JOURNAL OF COMPUTATIONAL PHYSICS 118, 261–268 (1995)

Low Complexity Algorithms for Electronic Structure Calculations

S. GOEDECKER

Cornell Theory Center, Ithaca, New York 14853-3801

Received November 8, 1993; revised October 14, 1994

Electronic structure calculations which are based on Wannier, like localized orbitals or the related density matrix, are an alternative to conventional calculations based on extended orbitals. For large systems this approach is potentially faster since it offers $O(N)$ scaling with respect to the number of atoms in the system. We derive a class of algorithms based on projection to calculate either the localized orbitals or the density matrix. © 1995 Academic Press, Inc.

1. INTRODUCTION

The complexity of an algorithm describes its scaling with respect to the size of the system N. If the number of operations necessary to do a certain calculation is given by cN^κ we speak of a low complexity algorithm if the exponent κ is small and of a high complexity algorithm if the exponent κ is large. It is an empirical observation that the prefactor c is in general much larger for low complexity algorithms. If there are two algorithms available for a certain problem, one of low complexity and with a high prefactor, the other of high complexity but with a low prefactor, there is therefore in general a crossover point where both algorithms require the same computer resources. If we go to larger systems, the low complexity algorithm will be faster; if we go to smaller systems the high complexity algorithm will be faster. In other words, because of larger prefactors, low complexity algorithms are only competitive if we have powerful computers at our disposal, which allow us to treat large systems. Computers become faster at a rapid rate and we therefore have to replace high complexity algorithms by low complexity algorithms. This trend can be observed in all domains of computing. Also, in electronic structure calculations some of the most important recent progress was related to the use of new lower complexity algorithms. In the Car Parrinello method [4] or conjugate gradient method [5–7], for instance, the numerical effort to calculate the matrix times the vector product of the Hamiltonian with an orbital, which is a sum of n plane waves, is not proportional to n^2 as it would be for an ordinary matrix times vector multiplication, but is proportional only to $n \log_2(n)$. Even the best available electronic structure calculation methods [4–7] still have a scaling which is cubic with respect to the number of electrons. This cubic scaling comes from the requirement that the orbitals be orthogonal. In the case of Gram–Schmidt orthogonalization, for instance, each orbital has to be orthogonalized to all the previous ones. As the system gets larger the number of the orbital increases and each orbital extends over a larger volume, requiring more basis set functions and therefore longer vectors, resulting altogether in a cubic operation count. The cubic terms arising from nonlocal pseudopotentials have recently been eliminated [8, 9]. Some proposals have recently been made for algorithms with linear scaling [1–3, 10–13]. We will derive a new family of algorithms with linear scaling based on projection, examine in depth the relation between the physics of the electronic system and the applicability of linear algorithms to it, and try to estimate the crossover points. To perform the analysis some new tools are required, such as the calculation of density matrices at finite temperature through complex contour integration.

Extended orbitals (which are eigenfunctions of the one-particle Hamiltonian) are the basic quantity in conventional electronic structure calculations. As the system grows, both the number of extended orbitals and the volume over which each orbital extends increase. If one could avoid the orthogonalization step, the best possible scaling which one could obtain with extended orbitals would therefore be quadratic. To obtain a linear scaling, the extended orbitals have to be replaced by the density matrix, whose physical behavior can be exploited to obtain a fast algorithm. This last point is essential. Mathematical and numerical analyses alone are not sufficient to construct a linear algorithm. They have to be combined with physical intuition. The loss of extended orbitals as the central quantity does not imply a loss of information about the system. In non-selfconsistent methods the basic quantity is the sum over all occupied eigenvalues. In the selfconsistent Kohn sham density functional theory [14] the two basic quantities are the sum of all the occupied Kohn sham eigenvalues and the charge density. These two quantities can be calculated directly from quantities related to the density matrix. It should also be pointed out that the absence of extended orbitals does not reduce the chemical understanding of big systems. Even though extended orbitals certainly give insight into chemical bonding of small molecules, looking at 1000 extended orbitals of a 100 atom system does not convey any useful information. It is, of course, also true that some selected orbitals (HOMO, LUMO) give useful information, but the calculation of a few orbitals is a relatively easy task.

0021-9991/95 $6.00

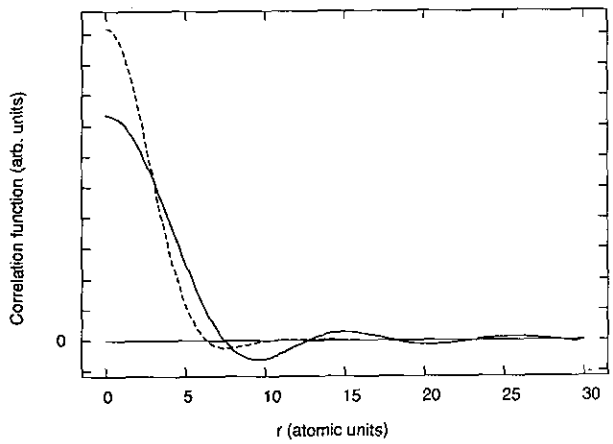

FIG. 1. The correlation function $F(0, r)$ for jellium with density $r_S = 2$. The solid line shows the zero temperature result, the dashed line the result at $T = 10^{-1}$ Hartree. The room temperature result ($T \approx 10^{-3}$ Hartree) is indistinguishable from the zero temperature result on the scale of this plot.

2. PHYSICAL PROPERTIES OF THE DENSITY MATRIX

Let us now briefly discuss the physical properties of the density matrix which are relevant for our numerical analysis. The (one-particle) density matrix is given by

$$F_{\mu,T}(\mathbf{r}, \mathbf{r}') = \sum_i \Psi_i(\mathbf{r}) f\left(\frac{\varepsilon_i - \mu}{kT}\right) \Psi_i^*(\mathbf{r}'), \qquad (1)$$

where $\Psi_i(\mathbf{r})$ and ε_i are the eigenfunctions and eigenvectors of an effective one-particle Hamiltonian and $f(x) = 1/(1 + e^x)$ denotes the Fermi distribution. The density matrix can be interpreted as a kind of correlation function which describes the importance of correlation between electronic properties at $\mathbf{r}$ and $\mathbf{r}'$. Electronic properties at $\mathbf{r}$ are fairly independent of the potential at $\mathbf{r}'$ if the distance between $\mathbf{r}$ and $\mathbf{r}'$ is greater than the correlation length, which equals the localization length of $F_{\mu,T}(\mathbf{r}, \mathbf{r}')$. Calculations which we performed illustrate how the volume of the localization region depends on two parameters. For metallic systems the localization volume decreases with increasing temperature, and for insulators it decreases with increasing bandgap and temperature. For jellium [15], the result can be calculated analytically at zero temperature. In this case $F_{\mu,T}$ evidently depends only on the difference $|\mathbf{r} - \mathbf{r}'|$.

$$F_\mu(r) = \frac{4\pi}{r^3}(\sin(k_F r) - (k_F r)\cos(k_F r)),$$

where k_F is the Fermi wave vector. Even though $F_{\mu T}$ decays, the decay is only algebraic and very slow. As shown in Fig. 1, higher temperatures improve the situation somewhat, but unless one goes to unphysically high temperatures the localization is rather poor. In Fig. 2 the decay properties of $F_{\mu,T}$ for some model insulators are shown. Even in this case the localization region is rather big and comprises several dozens of atoms.

The fact that a linear algorithm can be obtained for a quantum Coulomb system might seem surprising. For a classical system it is already a difficult task to construct such an algorithm since one cannot simply cut off the long-range Coulomb potential at an arbitrarily large radius. In the case of a metal the long-range Coulomb coupling enters through the interaction between the Fermi level and the Hartree potential. In a selfconsistent calculation a selfconsistent Hartree potential is found by a process called mixing [16], which is not discussed in this paper, and which might well turn out to be very difficult for large metallic systems, reflecting the difficulties in the corresponding classical system. In the case of an insulator, however, there is strong local charge neutrality and determining the selfconsistent potential is rather easy.

3. DIRECT METHODS FOR THE CHARGE DENSITY

Direct methods calculate the charge density directly on each grid point without using localized orbitals as intermediate quantities. Introducing the operator $F_{\mu,T}$,

$$F_{\mu,T} = f\left(\frac{H - \mu}{kT}\right),$$

which is diagonal in the eigenfunction representation. Equation (1) can be written in Diracs Bra Ket notation [17] as

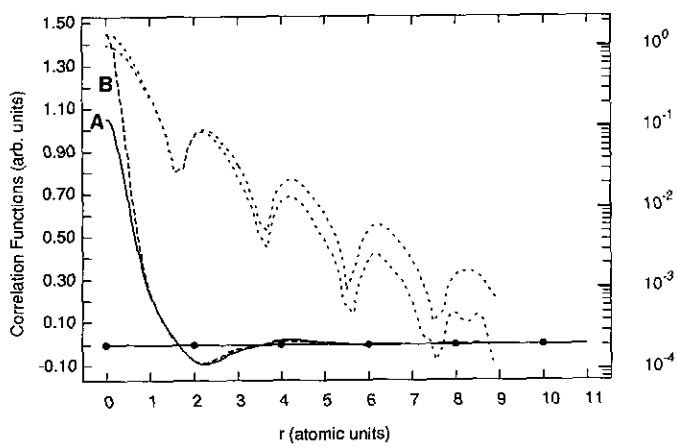

FIG. 2. The correlation function $F(0, x)$ for two simple cubic model insulators denoted A (solid line) and B (dashed line) along the bonding direction x. Insulator A has a valence band width of 1.7 Hartree and a band gap of 0.6 Hartree, insulator B a valence band width of 1.2 Hartree and a band gap of 1.5 Hartree. The two dotted lines, keyed to the right axis, are logarithmic plots of the two functions. It shows the exponential decay of the correlation function which is faster for the insulator with a wider bandgap (B). The region near the zeroes is cut out in the logarithmic plot. The positions of the atoms are denoted by solid circles.

$$F_{\mu,T}(\mathbf{r},\mathbf{r}') = \sum_{i,j} \langle \mathbf{r}|\Psi_i\rangle\langle\Psi_i|F_{\mu,T}|\Psi_j\rangle\langle\Psi_j|\mathbf{r}'\rangle. \quad (2)$$

The charge density is given by the diagonal elements of the density matrix.

$$\begin{aligned}\rho(\mathbf{r}) &= 2\sum_i \Psi_i(\mathbf{r})^* f\left(\frac{H-\mu}{kT}\right)\Psi_i(\mathbf{r}) \\ &= 2\sum_{i,j}\langle\mathbf{r}|\Psi_i\rangle\langle\Psi_i|F_{\mu,T}|\Psi_j\rangle\langle\Psi_j|\mathbf{r}\rangle.\end{aligned} \quad (3)$$

The factor of two in front comes from the fact that each orbital is occupied by two electrons.

Since Eqs. (2) and (3) are invariant under unitary transformations, we can go from the set of eigenfunctions to any other orthogonal complete set such as the set of delta functions located at **R**. In this representation the density matrix is given by

$$\begin{aligned}F_{\mu,T}(\mathbf{r},\mathbf{r}') &= \int d\mathbf{R}\int d\mathbf{R}'\langle\mathbf{r}|\mathbf{R}\rangle\langle\mathbf{R}|F_{\mu,T}|\mathbf{R}'\rangle\langle\mathbf{R}'|\mathbf{r}'\rangle \\ &= \langle\mathbf{r}|F_{\mu,T}|\mathbf{r}'\rangle,\end{aligned} \quad (4)$$

as it should be. Baroni [3] *et al.* used finite differences, which can be considered to be numerical approximations to delta functions, to calculate the charge density at zero temperature. In a selfconsistent density functional calculation the charge density has to be calculated on a very dense grid and the calculation for a single point requires a heavy computation. As he himself [3] remarked, a method which takes the smoothness of the charge density into account would be desirable to reduce the prefactor. One can profit by the smoothness if one uses localized but continuous functions as the basis set. The use of continuous basis functions requires, however, the knowledge of off diagonal elements of the density matrix. These off-diagonal elements can unfortunately not be calculated with the recursion method, which was used by Baroni and others. In the following a method which allows us to calculate these off-diagonal elements is presented.

Let us introduce a set of localized continuous atomic-like basis functions $u_{\mathbf{R}}(\mathbf{r}) = \langle\mathbf{r}|u(\mathbf{R})\rangle$. The index **R** denotes the location of the basis function, but it can as well include other properties such as angular momentum. Let us first assume that they are orthogonal and the identity operator is therefore given by

$$I = \sum_{\mathbf{R}} |u(\mathbf{R})\rangle\langle u(\mathbf{R})|.$$

For the charge density we therefore obtain

$$\rho(\mathbf{r}) = \sum_{\mathbf{R},\mathbf{R}'}\langle\mathbf{r}|u(\mathbf{R})\rangle\langle u(\mathbf{R})|F_{\mu,T}|u(\mathbf{R}')\rangle\langle u(\mathbf{R}')|\mathbf{r}\rangle. \quad (5)$$

If the matrix $F_{\mu,T}$ were known, the charge density could be calculated by several simple matrix times vector multiplications. Since the basis functions $\langle\mathbf{r}|u(\mathbf{R})\rangle$ are localized, only a small number of **R**, **R**′ need be included in the summation of Eq. (5) for one point. Evidently, the density matrix is also in this representation localized and it is not necessary to calculate it for the whole system when evaluating the charge density at one point. The method is therefore linear.

It remains to be shown how to calculate the matrix $\langle u(\mathbf{R})|F_{\mu,T}|u(\mathbf{R}')\rangle$. A method which only requires one to find the inverse of the Hamiltonian matrix, or equivalently the solution of linear systems of equations, is contour integration. As has been shown [18], $F_{\mu,T}$ has an integral representation:

$$F_{\mu,T} = \int dz \frac{1}{H-z}.$$

The number of integration points needed to approximate the above integral with a precision of four digits is of the order of 10. The charge density is therefore given by

$$\rho(\mathbf{r}) = \int dz \sum_{\mathbf{R},\mathbf{R}'}\langle\mathbf{r}|u(\mathbf{R})\rangle\left\langle u(\mathbf{R})\left|\frac{1}{H-z}\right|u(\mathbf{R}')\right\rangle\langle u(\mathbf{R}')|\mathbf{r}\rangle. \quad (6)$$

Localized orthogonal orbitals are difficult to construct, because the orthogonality requirement necessitates slowly decaying oscillations over distances of several atoms. Nonorthogonal orbitals are much better localized and therefore recommended in this context. If the orbitals are not orthogonal the overlap matrix $S_{\mathbf{R},\mathbf{R}'} = \langle u(\mathbf{R})|u(\mathbf{R}')\rangle$ comes into play and Eq. (6) becomes (see Appendix 1)

$$\rho(\mathbf{r}) = \oint dz \sum_{\mathbf{R},\mathbf{R}'}\langle\mathbf{r}|u(\mathbf{R})\rangle(H-Sz)^{-1}_{\mathbf{R},\mathbf{R}'}\langle u(\mathbf{R}')|\mathbf{r}\rangle. \quad (7)$$

4. DIRECT METHODS FOR THE EIGENVALUE SUM

A central quantity in both selfconsistent and non-selfconsistent calculations, such as calculations based on the Harris functional [19] or on the tight binding method, is the sum E over all eigenvalues:

$$E = \sum_i \varepsilon_i f\left(\frac{\varepsilon_i-\mu}{kT}\right). \quad (8)$$

Introducing the operator $G_{\mu,T}$ which is again diagonal in the eigenfunction representation

$$G_{\mu,T} = Hf\left(\frac{H-\mu}{kT}\right),$$

we can express E as the trace of $G_{\mu,T}$,

TABLE I

Comparison of CPU Times Required to Calculate the Sum over All Occupied Eigenvalues for a System of n Atoms, Treated with the Tight Binding Method

Number of atoms, n:	64	216	512	1000	1700	2744
Time with LAPACK:	5 sec	2.7 min	34 min	180 min	No timing	No timing
Time with linear method:	2.6 sec	32 sec	1.8 min	3.8 min	7.1 min	11.7 min

$$E = \sum_i \langle \Psi_i | G_{\mu,T} | \Psi_i \rangle = \mathrm{Tr}[G_{\mu,T}].$$

Let us now discuss how to calculate the trace. In addition to the contour integration method discussed above there are two other methods which are promising.

(a) *Polynomial Approximation*

Let us denote the extremal eigenvalues of H by $\varepsilon_{\max}$ and $\varepsilon_{\min}$. A polynomial $p_{\mu,T}$ can be constructed (e.g., Chebyshev expansion) which approximates the Fermi distribution between $\varepsilon_{\min}$ and $\varepsilon_{\max}$. To get a good approximation the order M of the polynomial has to be of the order of $(\varepsilon_{\max} - \varepsilon_{\min})/(kT)$. If the temperature is high or if, in the case of an insulator, it can be chosen to be high, M can be reasonably small and we have a very simple and efficient method. $G_{\mu,T}$ is then given by $G_{\mu,T} = Hp_{\mu,T}(H)$. Let us now assume that we have a set of localized orthogonal basis functions $u(\mathbf{R})$ where $\mathbf{R}$ denotes the position of the localized orbital, as well as other indices such as angular momentum if necessary. To calculate the trace we do not explicitly construct $G_{\mu,T}$ but we calculate each diagonal element $\langle u(\mathbf{R})|G_{\mu,T}|u(\mathbf{R})\rangle$ by first calculating $G_{\mu,T}|u(\mathbf{R})\rangle$ recursively by simple matrix multiplication with H and then take the scalar product with $\langle u(\mathbf{R})|$. Since the matrix $G_{\mu,T}$ is also sparse the method is again linear.

The above described method was applied to supercells of carbon atoms in the diamond equilibrium geometry. The tight binding method [20] was used as the electronic structure method. In this scheme a window $(\varepsilon_{\max} - \varepsilon_{\min})$ of 50 eV is sufficient. Since carbon is an insulator with a band gap of 4 eV we can choose a high electronic temperature of 1 eV and the resulting polynomial is of degree 50. Table I shows the timing results for an accuracy of 10 meV per atom. The linear scaling is clearly visible and dramatic speedups were obtained for large clusters. But also for relatively small clusters of 64 atoms the method was faster than exact diagonalization using Lapack [21] routines. The tests were performed on an IBM RS6000/550 workstation and a precision of 1 meV per atom was required.

The program was also parallelized using PVM [22]. Parallelization can be easily achieved, since each column of the matrix $G_{\mu,T}$ can be calculated independently. On a cluster of eight IBM RS6000 workstations we obtained a speedup of 6.4 compared to the serial version for the 1000 atom system.

(b) *Haydock's Method*

As has been shown by Haydock [23], diagonal elements of the Green function can easily be calculated by the recursion scheme:

$$\left\langle u_0(\mathbf{R}) \left| \frac{1}{H - z} \right| u_0(\mathbf{R}) \right\rangle = 1 \Big/ \left(z - a_0 - b_1^2 \Big/ \left(z - a_1 - \frac{b_2^2}{z - a_2 - \cdots} \right) \right),$$

where the coefficients a_i and b_i are defined by the recursion

$$b_1 \mathbf{u}_1 = H\mathbf{u}_0 - a_0 \mathbf{u}_1$$

$$b_{n+1}\mathbf{u}_{n+1} = H\mathbf{u}_n - a_n \mathbf{u}_n - b_n \mathbf{u}_{n-1}.$$

a_n is chosen such that $\mathbf{u}_{n+1}$ is orthogonal to $\mathbf{u}_n$ and b_{n+1} normalizes $\mathbf{u}_{n+1}$ to unity. One diagonal element of the matrix $G_{\mu,T}$ can therefore easily be calculated:

$$\langle u_0(\mathbf{R})|G_{\mu,T}|u_0(\mathbf{R})\rangle = \oint \left\langle u_0(\mathbf{R}) \left| \frac{H}{H - z} \right| u_0(\mathbf{R}) \right\rangle dz = \oint z \Big/ \left(z - a_0 - b_1^2 \Big/ \left(z - a_1 - \frac{b_2^2}{z - a_2 - \cdots} \right) \right) dz.$$

As has been shown in [18], the integration path can either represent zero or finite temperature conditions. If the Hamiltonian matrix is expressed with respect to a localized basis set, the localized function u_0 will spread over roughly one additional shell of neighbors in each step of the recursion scheme. If a good terminator can be found only a small number M of iterations is required to obtain sufficient accuracy, and the function u_M will be spread out only over a region which is small compared to the total volume of a big system, and the method is therefore linear. The generalization to nonorthogonal basis functions is also given in Haydock's paper [23].

5. DIRECT METHODS FOR THE MOMENTS OF THE DENSITY OF STATES

The density of states can be reconstructed from its moments M_ν, where

$$M_\nu = \sum_i \varepsilon_i^\nu. \tag{9}$$

This can be done by either the Pade or the maximum entropy method [24]. For reasons of numerical stability, it is preferable not to use the moments from Eq. (9) but, instead, moments with respect to well behaved polynomials such as the Legendre polynomials L_ν. We will assume in the following that the Hamiltonian matrix was scaled in such a way that all eigenvalues are in the interval $[-1; 1]$. The numerically stable moments are then

$$M_\nu = \sum_i L_\nu(\varepsilon_i). \tag{10}$$

Let us now expand the density of states $D(\varepsilon)$ in Legendre polynomials.

$$D(\varepsilon) = \sum_i c_\nu L_\nu(\varepsilon). \tag{11}$$

The expansion coefficients c_ν are then given by [25]

$$c_\nu = \frac{2\nu + 1}{2} \int_{-1}^{1} D(\varepsilon) L_\nu(\varepsilon)\, d\varepsilon. \tag{12}$$

Using the properties of the trace and the definition of the density of states, we obtain

$$\mathrm{Tr}[L_\nu(H)] = \sum_i L_\nu(\varepsilon_i) \approx \int_{-1}^{1} D(\varepsilon) L_\nu(\varepsilon)\, d\varepsilon. \tag{13}$$

This means that the moments M_ν are the expansion coefficients of the density of states and they can be calculated by taking the trace of the Legendre polynomial of the Hamiltonian matrix. The approximate equal sign in Eq. (13) comes from the fact that the sum can be replaced by the integral only in the limit of an infinitely large system. Since we are, however, interested in very large systems, the approximation is very good. The matrix $L_\nu(H)$ is much less diagonally dominant than the matrices $F_{\mu,T}$ and $G_{\mu,T}$. We have therefore in the present implementation not neglected any matrix elements and the method is therefore quadratic. If one needs, however, only moderate accuracy one could neglect these elements, and the method then becomes linear. The method evidently also scales linearly if one needs to calculate only the local density of states for one or a few selected atoms such as in the case of an impurity in a large system. Even in the quadratic implementation the calculation of the first 100 moments takes much less time than a classical diagonalization. In Table II we show the timings for conventional diagonalization with LAPACK and for our method when the two methods are applied to clusters of silicon atoms treated with the tight binding method [26]. The test was again done on an IBM RS6000/550 workstation.

In Fig. 3 we show the resulting density of states as given by Eq. (11). No attempt was made to improve it with the above-mentioned Pade or maximum entropy methods.

6. LOCALIZED ORBITALS IN INSULATORS AT ZERO TEMPERATURE

The function $F_\mu(\mathbf{r}, \mathbf{r}')$ actually contains not only information about the charge density at the point $\mathbf{r} = \mathbf{r}'$ but about a whole region around this point. At zero temperature, the density matrix of an insulator is an idempotent projection operator, which has only n_{el} eigenvalues which are different from zero, where n_{el} is the number of electrons. The space $\mathrm{Sp} = \{\Psi_1, \Psi_2, ..., \Psi_{n_{el}}\}$ spanned by the density matrix can therefore be generated by applying the density matrix to n_{el} linearly independent functions $\phi_1, \phi_2, ..., \phi_{n_{el}}$ whose overlap with Sp is nonzero. The functions $\langle \mathbf{r} | \Phi_i \rangle = \langle \mathbf{r} | F_\mu | \phi_i \rangle$ which represent the projection of the original set ϕ_i onto the space of eigenfunctions Sp can be calculated by contour integration, which numerically is approximated by a sum. Since this method does not require any special set of localized basis functions, all operators are given in the $\mathbf{r}$ representation.

$$\begin{aligned} \langle \mathbf{r} | \Phi_i \rangle = \langle \mathbf{r} | F_\mu | \phi_i \rangle &= \oint dz \int d\mathbf{r}' \left\langle \mathbf{r} \left| \frac{1}{H - z} \right| \mathbf{r}' \right\rangle \langle \mathbf{r}' | \phi_i \rangle \\ &= \oint dz \, \langle \mathbf{r} | \chi_z^i \rangle, \end{aligned} \tag{14}$$

where we denoted $\int d\mathbf{r}' \langle \mathbf{r} | 1/(H - z) | \mathbf{r}' \rangle \langle \mathbf{r}' | \phi_i \rangle$ by $\langle \mathbf{r} | \chi_z \rangle$. The function $\langle \mathbf{r} | \chi_z^i \rangle$ is a solution of the linear system of equations

$$\int d\mathbf{r}' \, \langle \mathbf{r} | H - z | \mathbf{r}' \rangle \langle \mathbf{r}' | \chi_z^i \rangle = \langle \mathbf{r} | \phi_i \rangle. \tag{15}$$

A set of orthonormal orbitals $\psi_1, \psi_2, ..., \psi_{n_{el}}$ is then given by

$$\psi_j(\mathbf{r}) = \sum_i S_{j,i}^{-1/2} \Phi_i(\mathbf{r}),$$

where $S_{j,i} = \langle \Phi_i | \Phi_j \rangle$. The charge density is given by

$$\rho(\mathbf{r}) = \sum_{j=1}^{n_{el}} \psi_j^*(\mathbf{r}) \psi_j(\mathbf{r}) = \sum_{i,j} \Phi_i^*(\mathbf{r}) S_{i,j}^{-1} \Phi_j(\mathbf{r}). \tag{16}$$

Equation (16) necessitates the calculation of the overlap matrix S (which is now an overlap matrix between orbital instead of basis functions) which would result in a cubic scaling for ex-

TABLE II

Comparison of CPU Times Required to Calculate the First 100 Moments of the Density of States for a System of n Atoms, Treated with the Tight Binding Method

Number of atoms, n:	64	216	512	1000	1728	2744
Time with LAPACK:	5 sec	2.7 min	34 min	180 min	No	No
Time with quadratic method:	5.4 sec	1.1 min	6.6 min	26 min	72 min	180 min

tended orbitals Ψ_i. If, however, the orbitals are localized, S is sparse and both the calculation and inversion of S scale linearly. Even if the inversion of S is not done with a method scaling linearly (polynomial inversion, conjugate gradient methods) this part of the calculation is negligibly small, and no overall cubic behaviour could be seen for systems of reasonable size. The numerically expensive part is the repeated solution of the linear system of Eq. (15). In the context of this work the preconditioned biconjugate gradient method [27] together with fast Fourier techniques for matrix multiplication [28, 29] was used in combination with plane waves as the basis set. The preconditioned biconjugate gradient method would probably also be the most efficient method if localized continuous basis functions such as gaussians are used. It is interesting to note that within this formulation the electronic structure problem is reduced to the solution of linear systems of equations for a set of complex energies z, whereas conventional formulations lead to eigenvalue problems.

In order to obtain well localized orbitals $\langle \mathbf{r} | \Phi_i \rangle$ a clever choice for $\langle \mathbf{r} | \phi_i \rangle$ has to be made. In this way it is possible to obtain orbitals whose localization is better than the one of $F_{\mu,T}(\mathbf{r}, \mathbf{r}')$. Let us consider the analytically treatable case of the free electron gas. The density matrix is given by

$$F_{\mu,T}(\mathbf{r}, \mathbf{r}') = \frac{1}{V} \int_{\mathrm{FS}} d\mathbf{k}\, e^{-i\mathbf{k}\mathbf{r}} e^{i\mathbf{k}\mathbf{r}'},$$

where FS denotes the Fermi sphere and V the volume. Applying the density matrix to an orbital ϕ gives

$$\Phi(\mathbf{r}) = \int d\mathbf{r}'\, F_{\mu,T}(\mathbf{r}, \mathbf{r}')\phi(\mathbf{r}') = \int_{\mathrm{FS}} d\mathbf{k}\, e^{i\mathbf{k}\mathbf{r}} \int d\mathbf{r}'\, \frac{1}{V}\, \phi(\mathbf{r}') e^{i\mathbf{k}\mathbf{r}'}$$

$$= \int_{\mathrm{FS}} d\mathbf{k}\, e^{i\mathbf{k}\mathbf{r}} \phi(\mathbf{k}).$$

If the Fourier transform $\phi(\mathbf{k})$ is sufficiently small for wavevectors larger than k_F, then $\Phi(\mathbf{r}) \approx \phi(\mathbf{r})$. If $\phi(\mathbf{r})$ were a broad Gaussian with a spatial extension greater than $2\pi/k_F$ the above conditions would be fulfilled. This means that for the free electron gas localized orbitals which have nearly the form of broad Gaussians can be chosen. It turns out that also in the case of an insulator the Gaussians give rather well localized orbitals. A pair $\Phi(\mathbf{r})$, $\phi(\mathbf{r})$ is shown in Fig. 4. From the above example we also see, however, that the asymptotic behaviour of a localized orbital cannot be changed. In the case of the free electron gas it is dictated by the discontinuity at the Fermi surface.

The above described method has been applied to linear molecules. Since in the present version of the program no semilocal pseudopotentials can be used, this molecule is not a highly accurate representation of a real molecule, but the local pseudo-

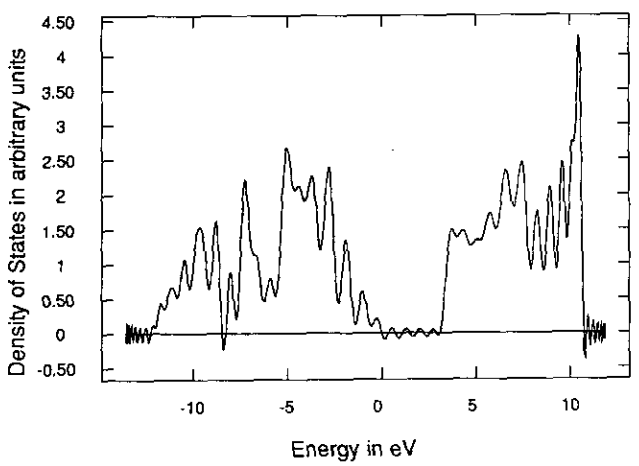

FIG. 3. The density of states of a cluster of 2744 silicon atoms as reconstructed from the first 100 moments.

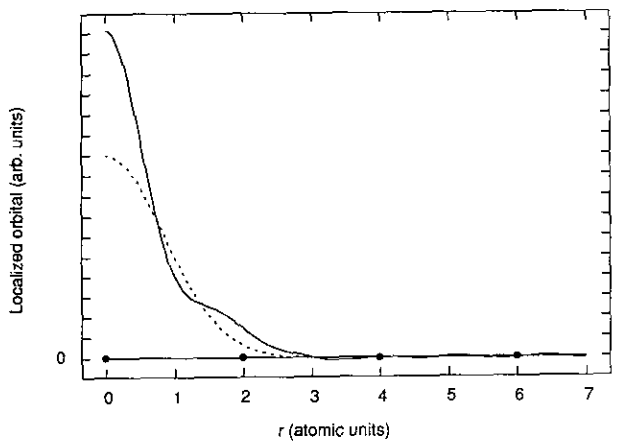

FIG. 4. A pair of $\Phi(\mathbf{r})$, $\phi(\mathbf{r})$ functions for model insulator A of Fig. 2. Note that the localized wavefunction is nearly zero for r greater than 3, whereas the correlation function vanishes beyond $r = 5$, roughly. The positions of the atoms are again denoted by solid circles.

TABLE III

Comparison of CPU Times Required to Calculate Eigenvalues and Charge Density with Iterative Diagonalization and the Linear Localized Orbital Method

Number of atoms, n:	10	20	30
Time with iter. diag.:	9 min	45 min	120 min
Time with linear method:	15 min	30 min	45 min

potential nevertheless represents the essential features of a silicon atom. We compared the new method with a state-of-the-art iterative block diagonalization method [30]. A localization length of 12 atomic units and an energy cutoff of 20 Hartrees were used. Eight integration points were required to evaluate the integral in Eq. (14) with an accuracy of 1 meV (see Table III).

Unfortunately the localized orbital method, unlike the direct methods described before, cannot be applied at nonzero temperatures and for systems having degenerate states at the Fermi level with fractional occupation, such as metals. In this case the density matrix has eigenvalues which are neither 0 nor 1 but are somewhere in the open interval]0, 1[and the density matrix is therefore not a projection operator. Unitary transformations of the orbitals cannot be done and localized orbitals can therefore not be constructed. This is physically intuitive. Electrical conductivity apparently requires extended orbitals at the Fermi level.

If necessary, the correlation function can easily be calculated from the localized orbitals. It is just the best approximation to a delta function which one can obtain within the space Sp of the occupied orbitals.

7. CONCLUSIONS

Algorithms with linear scaling exist for electronic structure calculations because the density matrix is localized. The localization is much better in the case of an insulator where the density matrix decays exponentially, whereas in metal the decay is only algebraic at zereo temperature.

Two classes of methods have been presented: direct methods which calculate all quantities, such as the charge density, directly without the intermediate step of orbitals and localized orbital methods.

Direct methods can be applied to a wide range of materials at any temperature. They require, however, localized basis sets such as gaussians, and of course they inherit in this way all the problems associated with these basis sets, such as basis set superposition errors.

In the present formulation, localized orbital methods jcan only be used at zero temperature for insulators. Their scaling is, however, not affected by the basis set. Plane waves whose periodicity volume equals the localization volume can for instance be used.

The prefactor for linear methods is unfortunately rather large. The linear behavior becomes visible only for systems larger than the localization volume. The radius at which one cuts off correlation effects depends of course on the precision one wants to obtain. It seems however that for bulk systems one needs at least a localization volume which contains on the order of 50 atoms to obtain reasonable precision. For linear molecules the method can however be faster for systems containing as few as 20 atoms.

Finally, it has to be pointed out that all these methods can be parallelized trivially since one localized orbital can be calculated independent of all the other ones, and they can thus fully take advantage of the emerging parallel supercomputers, aiming at teraflop rates.

APPENDIX 1

In Eq. (7) the notation $(H - Sz)^{-1}_{\mathbf{R},\mathbf{R}'}$ has been used to make clear that the operator $H - Sz$ is first expressed in the basis $|u(\mathbf{R})\rangle$ and only then inverted. In the case of nonorthogonal basis functions this is not the same as representing the operator $1/(H - Sz)$ in the basis $|u(\mathbf{R})\rangle$ which would be denoted by $\langle u(\mathbf{R})|1/(H - Sz)|u(\mathbf{R}')\rangle$. After this introductory remark let us now derive Eq. (7). Let $U_{\mathbf{R},i}$ be the matrix whose ith column contains the ith eigenvector of the generalized eigenvalue problem

$$\sum_{\mathbf{R}'} (H_{\mathbf{R},\mathbf{R}'} - \lambda_i S_{\mathbf{R},\mathbf{R}'}) U_{\mathbf{R}',i} = 0.$$

The ith orbital is then given by

$$\psi_i(\mathbf{r}) = \sum_{\mathbf{R}} U_{\mathbf{R},i} u_{\mathbf{R}}(\mathbf{r}).$$

Inverting the above equation gives

$$\mathbf{u}(\mathbf{r}) = U^{-1}\boldsymbol{\psi}(\mathbf{r}),$$

where $\mathbf{u}(\mathbf{r})$ denotes the vector whose components are $u_{\mathbf{R}}(\mathbf{r})$ and $\boldsymbol{\psi}(\mathbf{r})$ the vector whose components are $\psi_i(\mathbf{r})$. Equation (7) then becomes

$$\rho(\mathbf{r}) = \boldsymbol{\psi}(\mathbf{r})^H U^{-1}(H - Sz)^{-1}(U^{-1})^H \boldsymbol{\psi}(\mathbf{r}).$$

It remains now to show that $U^{-1}(H - Sz)^{-1}(U^{-1})^H$ is a diagonal matrix D^{-1} with matrix elements $\delta_{i,j}(\lambda_i - z)^{-1}$. This is, however, very easy. By elementary reversible matrix transformations we obtain

$$U^H(H - Sz)U = D,$$

where $D_{i,j} = \delta_{i,j}(\lambda_i - z)$. This expresses the well known fact that the eigenvectors of the generalized eigenvalue problem diagonalize both H and S.

ACKNOWLEDGMENTS

Financial support for this work came from the Swiss National Science Foundation. Additional support was provided by the U.S. National Science Foundation, the Cornell Theory Center, the Cornell Materials Science Center, and IBM. I thank Giovanni Bachelet and Klaus Maschke for their suggestions which made the paper more understandable, as well as Mike Teter, Stefano Baroni, David Vanderbilt, and Giulia Galli for interesting discussions. Luciano Colombo kindly gave me his program to calculate tight binding Hamiltonian matrices.

REFERENCES

1. W. Kohn, *Phys. Rev. B* **7,** 4388 (1973).
2. G. Galli and M. Parrinello, *Phys. Rev. Lett.* **69,** 3547 (1992).
3. S. Baroni and P. Gianozzi, *Europhys. Lett.* **17,** 547 (1992).
4. R. Car and M. Parrinello, *Phys. Rev. Lett.* **55,** 2471 (1985).
5. I. Stich, R. Car, M. Parrinello, and S. Baroni, *Phys. Rev. B* **39,** 4997 (1989).
6. M. P. Teter, M. C. Payne, and D. C. Allan, *Phys. Rev. B* **40,** 12255 (1989).
7. T. A. Arias and J. D. Joannopoulos, *Phys. Rev. Lett.* **73,** 680 (1994).
8. R. D. King-Smith, M. C. Payne, and J. S. Lin, *Phys. Rev. B* **44,** 13063 (1991).
9. M. Saito, O. Sugino, and A. Oshiyama, *Phys. Rev. B* **46,** 2606 (1992).
10. W. Yang, *Phys. Rev. Lett.* **66,** 1438 (1991).
11. X. P. Li, W. Nunes, and D. Vanderbilt, *Phys. Rev. B* **47,** 10891 (1993).
12. W.-L. Wang and M. Teter, *Phys. Rev. B* **46,** 12798 (1992).
13. P. Ordejon, D. Drabold, M. Grunbach, and R. Martin, *Phys. Rev. B* **48,** 14646 (1993).
14. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford Univ. Press, New York, 1989).
15. N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976).
16. M. Payne, M. Teter, D. Allan, T. Arias, and J. Joannopoulos, *Rev. Mod. Phys.* **64,** 1045, (1992); W. Pickett, *Comput. Phys. Rep.* **9,** 115 (1989).
17. E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970).
18. S. Goedecker, *Phys. Rev. B* **48,** 17573 (1993).
19. J. Harris, *Phys. Rev. B* **31,** 1770 (1985).
20. C. Xu, Z. Wang, C. Chan, and K. Ho, *J. Phys.: Condens. Matter* **4,** 6047 (1992).
21. Lapack User Guide (SIAM, Philadelphia, 1992).
22. *PVM 3 User Guide* (Oak Ridge National Laboratory, Oak Ridge, TN, 1993).
23. R. Haydock, *Solid State Phys.* **35,** 3547 (1980).
24. L. Mead and N. Papanicolaou, *J. Math. Phys.* **25,** 2404 (1984).
25. M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).
26. L. Goodwin, L. Skinner, and A. J. Pettifor, *Europhys. Lett.* **9,** 701 (1989).
27. W. Press and S. Teukolsky, *Comp. in Phys.* **6,** 400 (1992).
28. J. L. Martins and M. L. Cohen, *Phys. Rev. B* **37,** 6134 (1988).
29. S. Goedecker, *Comp. Phys. Comm.* **76,** 294 (1993).
30. E. Davidson, *Comp. Phys. Comm.* **53,** 49 (1989).