. Phys. Lett. **208**, 167 (1993).