Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set

G. Kresse
Institut für Theoretische Physik, Technische Universität Wien, Wiedner Hauptstraße 8-10/136, A-1040 Wien, Austria

J. Furthmüller
Institut für Festkörpertheorie und Theoretische Optik, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, D-07743 Jena, Germany

(Received 13 December 1995; revised manuscript received 30 May 1996)

We present an efficient scheme for calculating the Kohn-Sham ground state of metallic systems using pseudopotentials and a plane-wave basis set. In the first part the application of Pulay's DIIS method (direct inversion in the iterative subspace) to the iterative diagonalization of large matrices will be discussed. Our approach is stable, reliable, and minimizes the number of order N_{atoms}^3 operations. In the second part, we will discuss an efficient mixing scheme also based on Pulay's scheme. A special ''metric'' and a special ''preconditioning'' optimized for a plane-wave basis set will be introduced. Scaling of the method will be discussed in detail for non-self-consistent and self-consistent calculations. It will be shown that the number of iterations required to obtain a specific precision is almost independent of the system size. Altogether an order N_{atoms}^2 scaling is found for systems containing up to 1000 electrons. If we take into account that the number of k points can be decreased linearly with the system size, the overall scaling can approach N_{atoms}. We have implemented these algorithms within a powerful package called VASP (Vienna *ab initio* simulation package). The program and the techniques have been used successfully for a large number of different systems (liquid and amorphous semiconductors, liquid simple and transition metals, metallic and semiconducting surfaces, phonons in simple metals, transition metals, and semiconductors) and turned out to be very reliable. [S0163-1829(96)00440-7]

I. INTRODUCTION

In the past few years *ab initio* calculations based on Kohn-Sham (KS) density functional theory[1] have gained an enormous interest not only among solid state physicists but also among chemists. In part, this is due to the great success of the local density approximation (LDA) (see, for instance, Ref. 2) which — although originally intended only for the application to solids — seems to be reasonably accurate for molecules and the adsorption of molecules on surfaces too. In addition, taking into account generalized gradient corrections has removed one of the most problematic deficiencies of the LDA, the strong overbinding of isolated molecules.

The biggest advantage of the KS density functional approach is definitely its simplicity. Forces, for instance, can be evaluated in principle and in practice using the well known Hellmann-Feynman theorem.[3] Although different basis sets can be used, plane waves in the broadest sense seem currently to be most advantageous. This basis set is complete and allows an easy analytical evaluation of the forces and of the stress tensor. Its biggest disadvantage is probably that the number of plane waves N_{plw} which must be included is usually an order of magnitude larger than the number of basis functions (centered at atomic sites) used for ''minimal'' numerical basis sets; but this is more than made up by the fact that the action of the Hamiltonian onto trial wave functions can be evaluated very efficiently. Using modern iterative algorithms the explicit calculation and storage of the $N_{\text{plw}} \times N_{\text{plw}}$ Hamilton matrix can be avoided, allowing the use of very large basis sets ($N_{\text{plw}} \approx 10\,000$) even on simple workstations. With the recent introduction of efficient pseudopotentials (see, e.g., Vanderbilt[4]) and the introduction of the projector-augmented-plane-wave method[5] the applicability of plane waves has increased even further. Now also transition metals and first row elements no longer pose any serious difficulty and can be treated almost as efficiently as conventional ''simple'' elements.

For the calculation of the KS ground state it is possible to distinguish two methods: (i) Methods for determining the minimum of the KS total-energy functional directly (in the future simply called direct methods) and (ii) iterative methods for the diagonalization of the KS Hamiltonian in conjunction with an iterative improvement (i.e., mixing) of the charge density or the potential [we will refer to these methods as self-consistency cycle (SC) methods].

Both methods require no explicit storage of the Hamilton matrix and should therefore show similar ''efficiency.'' The direct methods (i) have been pioneered by Car and Parrinello.[6] They are based on the fact that the Kohn-Sham energy functional is minimal at the electronic ground state. Therefore, minimization with respect to the variational degrees of freedom leads to a convenient scheme for calculating the electronic ground state. The only problem to be solved is the inclusion of the orthonormality constraints on the wave functions, which is done with a Lagrange formalism in the original work of CP. Generally the standard CP algorithm is relatively slow if it is applied to the electrons only. Small improvements might be obtained by integrating the equations of motion analytically,[7] by introducing an improved preconditioning for the gradient[8] or by replacing the second-order CP equations by first-order steepest-descent[9,10] equations. Recently Tassone, Mauri, and Car[8] showed that

 © 1996 The American Physical Society

the most efficient algorithm is a preconditioned damped second-order equation of motion for the electrons, a scheme first proposed in Ref. 7. It is interesting to point out that the damped second-order equation of motion used in Ref. 8 is closely related to an acceleration scheme for slowly converging series by Tchebycheff which has been used for the mixing of charge densities by Akai and Dederichs.[11] The only drawback of this method is that it requires the determination of two parameters to obtain fast convergence.

In this respect, conjugate gradient (CG) schemes[12,13] are more promising because they are entirely parameter free. Within these schemes the KS functional is minimized along a given search direction exactly (which is usually not done within the techniques mentioned above), and in successive steps the new search direction is conjugated to previous search directions. The main problem within the CG methods is that the orthonormality constraints are not easy to incorporate. For semiconductors and insulators Teter, Payne, and Allan[14] proposed a reliable algorithm which optimizes the total energy for a single orbital within the subspace orthonormal to the current set of trial wave functions. Despite the advantage of small storage requirements, the algorithm is relatively slow because only a limited number of CG steps per orbital can be done, and because the charge density and the potential must be recalculated after each single update of each orbital. Therefore, algorithms which update all orbitals simultaneously should be superior. These algorithms were pioneered independently by Stich, Car, Parrinello, and Baroni[15] and by Gillan.[16] The most systematic and elegant way to incorporate the orthonormality constraints in this case is to generalize the KS functional to nonorthogonal orbitals,[17] but successful applications of this approach are still rare.[18–20]

In contrast to the direct methods the traditional methods (ii) try to split the evaluation of the KS ground state into the determination of a self-consistent charge density (or potential) and the diagonalization of the KS Hamiltonian for a fixed potential. Mathematically this is definitely less elegant, and especially in the solid state community a tendency towards direct minimization of the KS functional is visible. But on the other hand, SC methods have been used for a long time and in addition they have proved to be quite reliable and efficient in most cases. We will show in this paper that an efficient matrix diagonalization and an efficient mixing scheme result in a method which is highly competitive with direct minimization schemes. For metals our implementation seems to outperform any other scheme which we are aware of. In this paper, we will also discuss the scaling properties of our algorithm in terms of computer time and in terms of the necessary number of iterations for increasing system sizes. We will demonstrate that the matrix diagonalization scheme proposed here requires always the same number of iterations independent of the system size. In addition, we will demonstrate that our implementation scales like $O(N^2)$ (in terms of computer time per iteration) for systems containing up to 1000 electrons, indicating that quite large systems can be treated efficiently.

We want to point out that the method discussed here has been used successfully for different systems, including liquid simple metals (Na, Ge),[21] liquid transition metals (V, Cu),[22,23] the transition from a liquid metal to an amorphous semiconductor by rapid quenching of Ge,[24,23] and the metal-nonmetal transition in expanded l-Hg.[25] One advantage of an efficient electronic minimization is that the relaxation of the ions to their instantaneous equilibrium position is much faster. Successful calculations for clean and hydrogenated C(100) surfaces,[26] the site-selective adsorption of C atoms on Al(111) surfaces,[27] and Rh surface properties[28] demonstrate the feasibility of our method in this respect. Finally, we have also performed calculations of phonon dispersion relations in insulators and metals (cubic diamond and graphite see Ref. 29), indicating that forces can be evaluated efficiently and very accurately within the SC methods.

Our paper is organized as follows: In Sec. II, we will discuss the Kohn-Sham energy functional if partial occupancies are allowed, our self-consistency cycle will be outlined and an improved way for the calculation of interatomic forces will be discussed. In Sec. III, we discuss an iterative matrix diagonalization scheme based on Pulay's DIIS method (direct inversion in the iterative subspace). Charge density mixing will be discussed in Sec. IV. Finally, we apply our scheme to several test systems (Sec. V). In this part, we will mainly concentrate on the scaling aspects of the method.

II. THE KOHN-SHAM ENERGY FUNCTIONAL FOR PARTIAL OCCUPANCIES

A. The Kohn-Sham energy functional

In general, the Kohn-Sham free energy functional for an ultrasoft (US) Vanderbilt pseudopotential (PP) at finite temperature can be written as[4,30,31]

$$F_{\mathrm{KS}}[\{\phi\},\{f\},\{\mathbf{R}\}]=\sum_n f_n\langle\phi_n|T+V_{\mathrm{NL}}^{\mathrm{ion}}|\phi_n\rangle+E^{\mathrm{H}}[\rho]+E^{\mathrm{xc}}[\rho]+\int \mathbf{d}^3\mathbf{r}V_{\mathrm{loc}}^{\mathrm{ion}}(\mathbf{r})\rho(\mathbf{r})+\gamma_{\mathrm{Ewald}}(\{\mathbf{R}\})-\sum_n \sigma S(f_n), \quad (1)$$

where $S(f_n)$ is the entropy of noninteracting electrons.[32] The partial occupancies f_n are $f_n=1$ for occupied bands and $f_n=0$ for unoccupied bands, at finite temperature f_n is allowed to vary continuously between 0 and 1. To simplify the notation we have dropped the k-point index. The first sum runs over all bands N_b included in the calculation. E^H is the Hartree energy, E^{xc} the exchange-correlation energy functional, $V_{\mathrm{loc}}^{\mathrm{ion}}$ the local ionic pseudopotential, $T=-(\hbar^2/2m_e)\nabla^2$ the kinetic energy operator, and γ_{Ewald} the Madelung energy of the ions. It is clear that the KS functional depends on the positions $\mathbf{R}_N$ of the ions, the electronic wave functions ϕ_n, and the partial occupancies f_n only.

For US-PP's the nonlocal part of the PP can be written as

$$V_{\mathrm{NL}}^{\mathrm{ion}}=\sum_{ij} D_{ij}^{\mathrm{ion}}|\beta_j\rangle\langle\beta_i|, \quad (2)$$

where D_{ij}^{ion} is the strength of the nonlocal part of the pseudopotential. The charge density $\rho(\mathbf{r})$ is given by

$$\rho(\mathbf{r})=\sum_n f_n|\phi_n(\mathbf{r})|^2+\sum_{n,ij} f_n\langle\phi_n|\beta_j\rangle\langle\beta_i|\phi_n\rangle Q_{ij}(\mathbf{r}), \tag{3}$$

where β_i are localized projection states and $Q_{ij}(\mathbf{r})$ localized augmentation functions.[4,30,31] The total energy has to be minimized subject to the generalized orthonormality constraints

$$\langle\phi_n|\mathbf{S}|\phi_m\rangle=\delta_{nm}, \tag{4}$$

where the overlap matrix $\mathbf{S}$ is defined as

$$\mathbf{S}=1+\sum_{ij} q_{ij}|\beta_j\rangle\langle\beta_i|, \tag{5}$$

with

$$q_{ij}=\int Q_{ij}(\mathbf{r})\mathbf{d}^3\mathbf{r}. \tag{6}$$

Ultrasoft pseudopotentials are discussed in detail in Refs. 4,30,31,33. Their general advantage is that they reduce the necessary energy cutoff for transition metals and first row elements by a factor of 2–4. The resulting basis sets are comparable in size with the basis sets for typical ''pseudopotential elements'' like Na, Al, Si, and Ge.

The most important property of the KS functional is to be extremal in the ground state with respect to arbitrary variations of the wave functions. Under the constraints of orthonormalization, variation with respect to the wave functions leads to the generalized KS eigenvalue equations,[31]

$$\mathbf{H}|\phi_n\rangle=\epsilon_n\mathbf{S}|\phi_n\rangle, \tag{7}$$

where $\mathbf{H}$ is the Kohn-Sham Hamiltonian

$$\mathbf{H}=T+V_{\mathrm{loc}}^{\mathrm{sc}}+V_{\mathrm{NL}}^{\mathrm{sc}}, \tag{8}$$

with

$$V_{\mathrm{loc}}^{\mathrm{sc}}=V_{\mathrm{loc}}^{\mathrm{ion}}+V^H[\rho]+V^{\mathrm{xc}}[\rho], \tag{9}$$

where $V^H[\rho]$ is the Hartree potential and $V^{\mathrm{xc}}[\rho]$ the exchange-correlation potential. For ultrasoft pseudopotentials the nonlocal part of the pseudopotential $V_{\mathrm{NL}}^{\mathrm{sc}}$ depends also on the total local potential and must be calculated accordingly via [compare Eq. (2)]

$$D_{ij}^{\mathrm{sc}}=D_{ij}^{\mathrm{ion}}+\int Q_{ij}(\mathbf{r})V_{\mathrm{loc}}^{\mathrm{sc}}\mathbf{d}^3\mathbf{r}. \tag{10}$$

B. Semiconducting systems versus metallic systems

For the calculation of semiconducting and insulating systems it is an usual practice to calculate the occupied orbitals only, i.e., $N_b=N_{\mathrm{electron}}/2$ and $f_n=1$ for all calculated bands. In this case it can be shown that the total energy is invariant under any unitary transformation of the wave functions ϕ_n, and it is sufficient to calculate a set of wave functions which fulfills the less stringent equation

$$\mathbf{H}|\phi_n\rangle=\sum_m \gamma_{nm}\mathbf{S}|\phi_m\rangle, \tag{11}$$

where γ_{nm} is a Hermitian matrix.

The situation differs significantly in metallic systems; here, one can take two different approaches: It is possible to insist on calculating the occupied orbitals only, but as correctly pointed out by Annett[34] the number of iterations which are necessary to converge to a specific precision will increase with the square root of the system size even for non-self-consistent calculations. The main problem lies in the determination of the highest occupied orbital which will be very close in energy to the lowest unoccupied orbitals (due to the finite system size, level spacing will always be finite). For any minimization or matrix diagonalization algorithm it will become progressively harder to determine the correct occupied orbitals if the virtual energy gap at the Fermi level decreases. Keep in mind that this problem also occurs in semiconductors and molecules with a very small or zero gap.

Because of this behavior, it is generally an advantage to include unoccupied orbitals above the Fermi level in practical calculations. For non-self-consistent calculations, we will show in Sec. III D that our method requires always the same number of iterations, independent of the system size. To ensure good energy-stability partial occupancies have to be included at the same time. This avoids that the crossing of two eigenvalues at the Fermi level causes discontinuities (of the band occupations) during a self-consistent calculation. In addition, partial occupancies also solve the level crossing problem for molecular dynamics or ionic relaxations. One disadvantage of including additional unoccupied orbitals above the Fermi level is that it is necessary to calculate the KS orbitals exactly [Eq. (7)], making the calculation of the electronic ground state slightly more complex.

Another problem which occurs only for self-consistent calculations is the charge sloshing. Charge sloshing arises from the quadratic divergence of the dielectric matrix with the cell length in metals. We will discuss this problem in Sec. IV.

C. Partial occupancies

The first one to study the influence of partial occupancies on the KS functional was Mermin,[32] who extended the LDA to finite temperatures. This approach becomes physically significant if the temperature of the system is comparable to characteristic excitation energies. The impact of partial occupancies on the forces has probably been first discussed independently by Weinert and Davenport[35] and by Wentzcovitch, Martins, and Allen.[36]

The second approach to the introduction of partial occupancies concentrates on the evaluation of the energy at zero temperature: In this case, partial occupancies are introduced as a tool for reducing the number of k points in the Brillouin zone which are necessary to evaluate the band-structure energy. One of the most successful approaches in this respect is the linear tetrahedron (LT) method, in which the one-electron energies $\epsilon_{n\mathbf{k}}$ are interpolated linearly between the k points defining the corners of the elementary tetrahedra filling the BZ, and the integrals—for instance, for the band-structure energy—are performed analytically within each tetrahedron.[37] Blöchl[38] has recently revised the linear tetrahedron method to give effective weights $f(\{\epsilon_{n\mathbf{k}}\})$ for each band and k point. This new formulation gives strictly the same results as the conventional tetrahedron method but it is

easier to implement in existing plane-wave codes like the Vienna *ab initio* simulation package (VASP). In a second step, Blöchl was also able to derive a correction formula which removes the quadratic error inherent in the LT method by going beyond the linear approximation and by including the effects of the curvature of the bands at the Fermi surface (we will refer to this method as LT-C, whereas LT is the standard linear tetrahedron method). The LT-C method converges very fast with the number of k points, and we consider this method to be the most accurate and most unambiguous method for calculating the total energy of bulk materials containing a small number of atoms. Nevertheless the method is not applicable to large supercells because usually only a very small number of k points is used in this case. In addition, one can show that the method is not variational with respect to the partial occupancies making the exact evaluation of forces at least inconvenient (see Refs. 38,39), if not impossible.

Because of this we usually use a variant of the so called ''smearing methods,'' which were first introduced by Fu and Ho[40] in the context of plane-wave pseudopotential calculations. These methods are, in principle, closely related to the finite-temperature approach of Mermin,[32] but their main aim is the evaluation of zero-temperature properties; i.e., the broadening of the one-electron energy levels is only introduced as a mathematical tool to improve the convergence with respect to the number of k points. Currently we are mainly using the method of Methfessel and Paxton (MP):[41] They expanded the step function (the zero-temperature limit of the Fermi-Dirac occupation probability) in a complete orthonormal set of functions. Within this approach the widely used error function (integral of the Gaussian function) is only the lowest order approximation ($N=0$) of the step function, further successive approximations ($N=1,2,\ldots$) can be obtained easily. In analogy to Mermin's finite-temperature method, the total energy is no longer variational with respect to the partial occupancies and has to be replaced by a generalized free energy functional with the correct form for the entropy term $S(f_n)$ (one feature missing in the original work of Methfessel and Paxton). The free energy is now given by

$$F=E-\sum_n \sigma S_N\left(\frac{\epsilon_n-\mu}{\sigma}\right), \tag{12}$$

where S_N is defined by

$$S_N(x)=\tfrac{1}{2}A_N H_{2N}(x)e^{-x^2}. \tag{13}$$

H_m are the Hermite polynomials of degree m and explicit formulas for A_m can be found in Ref. 41.

In contrast to the standard Gaussian method ($N=0$) or to Mermin's finite-temperature approach the entropy term $\Sigma_n \sigma S_N[(\epsilon_n-\mu)/\sigma]$ for higher-order approximants ($N=1,2,\ldots$) will be very small for a reasonable choice of σ, and the deviations from $E_{\sigma=0}$ are only of the order $(2+N)$ in σ

$$F(\sigma)=E_{\sigma=0}+O(\sigma^{2+N}). \tag{14}$$

It is interesting to note that an analytical extrapolation to $\sigma=0$ is possible[16,42] by using

$$E_{\sigma=0}\approx\widetilde{E}(\sigma)=\frac{1}{N+2}[(N+1)F(\sigma)+E(\sigma)]. \tag{15}$$

This shows that the difference between the physically meaningful quantity $E_{\sigma=0}$ and $F(\sigma)$ is proportional to $\approx\Sigma_n \sigma S_N[(\epsilon_n-\mu)/\sigma]$. For the MP method, the entropy term is a simple error estimate for the difference between the free energy F and the ''physically'' correct energy $E_{\sigma=0}$. σ can be increased until this error estimate becomes larger than an allowed threshold (usually 1 meV). Because the free energy and the ''physical'' energy $E_{\sigma=0}$ are the same except for this small error the forces calculated as a derivative of the free energy are also correct and can be used to determine the zero- ''temperature'' ground state. For more details we refer to Ref. 39.

D. Self-consistency loop and iterative methods

The basic idea of the methods based on the self-consistency loop is to split the calculation of the KS-ground-state into two independent subproblems: one is the determination of the eigenfunctions and eigenvalues (i.e., minimization of the band-structure energy for a fixed charge density or potential), the other the calculation of the self-consistent charge density (or potential). One reason why these algorithms are more efficient lies probably in the fact that both subproblems are now rather easy to solve and can be ''preconditioned'' separately.

At the beginning of a calculation we choose an appropriate set of trial wave functions (usually seeded with a random number generator) $\{\phi_n;n=1,\ldots,N_b\}$ and a reasonable input charge density $\rho_{\rm in}$. The initial charge density corresponds to the superposition of the atomic pseudocharge densities of the constituents. From the input charge density, the local potential

$$V_{\rm loc}=V_{\rm loc}^{\rm ion}+V^H[\rho_{\rm in}]+V^{\rm xc}[\rho_{\rm in}] \tag{16}$$

and the corresponding double counting corrections

$$E_{\rm d.c.}[\rho_{\rm in}]=-E^H[\rho_{\rm in}]+E^{\rm xc}[\rho_{\rm in}]-\int \mathbf{d}^3\mathbf{r}V^{\rm xc}(\mathbf{r})\rho_{\rm in}(\mathbf{r}) \tag{17}$$

are evaluated. For ultrasoft pseudopotentials the nonlocal part of the pseudopotential depends also on the local potential and must be calculated accordingly [Eq. (10)]. In the next step the N_b trial wave functions are improved using an iterative method, and the new eigenenergies are used to calculate a new Fermi energy and new partial occupancies. The total free energy for the current iteration is calculated as the sum of the band-structure energy plus the entropy term plus double counting corrections,

$$F=\sum_n f_n\epsilon_n^{\rm app}-\sum_n \sigma S\left(\frac{\epsilon_n-\mu}{\sigma}\right)+E_{\rm d.c.}[\rho_{\rm in}]+\gamma_{\rm Ewald}. \tag{18}$$

Conceptually, the calculated energy corresponds to the energy evaluated from the Harris-Foulkes (HF) functional,[43–45] which is non-self-consistent—in contrast to the KS functional: the HF functional [defined in Eq. (18)] requires the calculation of the band-structure energy for a fixed charge

density ρ_{in}. With our code it is easy to evaluate this energy keeping the initial charge density fixed (for instance, to the superposition of atomic pseudocharge densities) and iterating only the eigenvectors until they are converged.

To get the exact KS-ground-state energy, self-consistency with respect to the input charge density requires that the charge density residual vector $R[\rho_{\text{in}}]$,

$$R[\rho_{\text{in}}]=\rho_{\text{out}}-\rho_{\text{in}}, \tag{19}$$

is zero where the output charge density ρ_{out} is calculated from the wave functions using Eq. (3). The residual vector $R[\rho_{\text{in}}]$ — and possibly information from previous mixing steps — allows one to calculate a new charge density ρ_{in} for the next self-consistency loop. In principle, it is necessary to evaluate the eigenfunctions ϕ_n exactly for each new input charge density, making ρ_{out} and the residual vector R functionals of the input charge density ρ_{in} only. Nevertheless, even in conjunction with elaborated Broyden-type mixing techniques, our results indicate that this is not necessary if the final (approximate) wave functions of the previous mixing iteration are used as new initial trial wave functions. In this case a few steps in the iterative matrix diagonalization are sufficient to get a reliable result for the charge density residual vector R. Mind that the HF energy functional has a saddle point at the KS-ground-state, i.e., it is *not minimal*. Therefore, it is not possible to use a standard conjugate gradient algorithm for the determination of the optimal ρ_{sc}.

In Sec. III, we will concentrate on different iterative methods for the diagonalization of the KS Hamiltonian, Sec. IV will discuss algorithms for the charge density mixing.

E. Forces

Forces for the finite-temperature KS functional can be obtained easily, and are, in principle, just given by

$$\mathbf{F}_N=\left.\frac{\partial \overline{F}}{\partial \mathbf{R}_N}\right|_{\text{ground state,}} \tag{20}$$

where $\overline{F}$ is the free energy of the system [Eq. (1)] plus the orthonormality constraints added with appropriate Lagrange multipliers.

This formula is exact and contains Hellmann-Feynman[3] as well as Pulay contributions[46] (for a plane-wave basis set no Pulay contributions exist, but mind that the term Pulay force is used quite ambiguously in literature). A similar formula also holds for the stress tensor, and derivatives with respect to the basis set are implicitly contained in this definition. It is now easy to show that the forces can be rewritten as (for the self-consistent case this equation was first derived in Ref. 47)

$$\mathbf{F}_N=\sum_{nqq'} f_n C^*_{nq'}\frac{\partial(\mathbf{H}^{\text{sc}}[\rho,\{\mathbf{R}\}]-\epsilon_n\mathbf{S}[\{\mathbf{R}\}])_{q'q}}{\partial \mathbf{R}_N}C_{nq}, \tag{21}$$

where C_{nq} is the expansion coefficient of ϕ_n for the plane-wave component $\mathbf{q}$, i.e.,

$$|\phi_n\rangle=\sum_q C_{nq}|\mathbf{q}\rangle, \tag{22}$$

and

$$\mathbf{H}_{qq'}=\langle \mathbf{q}'|\mathbf{H}|\mathbf{q}\rangle. \tag{23}$$

In Eq. (21) changes of the Hamiltonian $\mathbf{H}$ due to changes in the self-consistent charge density ρ^{sc} must *not* be calculated (because of the augmentation charges the charge density ρ^{sc} depends explicitly on the ionic positions for US-PP). For further details we refer to Refs. 30 and 31.

One complication arises from the fact that we are using the Harris-Foulkes functional instead of the exact Kohn-Sham functional, but looking carefully at the Harris-Foulkes functional it is possible to obtain a correction formula for the forces. If the input charge density ρ_{in} for the Harris-Foulkes functional is calculated from the atomic charge densities of the constituents (i.e., for non-self-consistent calculations), a Pulay-like additional term arises that accounts for the fact that the input charge density depends explicitly on the atomic coordinates. In this case, $\mathbf{H}$ in Eq. (21) has to be replaced by the Hamiltonian calculated from the superposition of the atomic charge densities, $\mathbf{H}[\rho_{\text{atom}},\{\mathbf{R}\}]$, and the term

$$\int \mathbf{d}^3\mathbf{r}\left(\frac{\partial V^H(\rho_{\text{atom}})+V^{\text{xc}}(\rho_{\text{atom}})}{\partial \mathbf{R}_N}(\mathbf{r})[\rho_{\text{out}}(\mathbf{r})-\rho_{\text{atom}}(\mathbf{r})]\right) \tag{24}$$

has to be added to the forces. In Eq. (21) changes of the Hamiltonian $\mathbf{H}$ due to changes in the input charge density ρ_{atom} have to be omitted as in the self-consistent case.

We have found that the analogous correction formula

$$\int \mathbf{d}^3\mathbf{r}\left(\frac{\partial V^H(\rho_{\text{atom}})+V^{\text{xc}}(\rho_{\text{atom}})}{\partial \mathbf{R}_N}(\mathbf{r})[\rho_{\text{out}}(\mathbf{r})-\rho_{\text{in}}(\mathbf{r})]\right) \tag{25}$$

also improves the convergence of the forces during a self-consistent calculation (this formula is very much in the spirit of Ref. 48). In this case, $\mathbf{H}$ in Eq. (21) has to be replaced by $\mathbf{H}[\rho_{\text{in}},\{\mathbf{R}\}]$, where ρ_{in} is the charge density obtained in the previous iteration. In principle, it is necessary to evaluate the change of ρ_{in} if the ions move (i.e., the first term in Eq. (25) should be $\partial[V^{\text{H}}(\rho_{\text{in}})+V^{\text{xc}}(\rho_{\text{in}})]/\partial\mathbf{R_N}$); it is not possible. However, simply replacing the change of ρ_{in} by the change of a superposition of atomic charges ρ_{atom} leading to Eq. (25) provides an excellent approximation. This correction formula improves the precision of the forces by almost two orders of magnitude and allows the self-consistency cycle to stop much earlier.

This is demonstrated in Fig. 1, where the convergence of the forces is compared for different algorithms for a long cell containing 16 Fe atoms (see Sec. V A 2). It can be seen that the optimized scheme (opt) explained here results in the best overall performance. Mind, that a similar convergence rate might be obtained by using the mixed charge density (mix) (Sec. IV) for the calculation of the local contribution to the forces, i.e.,

$$\sum_{nqq'} f_n C^*_{nq'}\frac{\partial V^{\text{ion}}_{\text{loc}q'q}}{\partial \mathbf{R}_N}C_{nq}\rightarrow\int \mathbf{d}^3\mathbf{r}\frac{\partial V^{\text{ion}}_{\text{loc}}}{\partial \mathbf{R}_N}(\mathbf{r})\rho_{\text{mixed}}(\mathbf{r}). \tag{26}$$

Especially this part of the forces is very sensitive to changes in the charge density. The use of the output charge density (out)—i.e., the left side of Eq. (26)—without the correction term [Eq. (25)] makes the forces worse by a factor 100 (see Fig. 1).

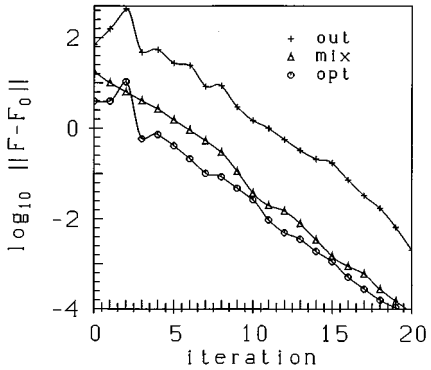

FIG. 1. Convergence of the forces (in eV/Å) for different algorithms for fcc-Fe (four cells). ''out''—output charge density was used for the calculation, ''mix''—mixed charge density was used, and ''opt''—is the optimized scheme explained in the text.

III. ITERATIVE METHODS FOR THE DIAGONALIZATION OF THE KS HAMILTONIAN

It is clear that a fast iterative matrix diagonalization is one of the key points within our algorithm. We have tested several methods: (i) the blocked Davidson scheme, which was first proposed by Davidson[49] and later modified by Liu for a simultaneous update of all bands,[50] (ii) the sequential CG algorithm proposed by Teter *et al.*,[14] and later used by Bylander, Kleinman, and Lee[51] for the iterative diagonalization of the KS Hamiltonian, and (iii) a new variant of the residual vector minimization scheme—direct inversion in the iterative subspace (RMM-DIIS).[53,54] We have found that our variant of the RMM-DIIS scheme is generally most efficient for large systems, whereas the sequential CG method is marginally more efficient for small systems. We will, therefore, concentrate on these two approaches. For a more elaborated comparison of different matrix diagonalization schemes we refer to Ref. 39.

A. Sequential conjugate gradient (CG) minimization

This algorithm proposed by Teter *et al.*[14] was originally used to minimize the total energy directly. It is straightforward to apply this algorithm to the diagonalization of large matrices.[51] The algorithm is strictly sequential, i.e., one band is optimized usually several times and then a move to the next band is done. For the optimization of each band a standard CG (Refs. 12 and 13) algorithm is used in which the expectation value of the Hamiltonian for one specific band is optimized:

$$\epsilon_{\rm app}=\frac{\langle\phi_m|\mathbf{H}|\phi_m\rangle}{\langle\phi_m|\mathbf{S}|\phi_m\rangle}. \tag{27}$$

This quantity is called Rayleigh quotient. Variation of the Rayleigh quotient with respect to $\langle\phi_m|$ leads to the residual vector defined as

$$|R(\phi_m)\rangle=(\mathbf{H}-\epsilon_{\rm app}\mathbf{S})|\phi_m\rangle, \tag{28}$$

if $\langle\phi_m|\mathbf{S}|\phi_m\rangle=1$. To ensure that orthonormality to all other bands is maintained, Lagrange multipliers must be introduced resulting in the following gradient vector:

$$|g(\phi_m)\rangle=|g_m\rangle=\left(1-\sum_n|\phi_n\rangle\langle\phi_n|\mathbf{S}\right)\mathbf{K}(\mathbf{H}-\epsilon_{\rm app}\mathbf{S})|\phi_m\rangle, \tag{29}$$

with $\mathbf{K}=\mathbf{1}$. Minimization is done along this gradient vector, respectively, the conjugated gradient vector.

To improve the efficiency we have adopted the preconditioning function proposed by Teter *et al.*:[14] In this case $\mathbf{K}$ is given by

$$\mathbf{K}=-\sum_q\frac{2|\mathbf{q}\rangle\langle\mathbf{q}|}{3/2E^{\rm kin}(R)}\frac{27+18x+12x^2+8x^3}{27+18x+12x^2+8x^3+16x^4},$$

$$\text{with}\quad x=\frac{\hbar^2}{2m_e}\frac{q^2}{3/2E^{\rm kin}(R)}, \tag{30}$$

where $E^{\rm kin}(R)$ is the kinetic energy of the residual vector. There are only two minor changes with respect to Ref. 14: First, we use $3/2E^{\rm kin}(R)$ instead of $E^{\rm kin}(R)$ in the definition of x, resulting in a slightly improved convergence for most elements. Second, we multiply the preconditioning function by a factor $2/[3/2E^{\rm kin}(R)]$. Using this factor the diagonal part of the preconditioning matrix $\mathbf{K}$ converges towards

$$\frac{2m_e}{\hbar^2q^2} \tag{31}$$

for large q resulting in a more convenient ''length'' of the correction vector $|g(\phi_m)\rangle$.

The CG scheme is actually very stable and reasonably efficient, but there is one minor complication. The sequential algorithm described here leads to ground-state eigenvalues and eigenvectors which represent only an arbitrary linear combination [in the sense of an unitary transformation, Eq. (11)] of the exact KS eigenstates. As already mentioned this is sufficient for semiconductors and insulators if one calculates only occupied bands. But for metals we need the true Kohn–Sham eigenstates. In this case it is convenient to add one additional step, which is called subspace rotation. In this step the final improved wave functions $\{\phi_m;m=1,\ldots,N_b\}$ are unitarily transformed in such a way that the Hamiltonian matrix becomes diagonal in the subspace spanned by the transformed wave functions (this is in the spirit of the Rayleigh-Ritz scheme[52]). This requires first the evaluation of the Hamilton matrix (and possibly the overlap matrix) in the subspace $\{\phi_m\}$

$$\overline{\mathbf{H}}_{nm}=\langle\phi_n|\mathbf{H}|\phi_m\rangle,\quad \overline{\mathbf{S}}_{nm}=\langle\phi_n|\mathbf{S}|\phi_m\rangle, \tag{32}$$

and then the diagonalization of this small Hamiltonian, using a conventional matrix diagonalization scheme, i.e.,

$$\sum_m\overline{\mathbf{H}}_{nm}\mathbf{B}_{mk}=\sum_m\epsilon_k^{\rm app}\overline{\mathbf{S}}_{nm}\mathbf{B}_{mk}. \tag{33}$$

The lowest eigenvalue-eigenvector pairs

$$\epsilon_k^{\rm app},|\overline{\phi_k}\rangle=\sum_m\mathbf{B}_{mk}|\phi_m\rangle \tag{34}$$

correspond to the best approximation of the exact lowest eigenvalues and eigenvectors within the subspace spanned by $\{\phi_m\}$.

Mind, that there is at least one alternative to the subspace rotation: If the orthonormalization of the search vector is restricted to those bands which are *lower or equal in energy* than the actual band (this is sufficient to ensure stability of the algorithm, but it requires of course a global reorthonormalization of all bands afterwards, e.g., applying a Gram-Schmidt orthogonalization), we would also observe convergence towards the real Kohn-Sham eigenstates. The reason is simple: Here no constraint is imposed on the lowest state guaranteeing that it converges to the absolute minimal eigenvalue, i.e., the true lowest KS eigenstate. The second-lowest state is only constrained to be orthonormal to the lowest state but again no further restrictions are applied, guaranteeing convergence towards the exact second-lowest KS state, and so on. Nevertheless, considering Sec. III D it becomes clear that this algorithm requires more and more iterations if the system size increases (i.e., if the spacing of the eigenvalues decreases).

B. Residual minimization method—direct inversion in the iterative subspace (RMM-DIIS)

The only remaining drawback of the CG algorithm is the necessity for an explicit orthonormalization of the preconditioned residual vector $\mathbf{K}|R(\phi_m)\rangle$ to the current set of trial wave functions [Eq. (29)]. This operation is the most time consuming part for large-size problems because one single vector must be orthonormalized to a large number of other vectors *for each single band update*. This requires a very high bandwidth from the main computer memory to the central processing unit (CPU), i.e., the peak bandwidth from the main memory to the CPU and not the peak floating-point performance of the CPU becomes often the limiting factor.

Unfortunately avoiding the orthonormalization is not possible applying the CG algorithm discussed above. An *unrestricted* CG scheme applied to the minimization of the Rayleigh quotient will always determine the lowest possible eigenvalue, i.e., if the orthonormalization is not done the algorithm will converge from any starting point towards the lowest eigenvector of the Hamiltonian. Ultimately, this property is connected with the fact that the Rayleigh quotient is stationary at each eigenvector but it does *not* possess a minimum except for the lowest eigenvalue (i.e., the Rayleigh quotient can be lowered by moving towards lower eigenvalues). Only the explicit orthonormalization makes it possible to achieve a stable and efficient convergence to a selected eigenvalue.

Fortunately, a solution to this problem exists which was first proposed by Wood and Zunger.[54] Minimizing *the norm of the residual vector* (hence the name residual minimization method) instead of the Rayleigh quotient makes the orthonormalization unnecessary (at least, in principle) because the norm of the residual vector has an *unconstrained* local minimum at each eigenvector (the norm of the residual vector is clearly positive definite).

In our implementation we use the original minimization method proposed by Pulay[53] and not the variant proposed by Wood and Zunger.[54] Wood's algorithm requires the additional calculation and storage of $\mathbf{S}|\phi\rangle$, and is therefore slower than our algorithm. We start with an evaluation of the preconditioned residual vector $\mathbf{K}|R_m^0\rangle=\mathbf{K}|R(\phi_m^0)\rangle$ for a selected band m. Then a Jacobi-like trial step along this direction is done

$$|\phi_m^1\rangle=|\phi_m^0\rangle+\lambda\mathbf{K}|R_m^0\rangle \tag{35}$$

and the new residual vector $|R_m^1\rangle=|R(\phi_m^1)\rangle$ is evaluated. Next a linear combination of the initial $|\phi_m^0\rangle$ and the trial wave function $|\phi_m^1\rangle$

$$|\overline{\phi^M}\rangle=\sum_{i=0}^{M}\alpha_i|\phi_m^i\rangle \quad \text{with } M=1 \tag{36}$$

is searched which minimizes the norm of the residual vector. Assuming linearity in the residual vector, i.e.,

$$|\overline{R^M}\rangle=|R(\overline{\phi^M})\rangle=\sum_{i=0}^{M}\alpha_i|R_m^i\rangle, \tag{37}$$

this requires the minimization of

$$\frac{\sum_{j,i=0}^{M}\alpha_i^*\alpha_j\langle R_m^i|R_m^j\rangle}{\sum_{j,i=0}^{N}\alpha_i^*\alpha_j\langle\phi_m^i|\mathbf{S}|\phi_m^j\rangle}. \tag{38}$$

This step is usually called DIIS and lies at the heart of Pulay's minimization scheme.[53] It is, for instance, possible that the trial step moves towards a band with lower energy lowering the Rayleigh quotient. In this case the DIIS ''corrector step'' will recognize the false move because the norm of the residual vector has increased and will correct the wrong trial step. The final move might even have a reversed sign from the trial move.

The problem stated in Eq. (38) is equivalent to determining the lowest eigenvector/eigenvalue from the Hermitian eigenvalue problem,

$$\sum_{j=0}^{M}\langle R_m^i|R_m^j\rangle\alpha_j=\epsilon\sum_{j=0}^{M}\langle\phi_m^i|\mathbf{S}|\phi_m^j\rangle\alpha_j. \tag{39}$$

The next trial step starts from $|\overline{\phi^M}\rangle$ along the direction $\mathbf{K}|\overline{R^M}\rangle$. In each iteration $M+1$ a new wave function $|\phi_m^{M+1}\rangle=|\overline{\phi^M}\rangle+\lambda\mathbf{K}|\overline{R^M}\rangle$ and a new residual vector $|R(\phi_m^{M+1})\rangle$ are added to the ''iterative'' subspace. The size of the trial step λ is a critical value for the stability of the algorithm. We have found that a reasonable choice for the trial step can be obtained from the minimization of the Rayleigh quotient along the search direction in *the first step*, this optimal λ is used until a move to the next band is performed. The line minimization can be done without additional computational requirements. Usually the optimal step size is between $0.3\leqslant\lambda\leqslant1$ for the preconditioning function given in Eq. (30). In rare cases—especially if the minimization of the Rayleigh quotient starts to go for the wrong band—the trial step might become very large. Therefore, we restrict the size of the trial step to a value between $0.1\leqslant\lambda\leqslant1$. Usually we perform several DIIS steps and one final trial step. The reason for finishing with the trial step is that the trial step is very cheap in comparison with the DIIS step, so even if the improvement in the trial step is small it is ''economical'' to do

it (for the majority of bands the trial step is already close to the exact position of the minimal residual vector).

The scheme explained in this section requires approximately the same number of iterations as the CG algorithm, but it avoids any explicit orthonormalization of the search vectors and is therefore much faster for very large problems where the orthonormalization is the leading factor. Even more important is the fact that the residual minimization is inherently local and it is, therefore, very easy to implement the algorithm on a parallel machine. For instance, each processor might handle a certain number of bands, information about other bands is not required (see, also, Sec. III D).

A remaining drawback of the RMM-DIIS method is that it always finds the vector which is closest to the initial trial vector. This leads, in principle, to serious problems because we have no guarantee to convergence to the correct ground state at all. Therefore initialization becomes a critical step, i.e., if the initial set of wave functions does not ''span'' the real ground state it might happen that in the final solution some eigenvectors are ''missing:'' To avoid this the initialization must be done with great care. We usually start with a set of random trial vectors and perform five sweeps over all bands. Each initial sweep consists of one subspace rotation and two *steepest–descent* steps into the direction of the preconditioned residual vectors [Eq. (35)] per band (see also Sec. III C). During this initial phase the Hamiltonian is also kept fixed, after this ''delay'' we switch to the RMM-DIIS scheme and start to update the potential. For problematic cases (for instance, if we use no unoccupied bands) we sometimes use the CG scheme described above for three initial steps.

As already explained subspace rotation and sequential update of the bands alternate. In the residual minimization scheme the final vectors are no longer orthogonal, we therefore reorthonormalize the vectors at the end after sweeping over all bands. We want to emphasize, that, in principle, the RMM-DIIS method *should* also converge without any explicit subspace rotation or orthonormalization, but we have found that the subspace rotation speeds up the calculations although it is an order $O(N^3)$ operation (see Sec. III D). The main problem is that the ''barrier'' in the norm of the residual vector between two neighboring eigenvectors with eigenvalues ϵ and $\epsilon+\delta\epsilon$ is only of the order $\delta\epsilon$.[55] Into directions of eigenvectors which differ significantly in energy the barrier will be very high. Therefore two eigenvectors which are close in energy are lying in one long steep valley and only a shallow hill separates them—a typical example of a badly conditioned minimization problem. The subspace rotation solves this problem because after the subspace rotation the residual vectors are orthonormal to the current trial set,

$$\langle \phi_n | R(\phi_m)\rangle = 0, \quad \forall m,n \tag{40}$$

and search vectors parallel to the long valleys are effectively suppressed. In this case the residual vector is exactly equivalent to the gradient defined in Eq. (29). (Another more rigorous way to look at this problem is given in Sec. III D.)

In several tests for large systems we have found that the subspace rotation and the reorthonormalization not only improve convergence but are indeed the only way to get good stability (especially if the spacing of the eigenvalues is small). The orthonormalization strictly avoids that two states converge to the same eigenvector, and the subspace rotation suppresses all ''unstable'' search directions towards wrong bands. Actually without these $O(N^3)$ steps it can easily occur that the RMM-DIIS algorithm incorrectly overcomes the small barrier between close eigenvectors making the whole scheme highly instable (orthonormalization to a number of bands in an energy window around the selected band might be another possible solution to this problem). Also mind that we get a reasonable trial step λ only because we use the subspace rotation, i.e., in the first step the residual vector is the same as the exact orthogonalized gradient vector [Eq. (29)] making the initial steepest-descent step efficient and rather stable (in the first steepest-descent step the danger to go for the wrong band is negligible). We want to point out here that the subspace rotation was missing in the original work of Wood and Zunger[54] and we suspect that their algorithm will be unstable for large systems with a small level spacing.

C. The complete algorithm

The complete self-consistency loop consists of several steps (the section where the algorithm has been discussed is given in brackets): (i) subspace rotation (III A), (ii) CG (III A) or RMM-DIIS (III B) minimization, (iii) orthonormalization using a Gram-Schmidt method (only required for the RMM-DIIS scheme), (iv) update of partial occupancies and charge density for a self-consistent calculation.

In each iteration the initial trial set $\{|\phi_n\rangle, n=1,\ldots,N_b\}$ is equivalent to the final set of the previous iteration, initialization is done with a random number generator. This loop is repeated until self-consistency is reached, for a non-self-consistent calculation no charge density update is done.

We have found that the subspace rotation should be performed between the update of the charge density and the RMM-DIIS or CG step, especially at the beginning of a self-consistent calculation. In this case the calculated residual vectors $|R(\phi_m)\rangle$ agree with the exact gradients $|g(\phi_m)\rangle$. For this reason and because the wave functions should be orthonormal for a recalculation of the charge density, it is necessary to separate the orthonormalization of all wave functions and the diagonalization of the subspace Hamilton matrix which is done at once in the Rayleigh-Ritz scheme.

In addition it is necessary to find an optimal break condition for the sequential RMM-DIIS and CG algorithms. A static criterion, for example two or three steps per band, is not a good choice, because lower bands converge usually much faster than higher bands. Therefore, we have adopted the following dynamic criterion (which is inspired by Ref. 56): (i) Both algorithms are stopped if the change in the total eigenvalue becomes smaller than $E_{\text{accuracy}}/N_b/4$, where E_{accuracy} is the required accuracy of the calculation and N_b is the number of bands included in the calculation. (ii) The RMM-DIIS method is stopped if the squared norm of the residual vector gets smaller than 30% of its initial value, and the minimization always stops with the trial step. (iii) The CG method is stopped if the change in the eigenvalue gets smaller than 30% of the change in the first, i.e., the steepest-descent step. (iv) The maximum number of steps is always four. For the RMM-DIIS the residual vector is minimized three times and finally a fourth trial step is performed. (v) Empty bands are optimized only twice.

By now, these criteria have been used for a large number of system and are very robust. In most cases two CG or two RMM-DIIS steps are done per band, but problematic eigenvalue/eigenvector pairs are iterated more frequently. Usually more iterations are done for the higher bands, and the total speed of convergence for all bands is very good.

D. Computational costs, convergence, and number of iterations

To make a fair comparison between the CG and RMM-DIIS scheme it is necessary to count the number of operations for each algorithm carefully. The CG minimization of the Rayleigh quotient requires always slightly less evaluations of the Hamiltonian multiplied with a wave function than the RMM-DIIS method, but for large systems the most expensive part is the orthonormalization of the wave functions. For our implementation the evaluation of $(\mathbf{H}-\epsilon_n\mathbf{S})|\phi_n\rangle$ is an order

$$T^H = N_b N_{\mathrm{plw}} \ln N_{\mathrm{plw}} \propto N^2 \ln N \tag{41}$$

operation, where N qualifies the system size. The limiting factors are to a smaller degree the fast Fourier transformations ($N_b N_{\mathrm{plw}} \ln N_{\mathrm{plw}} \propto N^2 \ln N$) and the evaluation of the nonlocal projection operators. For large systems we calculate the nonlocal projection operators in real space[57] and therefore the number of operations per band increases linearly with the system size (N). For all bands this is only an $O(N^2)$ operation. We want to point out that this fact is of considerable importance for the speed of the RMM-DIIS method. If T^H is an $O(N^3)$ operation the improvement of the RMM-DIIS method over other minimization methods would be negligible. The main advantage of the RMM-DIIS method is to reduce $O(N^3)$ operations (like orthonormalization) to an absolute minimum.

The Gram-Schmidt orthonormalization takes

$$T^{\mathrm{GS}} = N_b^2 \times N_{\mathrm{plw}} \propto N^3 \tag{42}$$

steps, whereas the explicit orthogonalization of the gradients of each band to all other bands in Eq. (29) takes twice as many steps,

$$T^{\mathrm{ort}} = 2N_b^2 \times N_{\mathrm{plw}} \propto 2N^3. \tag{43}$$

But even worse, the explicit orthogonalization makes any efficient memory caching impossible. The CG algorithm is strictly sequential and at each iteration the new gradient must be orthogonalized to all other bands, requiring a large bandwidth from the main memory. For the Gram-Schmidt orthonormalization a routine with good data locality which avoids this problem can be found easily and T^{ort} is therefore, depending on the computer system, 4–8 times larger than T^{GS} (see Appendix A). Efficient routines with good data locality can also be found for the subspace rotation, and the number of operations is

$$T^{\mathrm{diag}} = T^H + 1.5 N_b^2 \times N_{\mathrm{plw}}. \tag{44}$$

For large systems, where the orthogonalization is the leading factor, one RMM-DIIS step is clearly much faster than one CG step, if we consider only $O(N^3)$ operations the improvement is close to a factor of 4 (assuming that $T^{\mathrm{ort}} \approx 4T^{\mathrm{GS}}$, and that each band is optimized twice). For the RMM-DIIS scheme, we have found a nearly quadratic scaling for systems containing up to 1000 electrons. This means that cells with approximately 200 ''simple'' atoms (Al, Si, C) and 100 transition metal atoms can be treated efficiently, for these system sizes the RMM-DIIS scheme is twice as fast as the CG scheme.

At this point we want to comment on the number of iterations required in the iterative matrix diagonalization. With ''number of iterations'' we always refer to the number of outer loops, the number of iterations in the inner loop is kept fixed, and we have discussed the break criterion for the inner loop previously. Let us assume for the moment that the Hamiltonian is kept fixed. The convergence behavior for some methods was discussed recently by Annett[34] and Tassone *et al.*[8] and a similar analysis can be done in our case. The convergence depends mainly on the eigenvalue spectrum of the Hamilton matrix. Assuming that the trial wave function ϕ_i is close to the real eigenvector ξ_i, it is possible to expand ϕ_i in terms of the real eigenvectors ξ_j

$$|\phi_i\rangle = |\xi_i\rangle + \sum_{i \neq j}^{N_{\mathrm{plw}}} c_j |\xi_j\rangle, \tag{45}$$

where c_j is small. To second order the residual vector is given by

$$|R(\phi_i)\rangle = \sum_{i \neq j}^{N_{\mathrm{plw}}} (\epsilon_j - \epsilon_i) c_j \mathbf{S}|\xi_j\rangle. \tag{46}$$

The required number of iterations depends only on the range of $\gamma = \epsilon_j - \epsilon_i$ and for a simple steepest-descent approach it is given by $\gamma_{\max}/\gamma_{\min}$, for the CG and RMM-DIIS scheme it is given by $\sqrt{\gamma_{\max}/\gamma_{\min}}$.[34] Actually our algorithm is some mixture in between because we do a quite small number of CG and RMM-DIIS steps for each band.

The value of $\gamma_{\max}$ is clearly independent of the system size and mainly determined by the plane-wave cutoff. Because of the dominance of the kinetic energy part $\gamma_{\max}$ is equal to $\gamma_{\max} \approx \hbar^2 G_{\mathrm{cut}}^2 / 2m_e$. The preconditioning [according to Eq. (30)] limits $\gamma_{\max}$, it effectively removes the kinetic energy dominance from the residual vector and makes the convergence practically independent of the chosen cutoff. The effect of the subspace rotation is to limit $\gamma_{\min}$. After the subspace rotation, each trial wave function ϕ_i does not contain any component of eigenvectors $\{\xi_j; j=1,\ldots,N_b; j \neq i\}$ up to second order and the residual vector is not pointing into the direction of another band included in the trial basis set [see Eq. (40)]. In other words, only components resulting from the bands above the highest band included in the calculation can contribute in Eq. (46), i.e., the smallest γ is then determined by $\epsilon_{(N_b+1)} - \epsilon_i$. The band which converges slowest is evidently the highest band $i = N_b$, and $\gamma_{\min}$ for this band is determined by the spacing of the eigenvalues at the top of the included number of bands (for a lower band index the energetical distance to the band $i = N_b + 1$ will be larger resulting in a faster convergence). If only occupied bands are taken into account in the calculation of a metal then $\gamma_{\min}$ will depend on the system size and the larger the system is the slower the convergence of the highest band (and hence of the total energy and the forces) will be. Mind

that a small γ means that the energy is varying very slowly in the corresponding direction, and actually these ''soft'' modes are also the reason why CP-MD's are inherently instable for metals, i.e., because of the weak curvature of the total energy the modes are oscillating with a very low frequency in the CP approach and couple strongly to the ionic degrees of freedom.

The key point is the inclusion of a sufficient number of empty bands so that for all occupied bands $\gamma_{\min}$ is sufficiently large. If the unoccupied bands span an energy range $E_{\text{unoccupied}}$ then $\gamma_{\min}$ will have a lower bound of $E_{\text{unoccupied}}$ for all occupied bands (resulting in a convergence of the total energy and of the forces which is independent of the system size). In this case the problematic region where the convergence is slow and system size dependent is moved upwards to the unoccupied bands, which are of no importance for the total energy and forces.

It is now rather clear that, both (i) the subspace rotation (no matter which system type) and (ii) a reasonable number of empty bands are required in order to get an efficient scheme for metals. In principle, it might seem that $O(N^3)$ operations can be omitted totally but this is shortsighted. Not only the overall stability will degrade without these operations (see Sec. III A), but also the savings in CPU time *per iteration* would be overcompensated by an increased number of iterations needed to achieve a certain accuracy (the real important quantity is the product of CPU time per iteration times the number of necessary iterations). Only for rather small cells with a large spacing of the eigenvalues merely any algorithm works reliably.

We will demonstrate that our approach is very stable even for transition metals (with a small level spacing) and we will demonstrate that the number of iterations is independent of the system size. Overall the scaling of the CPU-time requirements of our iterative matrix diagonalization scheme is nearly quadratic with respect to the system size, i.e., $O(N^2)$, for systems containing up to 1000 electrons. For self-consistent calculations things are more complicated and in this case the number of iterations is often mainly determined by the convergence of the charge density mixing scheme. This topic will be addressed in Sec. IV B.

IV. CHARGE DENSITY MIXING

The second key step within our algorithm is an efficient mixing of the input and output charge densities. The central quantity, which has to be minimized in this case is the charge density residual vector $R[\rho_{\text{in}}]$ [see Eq. (19)],

$$R[\rho_{\text{in}}]=\rho_{\text{out}}[\rho_{\text{in}}]-\rho_{\text{in}}. \tag{47}$$

Linear mixing (or one might call it steepest-descent mixing) adds a certain amount of R to the current input charge density,

$$\rho_{\text{in}}^{m+1}=\rho_{\text{in}}^{m}+AR[\rho_{\text{in}}^{m}]. \tag{48}$$

This is usually a quite slow approach. If it is assumed that $R[\rho]$ can be linearized around its root ρ_{sc},

$$R[\rho]=-\mathbf{J}(\rho-\rho_{\text{sc}}), \tag{49}$$

then it can be shown that the convergence depends only on the eigenvalue spectrum Γ of the Jacobian matrix $\mathbf{J}$, respectively, on the eigenvalue spectrum $\Gamma=1/\Gamma$ of the inverse of the Jacobian matrix $\mathbf{J}^{-1}$. A thorough analysis of the linear mixing approach was done by Dederichs and Zeller[11] — they showed that the maximal linear mixing parameter A is determined by $A<2/\Gamma_{\max}=2\Gamma_{\min}$, and the number of iterations required to reach a specified precision is given by $\Gamma_{\max}/\Gamma_{\min}=\Gamma_{\max}/\Gamma_{\min}$.

It is easy to show that the matrix $\mathbf{J}$ is the charge dielectric matrix which describes the total self-consistent change of the charge density $\rho_{\text{tot}}=\rho_{\text{ext}}+\rho_{\text{ind}}$ for an external charge perturbation ρ_{ext}, i.e.,

$$\rho_{\text{tot}}=\mathbf{J}^{-1}\rho_{\text{ext}}.$$

Following Vanderbilt and Louie,[59] $\mathbf{J}$ is given by

$$\mathbf{J}=1-\chi\mathbf{U}, \tag{50}$$

where χ is the dielectric susceptibility and $\mathbf{U}$ is an operator describing the change of the potential due to a change of the charge density; considering the Hartree potential only (which is usually of much higher importance than the exchange-correlation potential), this operator is given by (see, also, Refs. 58, 11, and 34)

$$\langle\mathbf{q}|\mathbf{U}|\mathbf{q}'\rangle=\delta_{\mathbf{qq}'}\frac{4\pi\mathbf{e}^2}{\mathbf{q}^2}. \tag{51}$$

For metals the dielectric matrix diverges quadratically for small q; this behavior is usually called ''charge sloshing.'' Mind that the dielectric matrix does not diverge in insulators making self-consistent calculations much easier in insulators.

Recently, Annett[34] discussed some aspects of the convergence of the charge density mixing approach. He points out that accelerated Broyden-like methods[62] exist, but incorrectly states that they should not have an improved scaling over the steepest-descent approach. There is at least one method (Tchebycheff acceleration scheme) proposed by Akai and Dederichs[11] for which it was shown explicitly that the number of iterations is proportional to $\sqrt{\Gamma_{\max}/\Gamma_{\min}}$. In addition, Blügel[60] has demonstrated for several examples that Broyden-like methods are even more efficient than the Tchebycheff acceleration scheme and it is well known that all Broyden or quasi-Newton schemes[62] show also a quadratic convergence.[13] We therefore assume that for these methods the number of iterations necessary to converge to a specific accuracy is proportional to $\sqrt{\Gamma_{\max}/\Gamma_{\min}}$ (we will come back to this topic in Sec. IV B and Sec. V B 2).

In practice, we are again resorting to the DIIS scheme proposed by Pulay.[53] Recently we were able to show[39] that this scheme is closely related to the charge density mixing scheme proposed by Johnson.[61] Johnson's scheme is a variant of the well known Broyden algorithms.[62] These algorithms try to build up an approximation of the Jacobian matrix $\mathbf{J}$ or the inverse of the Jacobian matrix $\mathbf{G}=\mathbf{J}^{-1}$ by updating an approximation of the Jacobian matrix at each iteration and were pioneered for electronic structure calculations by Bendt and Zunger[63] and Ho, Ihm, and Joannopoulos.[58] Storing the full $N\times N$ Jacobian matrix is rarely possible for large self-consistency problems, but in the

past few years several authors were able to derive modified algorithms which require only the storage of a few N-dimensional vectors for each iteration.[64,60] Johnson's approach is such a reformulation of a modified Broyden scheme proposed by Vanderbilt and Louie,[59] in which all information obtained in previous steps is kept during the update of the inverse of the Jacobian matrix. It was demonstrated, for instance, in Ref. 59 that this scheme is superior to Broyden's second method, and in several tests we have seen that especially for surfaces the improvement of Pulay's method over Broyden's second method can be substantial (see also our recent work[39]).

A. Pulay mixing

In this section we will shortly review Pulay's approach, using a notation which is more natural for the problem of charge density mixing. In the scheme of Pulay[53] the input charge density and the residual vectors are stored for a number of mixing steps. A new optimal input charge density is obtained in each step as a linear combination of the input charge densities of all previous steps,

$$\rho_{\mathrm{in}}^{\mathrm{opt}}=\sum_i \alpha_i \rho_{\mathrm{in}}^i. \tag{52}$$

Assuming linearity of the residual vector with respect to the input charge density ρ_{in}, the residual vector at $\rho_{\mathrm{in}}^{\mathrm{opt}}$ is given by

$$R[\rho_{\mathrm{in}}^{\mathrm{opt}}]=R\left[\sum_i \alpha_i \rho_{\mathrm{in}}^i\right]=\sum_i \alpha_i R[\rho_{\mathrm{in}}^i]. \tag{53}$$

The optimal new charge density must minimize the norm of the residual vector,

$$\langle R[\rho_{\mathrm{in}}^{\mathrm{opt}}]|R[\rho_{\mathrm{in}}^{\mathrm{opt}}]\rangle, \tag{54}$$

with respect to α_i subject to the constraint

$$\sum_i \alpha_i=1, \tag{55}$$

which conserves the number of electrons. These equations are very similar to those given in Sec. III B, only the functional form of the constraint has changed. The optimal α_i is now given by

$$\alpha_i=\frac{\Sigma_j A_{ji}^{-1}}{\Sigma_{kj} A_{kj}^{-1}} \quad \text{with} \quad A_{ij}=\langle R[\rho_{\mathrm{in}}^j]|R[\rho_{\mathrm{in}}^i]\rangle. \tag{56}$$

We were able to show[39] that this update is equivalent to a quasi-Newton scheme in which the inverse of the Jacobian matrix is given in each iteration by

$$\mathbf{G}^m=\mathbf{G}^1-\sum_{k,n=1}^{m-1} \overline{A_{kn}^{-1}}(\mathbf{G}^1|\Delta R^n\rangle+|\Delta\rho^n\rangle)\langle\Delta R^k|, \tag{57}$$

where

$$\overline{A_{kn}}=\langle\Delta R^n|\Delta R^k\rangle \tag{58}$$

and

$$\Delta\rho^i=\rho_{\mathrm{in}}^{i+1}-\rho_{\mathrm{in}}^i \;\; \Delta R^i=R[\rho_{\mathrm{in}}^{i+1}]-R[\rho_{\mathrm{in}}^i], \tag{59}$$

with $i<m$. Having an explicit formula for the inverse of the Jacobian matrix is convenient because it has some physical meaning (it is the inverse of the charge dielectric matrix discussed above). It is easy to see that $\mathbf{G}^m$ defined in Eq. (57) fulfills the equation

$$\mathbf{G}^m|\Delta R^i\rangle=-|\Delta\rho^i\rangle, \quad \forall i<m. \tag{60}$$

This indicates that $\mathbf{G}^m$ is the best approximation of the exact inverse of the Jacobian matrix $\mathbf{J}^{-1}$ in the subspace searched up to now. This can be seen by subtracting Eq. (49) for iteration i and $i+1$ and multiplying on the left with $\mathbf{J}^{-1}$ giving $\mathbf{J}^{-1}|\Delta R^i\rangle=-|\Delta\rho^i\rangle$.

B. Preconditioning, metric, and convergence

There are some subtle details which can help to improve convergence: The choice of the trial step, which is determined by $\mathbf{G}^1$, and an optimized metric for evaluating the scalar products $\langle\ |\ \rangle$ are important.

Regarding Eq. (57) we can see that $\mathbf{G}^1$ is the initial approximation for $\mathbf{J}^{-1}$. It has the same ''meaning'' as the preconditioning matrix in the iterative matrix diagonalization, the closer $\mathbf{G}^1$ is to the exact inverse of the Jacobian matrix, the faster the convergence will be. For bulk materials we usually resort to a matrix proposed by Kerker,[65]

$$\mathbf{G}_q^1=A\frac{q^2}{q^2+q_0^2}, \tag{61}$$

which is a simple diagonal model for the inverse of the dielectric matrix in metallic systems and can be derived easily combining Eqs. (50) and (51). For large q (rapid variations of the charge density) the exact dielectric matrix converges to 1, therefore A should be close to one (rapid oscillations in the potential are not screened). This matrix has the advantage of correctly damping the oscillations in the low-q components of the charge density, i.e., for small wave vectors the function behaves like Aq^2/q_0^2 and mixes only a small amount of the output charge density to the input charge density. For large wave vectors q, a simple linear mixing with the linear mixing parameter A is done. Pulay's method is rather insensitive to the choice of the parameters for the initial mixing: $A=0.8$ and $q_0=1.5$ Å^{-1} is usually satisfactory. For magnetic systems, molecules, and surfaces an initial linear mixing with $A=0.2$ is sometimes faster.

In addition, it is relatively easy to optimize the preconditioning matrix if optimal performance is required. As stated previously the convergence depends strongly on the eigenvalue spectrum of the Jacobian matrix (49). To be more precise, for a preconditioned algorithm it depends on the eigenvalue spectrum of $\mathbf{G}^1\mathbf{J}$, respectively, $\mathbf{G}^{1^{-1}}\mathbf{J}^{-1}$. This is equivalent to determining the eigenvalues of the generalized non-Hermitian eigenvalue problem,

$$\mathbf{G}|x^i\rangle=\Gamma^i\mathbf{G}^1|x^i\rangle. \tag{62}$$

During the self-consistency steps we get a steadily improving approximation for $\mathbf{G}$ and for each iteration we solve the $(m-1)$ dimensional eigenvalue problem,

$$(\mathbf{G}^m-\mathbf{G}^1)|x^i\rangle=(\Gamma^i-1)\mathbf{G}^1|x^i\rangle, \tag{63}$$

to get the power spectrum Γ. $\mathbf{G}^1$ is optimal if the mean eigenvalue is 1, and if the width of the eigenvalue spectrum is minimal. For an initial linear mixing an optimal setting for A can be found easily by setting $A_{\rm opt}=A_{\rm current}/\Gamma_{\rm mean}$. The optimal $A_{\rm opt}$ is always orders of magnitudes larger than that one used in a straight linear mixing without an update of $\mathbf{G}$. For the Kerker scheme we usually optimize only q_0 and keep A fixed to 0.8. But we want to emphasize once again that the optimization of the mixing parameter is rarely necessary as we have demonstrated also in Ref. 39. For an initial linear mixing the linear mixing parameter can vary by a factor of 10 without significantly changing the convergence. Only very small mixing parameters ($A<0.02$) usually used in a straight mixing should be avoided because this would result in a significantly slower search of the Hilbert space and then rapid oscillations, which are not screened, will converge very slowly.

The second thing which is usually important is the introduction of a reasonable metric. We have found that the inclusion of a weighting factor

$$f_q=\frac{q^2+q_1^2}{q^2} \tag{64}$$

in the evaluation of the scalar products, i.e.,

$$\langle A|B\rangle:=\sum_q f_q A_q^* B_q, \tag{65}$$

can improve the convergence considerably for complex metallic systems. This function is inspired by the fact that due to the Hartree potential the contributions for small wave vectors are more important and critical than contributions for large wave vectors. The choice of q_1 is relatively unimportant and we set q_1 in a way that the shortest wave vector is weighted 20 times stronger than the longest wave vector. We have checked that the metric and the initial approximation for the inverse of the Jacobian matrix improve the convergence independently. Anyway the introduction of a metric seems to be more important than the setting for the initial matrix $\mathbf{G}^1$.

At this point, we also want to make clear that a considerable difference between charge density mixing and potential mixing exists. Taking into account only the Hartree term the potential is given by

$$V(q)\propto\frac{1}{q^2}\rho(q),$$

therefore the metric for the evaluation of scalar products differs by a factor of $1/q^4$ in both cases.

Finally, we are frequently confronted with very large systems with FFT grids containing up to $64\times64\times64$ points, which are necessary to describe the rather hard augmentation charges of transition metals. Storage of all information from all previous steps would exceed the central memory capacities even for the new efficient mixing schemes. A rather simple solution to this problem exists: Because the dielectric matrix converges to 1 for large q, no mixing is necessary for large wave vectors q, i.e., it is possible to set

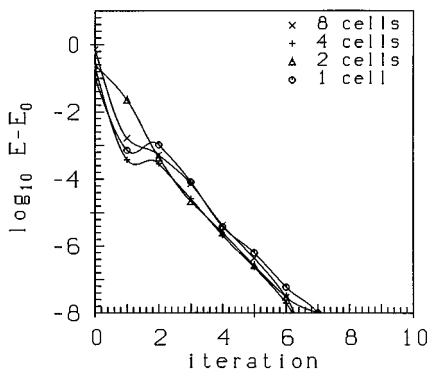

FIG. 2. Convergence of the total free energy E (in eV) for the RMM-DIIS algorithm for the cd-C, non-self-consistent case (no division by the number of atoms has been done). ''one cell'' corresponds to a cubic supercell containing eight atoms, for ''x cells'' the cell has been multiplied in one direction x times. The symmetry was destroyed by adding random vectors to the positions (see text).

$$\rho_{\rm in}^{m+1},_q=\rho_{\rm out}^m,_q \tag{66}$$

without any loss of efficiency, and only a relatively small number of grid points must be treated with Pulay's method; usually we take all grid points which are also contained in the plane-wave basis set $[\hbar^2|\mathbf{q}|^2/(2m_e)<E_{\rm cut}$, for reasons of simplicity we take a box instead of this sphere]. In Sec. V B, we will demonstrate for two cases that the number of iterations required to converge to a specific precision is almost independent of the system size.

V. CONVERGENCE FOR INSULATING AND METALLIC SYSTEMS

A. Convergence behavior of the iterative matrix diagonalization method

In this section, we will demonstrate that for our matrix diagonalization scheme the required number of iterations does not depend on the system size. We have chosen one metallic and one insulating example. For all calculations presented in Sec. V A the charge density and hence the Hamiltonian were kept fixed.

1. Insulating system

The first system is cubic diamond (cd), our smallest test cell consists of eight atoms in a cubic supercell. We increased the system size by multiplying the cell in one direction twice, four times, and eight times. All symmetry (including the translational symmetry due to the replication of the original cell) was destroyed by randomizing the initial positions by 2% of the lattice parameter. For all cells only the Γ point was used and a cutoff of 280 eV was chosen. With an US-PP this cutoff is sufficient to obtain an excellent description of the equilibrium ground-state properties of cubic diamond.[26] Only occupied bands were used in the calculation. As already stated this is not our usual (and also not the generally recommended) practice—normally we include a certain number of empty bands even for semiconductors and insulators. To show that our method also works for the minimum number of bands, we have changed the setup accordingly. Nevertheless, due to this setup we had some dif-

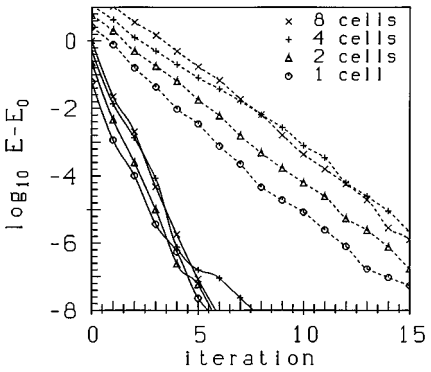

FIG. 3. Convergence of the total free energy E (in eV) for the RMM-DIIS algorithm (full curve) and for the CGa scheme (broken curve) for fcc-Fe, non-self-consistent case. ''one cell'' corresponds to a cubic supercell containing four atoms, for ''x cells'' the cell has been multiplied in one direction x times. The symmetry was destroyed by adding random vectors to the positions (see text). The RMM-DIIS algorithm is a band-by-band algorithm, whereas the CGa scheme optimizes all wave functions at the same time (see Sec. V B 2).

ficulties in the initialization of the wave functions. For the RMM-DIIS method it is much safer to include a certain number of empty bands, because this guarantees that the correct ground state is spanned by the initial wave functions.

In Fig. 2 the convergence for different system sizes is shown, and it is clearly visible that there is no dependence on the length of the cell. For this calculation the initial electronic configuration was calculated using a random initialization and 3 CG sweeps over all bands on the wave functions (i.e., three subspace rotations, three sweeps over all bands, two optimizations of each wave function per sweep with the CG algorithm).

2. Open-shell transition metals

The second test system is an open-shell transition metal. We have chosen paramagnetic fcc-Fe and a cutoff of 250 eV. The smallest test cell contains four atoms in a cubic supercell. The system size was increased by multiplying the cell in one direction twice, four times, and eight times. For the smallest cell a k-point mesh with $4\times4\times4$ Monkhorst Pack special points[66] was used. The number of k points was decreased to $4\times4\times2$ for the double cell, $4\times4\times1$ for the four times repeated cell, and $2\times2\times1$ for the largest cell, a smearing of $\sigma=0.3$ was used in the calculation (for the large cell we used this k-point setting to get results reasonably fast). As previously all symmetry was destroyed by randomizing the initial positions by 5% of the lattice parameter. The initial wave functions were chosen in the same way as above. To get a reasonable convergence it was necessary to include $1.5N_{\text{ions}}$ empty bands above the Fermi level. It can be seen in Fig. 3 that the convergence is once again practically independent of the system size (the results for the CGa calculation will be discussed later in Sec. V B 2).

B. Convergence behavior for self-consistent calculations

1. Insulating system

For the insulating system, the number of iteration steps does not increase seriously if self-consistency is switched on

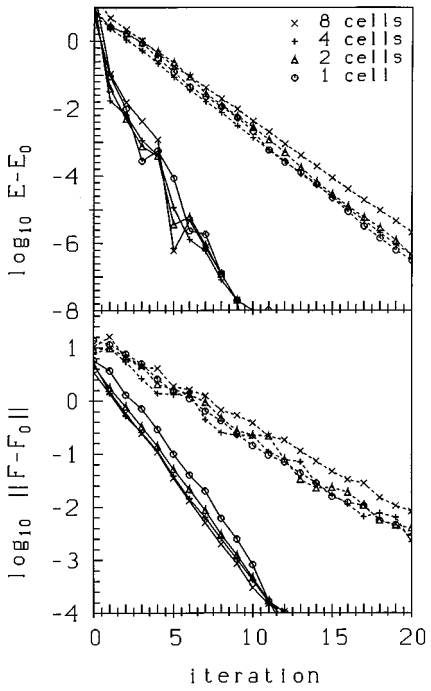

FIG. 4. Top panel—convergence of the total free energy E (in eV) for the RMM-DIIS (full curve) and the CGa algorithm (broken curve) for cd-C, self-consistent case. The RMM-DIIS algorithm uses a self-consistency cycle, whereas the CGa algorithm minimizes the KS functional directly. Lower panel—convergence of the forces (in eV/Å) for both methods.

(see Fig. 4). The charge sloshing is rather weak in this system, so the overall convergence is mainly determined by the convergence progress of the band-structure term. We used the Kerker matrix for the initial approximation of the inverse Jacobian matrix (61) and found an optimal mixing parameter of $q_0=1.0$ Å^{-1}. Mind that the Kerker matrix does not possess the correct limit for small q for insulating systems, but as long as the length of the system is not too large this does not pose a problem (for very large system a lower bound for the model dielectric function might be required for insulating systems). With this setup the range of eigenvalues of Eq. (63) remained practically unchanged with the system size, and ranged between 0.5–1.4. We also tested a setup in which $\mathbf{G}^1$ was initially set to a linear mixing with $A=1.0$. In this case the obtained eigenvalue spectrum corresponds to the eigenvalue spectrum of the inverse of the real dielectric matrix $\mathbf{J}^{-1}$. We found a spectrum ranging between 0.35 and 1 for the smallest cell and 0.15–1 for the largest cell (the minimal eigenvalue of $\mathbf{G}$ clearly converged to a finite value of ≈0.15). The convergence remained quite good. Instead of 12 iterations, 17 iterations were required for the largest cell to obtain an accuracy of 10^{-8} eV. In accordance with the range of eigenvalues the maximum linear mixing parameter for a straight mixing without an update of $\mathbf{G}$ was 0.7 for the smallest cell and 0.3 for the largest cell (i.e., twice the minimum eigenvalue of $\mathbf{G}$), the convergence remained reasonable even with a simple straight mixing without an update of $\mathbf{G}^1$. We have also done calculations for a $3\times3\times3$ cell containing 216 atoms. For this large cell also only 12 iterations were required to converge to the same precision.

TABLE I. Time necessary to perform one iteration for a carbon ensemble containing N_{cell} cells, respectively, N_{ions} ions for several algorithms on an IBM RS 6000/Model 590 (a CRAY C90 is approximately four times faster). The timing is given for a code which takes into account that $C_q = C^*_{-q}$, if the Γ point only is used for the k-point sampling. RMM refers to the residual minimization band-by-band scheme with mixing (Sec. III B), CG to the conjugate gradient band-by-band scheme with mixing (Sec. III A), CGa is the conjugate gradient scheme applied directly to all degrees of freedom of the KS functional.

N_{cell}	N_{ions}	RMM	CG	CGa
1	8	1.0	1.0	1.2
2	16	3.0	3.0	3.2
4	32	10.0	10.0	9.0
8	64	35.0	50.0	32.0
$3\times3\times3$	216	410.0	800.0	

To show that our method compares very well with methods which minimize the Kohn-Sham functional directly, we have included results for a method which optimizes all degrees of freedom simultaneously using a conjugate gradient algorithm (we will call this algorithm CGa not to be mixed up with the CG algorithm discussed in Sec. III A). A detailed discussion of our ''direct'' algorithm can be found in Ref. 39. It is very similar to the algorithms proposed by Gillan[16] and Arias, Payne, and Joannopoulos,[17] but it includes a consistent update of the partial occupancies and performs a subspace rotation at each step so that metallic systems can be treated too (all degrees of freedom are treated with the CG algorithm). One big advantage of this CGa algorithm over several other direct algorithms (for instance, that one of Tassone *et al.*[8]) is that it is entirely parameter free.

In Fig. 4 it can be seen that the direct algorithm requires approximately twice to three times as many iterations as the algorithm based on the self-consistency cycle. The initial error for the CGa method is slightly larger than for the mixing methods because the initial charge density has to be set equal to the charge density calculated from the initial wave functions. In the mixing methods we can start with the linear superposition of the atomic pseudocharge densities of the constituents. In Table I we show the timing for both schemes. It can be seen that the scaling is better than quadratic, and that both schemes take approximately the same time. The reason for the ''super''-quadratic behavior for small cells lies in the fact that all operations related to the augmentation part scale linearly with the system size, and dominate for very small systems. Because the CGa scheme has to evaluate the augmentation terms twice, it is slower for small systems and takes approximately 90% of the time of the RMM algorithm for large systems. We want to point out that using the Γ point only for small systems containing only 8–32 atoms is rather unrealistic. If several k points are used or if the system size is sufficiently large the costs for the treatment of the augmentation part are negligible. In Table I we have also included the timing for the sequential CG scheme. For small cells it requires the same CPU time as the RMM scheme because the costs for the orthonormalization are negligible, whereas for the $3\times3\times3$ cell the RMM scheme is faster by a factor two.

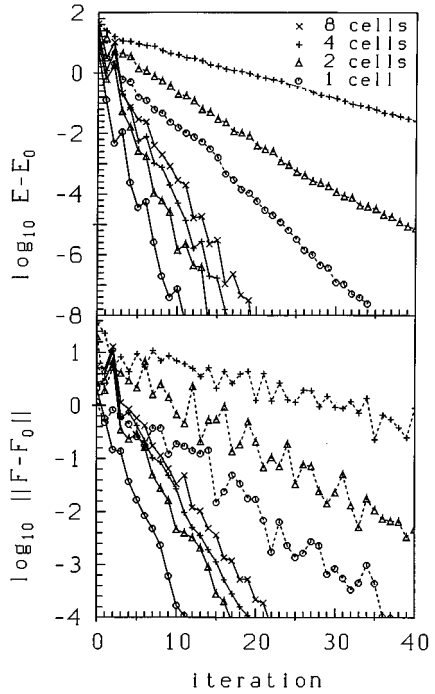

FIG. 5. Top panel—convergence of the total free energy E (in eV) for the RMM-DIIS (full curve) and the CGa algorithm (broken curve) for fcc-Fe, self-consistent case. The RMM-DIIS algorithm uses a self-consistency cycle, whereas the CGa algorithm minimizes the KS functional directly. Lower panel—convergence of the forces (in eV/Å) for both methods.

To demonstrate that our scheme also gives very accurate forces we have included one panel in Fig. 4 which shows the convergence of the forces for both methods. It can be seen that the convergence of the forces is extraordinary for our method. After ten iterations more than three digits after the comma are correct. To get the same accuracy from the CGa scheme, the number of iterations has to be at least doubled.

2. Open-shell transition metal

For the transition metal, the charge sloshing is definitely much stronger. Once again the parameter q_0 was optimized, in this case we found $q_0 = 4.0$ Å^{-1} to be the optimal choice, the range of the eigenvalues of Eq. (63) was, except for the small cell, between 0.2 and 3.0 and remained almost *unchanged* when the system size was increased.

For the open-shell transition metal (see Fig. 5) it can be seen that the convergence slows down slightly when the size of the system is increased. Going from the smallest system to the largest system the number of iterations increases by a factor of two. The smallest cell actually shows only very weak charge sloshing, difficulties with the sloshing start in the two times repeated cell (see the strong increase of the error of the energy and the forces in the second iteration). If we compare this cell with the largest cell, we can see that the increase in the number of iterations is very small. We also performed a calculation for a $3\times3\times3$ cell containing 108 atoms using the Γ point only. In this case, the convergence was similar to the eight times repeated cell.

Our results differ significantly from the theoretical predictions of Annett,[34] who stated that the number of iterations

should increase quadratically with the length of the system for SC methods. The main reasons for the difference between our observed convergence behavior and the theoretical predictions of Annett have their origin in the following two points: First, we use a quadratically convergent method, second, we use a diagonal approximation for the dielectric matrix, which, in practice, removes problems arising from the quadratic divergence of the Hartree term in **J**.

To illustrate this, we tried a simple mixing without an update of **G** for this system. For the smallest cell the maximum linear mixing parameter is 0.08 and 80 iterations are required to converge to an accuracy of 10^{-8} eV. For the double cell the mixing parameter has to be decreased by a factor of 2 to 0.04 and now almost 160 iterations are required. For the four times repeated cell, the mixing parameter must be less than 0.015 and more than 500 iterations are necessary. This scenario is indeed similar to that discussed in Ref. 34. The situation already improves if a fixed Kerker-like approximation is used for **G** (no update of **G**). Optimal performance is found with $q_0=6.0$ and $A\approx 0.8-1$, now 24 and 34 iterations are necessary to converge to the required precision for the small, respectively, the doubled cell.

We have also checked the magnitude of the maximum linear mixing parameter by performing calculations with an update of **G** for an initial linear mixing ($\mathbf{G}^1=A$). The obtained eigenvalue spectrum of **G** was in agreement with the maximal linear mixing parameter. We found an eigenvalue spectrum of 0.045–1 for the small cell, 0.025–1 for the doubled cell, 0.008–1 for the four times repeated cell, and 0.0025–1 for the largest cell. This behavior can be explained assuming an asymptotic behavior of $\mathbf{J}^{-1}$ according to Eq. (61). The number of iterations in this case was 17 for the smallest and 35 for the largest cell. One might expect in this case that the number of iterations increases with the square root of the eigenvalue spectrum and therefore linearly with system length (see Sec. IV B), but this is not the case, since the eigenvalue spectrum is not continuous, i.e., doubling the cell creates only two additional small eigenvectors (corresponding to the new shortest reciprocal wave vector of the cell). All other new eigenvectors remain within the range of the eigenvalue spectrum already found for the nondoubled cell.

At this point we also want to make a comment on why the introduction of the metric improves the performance. In principle, we are dealing with a rather ill-conditioned problem — the eigenvalue spectrum of **G** can be very broad. The metric forces the mixing algorithm to converge the charge densities for short wave vectors (long range oscillations—corresponding to small eigenvectors) first. If the progress of the eigenvalue spectrum is monitored carefully it can be seen that the eigenvalue spectrum is indeed built up from the smallest (and most critical) eigenvalues, these eigenvalues are converged after a few steps, and after modeling this complicated ''charge-sloshing'' part of the dielectric matrix the short range behavior where no sloshing occurs is determined. If no metric is included all eigenvalues converge more or less at the same time, resulting in a significantly slower convergence for large cells.

We have also tested the CGa algorithm for metallic systems and found a quite bad performance especially when the cell length was increased (results for the largest cell are not included because the convergence was too bad). The optimal step size for the line minimization becomes gradually smaller with the system length (it decreases approximately by a factor of two if the cell length is doubled). The number of required steps is almost proportional to the system length, making the scheme generally much slower than our scheme based on the SC cycle (the scaling is the same as that one predicted in Ref. 34). We have checked that the only reason for the slow convergence is the charge sloshing, i.e., if we use the CGa algorithm to determine the eigenvalues for a fixed Hamiltonian then the required number of steps is independent of the system size (see Fig. 2). But even in this case the band-by-band RMM-DIIS or band-by-band CG scheme is superior. This is not astonishing and the reason for this is that the RMM-DIIS (or band-by-band CG scheme) iterates the highest occupied bands usually four times per step, whereas the CGa scheme optimizes each band only once per step.

To solve the charge-sloshing problem in the CGa scheme, it is necessary to build in a similar ''charge preconditioning'' as in the methods based on the SC cycle. But because a separate treatment and preconditioning of the input charge density would destroy the extremal properties of the KS functional (see Sec. II D) this could be done, for instance, only within a steepest-descent scheme or Tchebycheff's acceleration scheme (merely a combination of Ref. 8 and Ref. 11). The main problem within this approach would be the determination of all convergence parameters, and therefore we have not tested this approach yet. On the other hand, it would be very easy to implement the required changes in an existing Car-Parrinello code.

Finally, it must be realized that the SC methods allow us to store the change of the charge density for a large number of steps, whereas the CGa method stores only the gradient and the search direction of the last step. If the line minimization is not done with very high accuracy the net convergence will slow down considerably. A scheme which overcomes this difficulty was introduced by Hutter, Lüthi, and Parrinello,[67] but it requires the storage of a large set of wave functions making it rather inconvenient for large systems.

TABLE II. Time necessary to perform one iteration for the Fe ensembles containing N_{ions} atoms for several algorithms on an IBM RS 6000/ Model 590 (a CRAY C90 is approximately four times faster). N_{cell} is the number of fcc supercells, N_{ions} is the number of ions, and $N_{k\ \text{points}}$ is the number of k points. RMM refers to the residual minimization method band-by-band scheme with mixing, CGa is the conjugate gradient scheme applied directly to the KS functional. The timing for the $3\times3\times3$ cell (Γ point only) is given for a version which takes into account that $C_q=C^*_{-q}$. The sequential CG algorithm with mixing would require 900 s/iteration. For the eight times repeated cell one iteration with the Γ point only would take 27 s for the RMM scheme.

N_{cell}	N_{ions}	$N_{k\ \text{points}}$	RMM	CGa
1	4	32	21.0	16.0
2	8	16	39.0	32.0
4	16	8	80.0	65.0
8	32	2	92.0	
$3\times3\times3$	108	1	360.0	

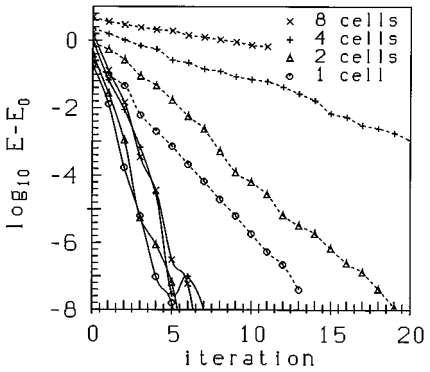

FIG. 6. Convergence of the total free energy E (in eV) for the RMM-DIIS (full curve) and the CGa (broken curve) algorithm for fcc-Al, self-consistent case. The RMM-DIIS algorithm uses a self-consistency cycle, whereas the CGa algorithm minimizes the KS functional directly.

In Table II, the timings for the calculations presented here are shown. Once again the RMM-DIIS scheme is only slightly slower than the CGa scheme. The scaling of our code is quadratic and if we take into account that the number of k points can be decreased linearly with the system size, then we can clearly see an almost linear scaling.

3. Simple metals

The mixing scheme of Kerker is especially efficient for nearly free-electron metals. Here, accurate models for the dielectric matrix exist which, for instance, allow us to calculate accurate pair potentials using a second-order perturbation theory.[68] Because the functional form of the inverse of the dielectric function of a nearly free-electron gas closely resembles Eq. (61), it is not surprising that a fast convergence can be reached even without an update of **G**. With an update of **G**, the convergence is at least as good as that for the insulating C system (although charge sloshing is considerable in simple metals). We have found such a favorable scaling for Al, l-Si, l-Ge, l-Te, and l-Se, and even for Cu or Ag we obtained a very rapid convergence to the ground state (the density of states at the Fermi level seems to be the critical quantity). To demonstrate the feasibility of the mixing scheme in this respect, we show the convergence for fcc-Al in Fig. 6. The cells and the k-point sets are the same as for fcc-Fe (all symmetry was once again destroyed by displacing the atoms by random vectors, compare Sec. V A 2). Only eight iterations are required to converge to the electronic ground state, and the convergence is independent of the system size. Mind once again, that one of the key points for fast convergence is in this case the use of the Kerker matrix for $\mathbf{G}^1$. The convergence gets significantly slower if a simple linear approximation is used for the initial approximation of the inverse of the Jacobian matrix ($\mathbf{G}^1 = A$), because the smallest eigenvalue of the inverse of the charge dielectric matrix decreases for simple metals almost strictly with the square of the length of the system. This is also the reason why the performance of the CGa algorithm drops significantly if the system size is increased.

VI. CONCLUSION

We have presented a detailed description of an efficient iterative matrix diagonalization scheme and an efficient charge density mixing scheme. Both schemes are based on the residual minimization method—direct inversion in the iterative subspace (RMM-DIIS). The advantage of the matrix diagonalization based on the RMM-DIIS scheme over the minimization of the Rayleigh quotient with a standard conjugate gradient scheme lies in the fact that all $O(N^3)$ operations can be reduced to their absolute minimum. Altogether our scheme scales almost quadratically with the system size up to systems containing 1000 electrons. At this point $O(N^3)$ operations become important. The application of the diagonalization scheme is of course not limited to a plane-wave basis set only, whenever the evaluation of $(\mathbf{H}-\epsilon\mathbf{S})|\phi\rangle$ is an $O(N^2)$ operation the RMM-DIIS scheme should outperform other techniques. For instance, testing this technique for the diagonalization of large sparse matrices might be interesting.

For the charge density mixing we have discussed the use of Pulay's DIIS minimization method and the importance of including a reasonable model for the dielectric matrix. It was shown for three examples (one insulating, two metallic) that the number of iterations does not increase dramatically with the length of the supercell. This result contradicts the theoretical investigation of Annett[34]—the main reasons for the contradiction are given by the following points: First, Annett has done an analysis at zero temperature only, we are using ''finite temperature'' LDA. Second, Annett did not realize that Broyden-like methods are indeed quadratically convergent. Third, it is relatively easy to deal with the $1/q^2$ divergence of the Hartree term by using an appropriate model for the dielectric function in the update of the charge density. We think that our findings have also significant impact on order $O(N)$ methods because we have demonstrated that methods relying on the determination of the self-consistent charge density can be very competitive to methods doing a direct minimization of the KS functional: at finite temperature an order $O(N)$ algorithm relying on SC methods should be possible even for metals.

We have also compared our SC method with methods doing a direct minimization of the KS functional. Especially for metals we have found that the SC methods are significantly faster and show a much better scaling with the system size than methods doing a direct minimization using, for instance, a CG algorithm. This is not only true for the convergence of the total energy but also for the convergence of the forces. We have not done a comparison of our method with the recent formulation of the Car-Parrinello technique given by Tassone,[8] but we expect that Tassone's algorithm behaves very similar to the CGa scheme used here. Indeed the CGa algorithm and Tassone's algorithm both show a quadratic convergence,[8,11,34] the disadvantage of Tassone's algorithm is that it requires the determination of two parameters to obtain optimal performance. Finally, if we consider that all direct methods also require much more storage capacity, it becomes questionable why these approaches have been emphasized so strongly during the past few years. Our program can be used efficiently for systems containing up to 100 transition metals or 200 simple elements on workstations with 256 Mbytes. Finally, we want to emphasize at the end that the algorithms discussed here are extremely reliable and have been used for more than two years; up to now we have not

found a system where no convergence to the correct ground state was obtained.

ACKNOWLEDGMENTS

We are indebted to J. Hafner; he has encouraged this work and he has supplied invaluable help in a lot of discussions. We also have to thank M.C. Payne, who has supplied a copy of his early *ab initio* plane-wave program (local pseudopotentials only, version July 1989). The final stage of this work has been undertaken within the ''GdR Dynamique Moléculaire Quantique Appliquée à la Catalyse,'' a joint project of CNRS, Technical University Vienna, and Institut Français du Pétrole.

APPENDIX: FAST GRAM-SCHMIDT ORTHONORMALIZATION

Usually the Gram-Schmidt orthonormalization is done by sequentially orthogonalizing all bands to one selected band. The outer loop starts with the first band, normalizes that one and then orthogonalizes all other bands to the current band. The second band is then normalized and all remaining bands are orthogonalized to the second band and so on. For this algorithm the limiting factor is not the maximum number of floating-point operations per second but the maximum data transfer bandwidth from the main computer memory to the central processing unit (CPU).

The solution is straightforward: We first calculate the overlap matrix $\mathbf{S}$,

$$\overline{\mathbf{S}}_{ij} = \langle \phi_i | \mathbf{S} | \phi_j \rangle, \tag{A1}$$

and then we perform a Choleski decomposition of $\overline{\mathbf{S}}$, i.e.,

$$\overline{\mathbf{S}} = \mathbf{L}\mathbf{U}. \tag{A2}$$

The final orthonormalized orbitals are then given by

$$|\overline{\phi_j}\rangle = \sum \mathbf{U}^{-1}_{ij} |\phi_i\rangle. \tag{A3}$$

For the matrix-matrix operations (and partially also for the less expensive Choleski decomposition of $\mathbf{S}$ and the inversion of the matrix $\mathbf{U}$) the cache can be used very efficiently, using ''blocked'' schemes. Mind that such optimizations are impossible for the sequential CG scheme (involving only vector-vector operations). Efficient use of the cache allows the CPU to run almost at its peak floating-point performance. The performance improvement over the standard sequential Gram-Schmidt scheme is usually considerable and ranges between a factor of two (for IBM/RS6000 workstation and CRAY C90 supercomputers) and four (for SGI, DEC ALPHA, and SUN workstations).

[1] W. Kohn and L. Sham, Phys. Rev. **140**, A1133 (1965).
[2] R.O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).
[3] R.P. Feynman, Phys. Rev. **56**, 340 (1939).
[4] D. Vanderbilt, Phys. Rev. B **41** 7892 (1990).
[5] P.E. Blöchl, Phys. Rev. B **50**, 17 953 (1994).
[6] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[7] M.C. Payne, J.D. Joannopoulos, D.C. Allan, M.P. Teter, and D.H. Vanderbilt, Phys. Rev. Lett. **56**, 2656 (1986).
[8] F. Tassone, F. Mauri, and R. Car, Phys. Rev. B **50**, 10 561 (1994).
[9] A. Williams and J. Soler, Bull. Am. Phys. Soc. B **32**, 562 (1987).
[10] Guo-Xin Qian, M. Weinert, G.W. Fernando, and J.W. Davenport, Phys Rev. Lett. **64**, 1146 (1990).
[11] P.H. Dederichs and R. Zeller, Phys. Rev. B **28**, 5462 (1983); H. Akai and P.H. Dederichs, J. Phys. C **18**, 2455 (1985).
[12] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes* (Cambridge University Press, New York, 1986).
[13] E. Polak, *Computational Methods in Optimization* (Academic Press, New York, 1971).
[14] M.P. Teter, M.C. Payne, and D.C. Allan, Phys. Rev. B **40**, 12 255 (1989).
[15] I. Stich, R. Car, M. Parrinello, and S. Baroni, Phys Rev. B **39**, 4997 (1989).
[16] M.J. Gillan, J. Phys. Condens. Matter **1**, 689 (1989).
[17] T.A. Arias, M.C. Payne, and J.D. Joannopoulos, Phys. Rev. Lett. **69**, 1077 (1992).
[18] M.P. Grumbach, D. Hohl, R.M. Martin, and R. Car, J. Phys. Condens. Matter **1**, 1999 (1994).
[19] R.D. King-Smith and D. Vanderbilt, Phys. Rev. B **49**, 5828 (1994).
[20] J.M. Holender, M.J. Gillan, M.C. Payne, and A.D. Simpson, Phys. Rev. B **52**, 967 (1995); I. Stich, M. Parrinello, and J.M. Holender, Phys. Rev. Lett. **76**, 2077 (1996).
[21] G. Kresse and J. Hafner, Phys. Rev. B **47**, RC558 (1993).
[22] G. Kresse and J. Hafner, Phys. Rev. B **48**, 13 115 (1993).
[23] G. Kresse, Proceedings of the 6th International Conference on the Structure of Non–Crystalline Materials (NCM6), Prague, 1994 [J. Non-Cryst. Solids **192/193**, 222 (1995)].
[24] G. Kresse and J. Hafner, Phys. Rev. B **49**, 14 251 (1994).
[25] G. Kresse, Proceedings of the 9th International Conference on Liquid and Amorphous Metals (LAM9), Chicago, 1995 [J. Non-Cryst. Solids (to be published)].
[26] J. Furthmüller, J. Hafner, and G. Kresse, Europhys. Lett. **28**, 659 (1994); Phys. Rev. B **53**, 7334 (1996).
[27] J. Furthmüller, G. Kresse, J. Hafner, R. Stumpf, and M. Scheffler, Phys. Rev. Lett. **74**, 5084 (1995).
[28] A. Eichler, J. Hafner, J. Furthmüller, and G. Kresse, Surf. Sci. **346**, 300 (1996).
[29] G. Kresse, J. Furthmüller, and J. Hafner, Europhys. Lett. **32**, 729 (1995).
[30] K. Laasonen, A. Pasquarello, R. Car, C. Lee, and D. Vanderbilt, Phys. Rev. B **47**, 10 142 (1992).
[31] G. Kresse, Ph.D. thesis, Technische Universität Wien, 1993.
[32] N.D. Mermin, Phys. Rev. **137**, A1441 (1965).
[33] G. Kresse and J. Hafner, J. Phys. Condens. Matter **6**, 8245 (1994).
[34] J.F. Annett, Comput. Mat. Sci. **4**, 23 (1995).
[35] M. Weinert and J.W. Davenport, Phys. Rev. B **45**, 13 709 (1992).
[36] R.M. Wentzcovitch, J.L. Martins, and P.B. Allen, Phys. Rev B **45**, 11 372 (1992).
[37] O. Jepsen and O.K. Andersen, Solid State Commun. **9**, 1763 (1971).

[38] P.E. Blöchl, O. Jepsen, and O.K. Andersen, Phys. Rev. B **49**, 16 223 (1994).
[39] G. Kresse and J. Furthmüller, Comput. Mat. Sci. (to be published).
[40] C.-L. Fu and K.-M. Ho, Phys. Rev. B **28**, 5480 (1983).
[41] M. Methfessel and A.T. Paxton, Phys. Rev. B **40**, 3616 (1989).
[42] A. De Vita and M.J. Gillan, J. Phys. Condens. Matter **3**, 6225 (1991); A. De Vita, Ph.D. thesis, Keele University, 1992; A. De Vita and M.J. Gillan (unpublished).
[43] J. Harris, Phys. Rev. B **31**, 1770 (1985).
[44] W.M.C. Foulkes, Ph.D. thesis, University of Cambridge, 1987.
[45] W.M.C. Foulkes and R. Haydock, Phys. Rev. B **39**, 12 520 (1989).
[46] P. Pulay, in *Modern Theoretical Chemistry*, edited by H.F. Schaefer (Plenum, New York, 1977); Mol. Phys. **17**, 197 (1969).
[47] S. Goedecker and K. Maschke, Phys. Rev. B **45**, 1597 (1992).
[48] P. Bendt and A. Zunger, Phys. Rev. Lett. **50**, 1684 (1983); M. Methfessel and M. van Schilfgaarde, Phys. Rev. B **48**, 4937 (1993).
[49] E.R. Davidson, in *Methods in Computational Molecular Physics*, Vol. 113 of *NATO Advanced Study Institute, Series C: Mathematical and Physical Sciences*, edited by G.H.F. Diercksen and S. Wilson (Plenum, New York, 1983), p. 95; Report on the Workshop *Numerical Algorithms in Chemistry: Algebraic Methods*, edited by C. Moler and I. Shavitt (University of California, Berkley, 1978), p. 49; J. Comput. Phys. **17**, 87 (1975).
[50] B. Liu, in Report on the Workshop *Numerical Algorithms in Chemistry: Algebraic Methods* (Ref. 49), p. 49.
[51] D.M. Bylander, L. Kleinman, and S. Lee, Phys. Rev. B **42**, 1394 (1990).
[52] B.N. Parlett, *The Symmetric Eigenvalue Problem* (Prentice Hall, Englewood Cliffs, NJ, 1980).
[53] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[54] D.M. Wood and A. Zunger, J. Phys. A **18**, 1343 (1985).
[55] P.E. Blöchl and R.M. Martin (private communication).
[56] T.A. Arias, M.C. Payne, and J.D. Joannopoulos, Phys. Rev. B **45**, 1538 (1992).
[57] R.D. King-Smith, M.C. Payne, and J.S. Lin, Phys. Rev. B **44**, 13 063 (1991).
[58] Kai-Ming Ho, J. Ihm, and J.D. Joannopoulos, Phys. Rev. B **25**, 4260 (1982).
[59] D. Vanderbilt and S.G. Louie, Phys. Rev. B **30**, 6118 (1984).
[60] S. Blügel, Ph.D. thesis, Aachen University, 1988.
[61] D.D. Johnson, Phys. Rev. B **38**, 12 087 (1988).
[62] C.G. Broyden, Math. Comput. **19**, 577 (1965).
[63] P. Bendt and A. Zunger, Phys. Rev. B **26**, 3114 (1982).
[64] G.P. Srivastava, J. Phys. A **17**, L317 (1984).
[65] G.P. Kerker, Phys. Rev. B **23**, 3082 (1981).
[66] H.J. Monkhorst and J.D. Pack, Phys. Rev. B **13**, 5188 (1976).
[67] J. Hutter, H.P. Lüthi, and M. Parrinello, Comput. Mat. Sci. **2**, 244 (1994).
[68] J. Hafner, *From Hamiltonians to Phase Diagrams* (Springer, Berlin, 1987).