

Model for energetics of solids based on the density matrix

Murray S. Daw

Theory Department, Sandia National Laboratories, Livermore, California 94551-0969

(Received 10 August 1992)

We propose a model for energetics of solids, based on a direct solution for the density matrix directly from the Hamiltonian. The approach is suitable for obtaining an approximate, finite-ranged density matrix in a basis of localized orbitals. This scheme scales linearly with the size of the system, and should represent an improvement over the much-used embedded-atom, cluster-functional, and bond-order potentials.

During the last decade there has been a flurry of activity in the field of atomistic simulations of materials properties. Most of these simulations are based on modeling the band structures of metals and semiconductors with tight-binding (TB) Hamiltonians.¹ Although reasonable model Hamiltonians exist for many materials, solving for the dynamics of large systems through direct diagonalization of the Hamiltonian suffers from a scaling problem: the time required for diagonalization scales like $O(N_a^3)$ where N_a is the number of atoms in the system.

The scaling problem has been addressed by the development of other schemes. The recursion method,² for example, provides an approximation to the density of states (DOS) and scales like $O(N_a)$. Other workers have modeled the energetics in terms of simple, explicit functions of the Hamiltonian.³ One such model, the embedded-atom method⁴ (EAM) (and the closely related Finnis-Sinclair method⁵ and effective-medium theory⁶) has been used quite successfully in the investigation of complicated defects (such as surfaces, grain boundaries, and dislocations) in the noble metals. Many-body interactions (sometimes in the form of "bond-order potentials") have likewise been devised for semiconductors.⁷⁻⁹ For metals from the central transition series, recent proposals¹⁰⁻¹² have focused on "cluster functionals" involving three- and four-body interactions.

This paper will describe a scheme based on the direct calculation of the density matrix from the Hamiltonian. This approach is suitable for obtaining an approximate, finite-ranged density matrix. The computational time for this method scales linearly with the size of the system. We demonstrate the method for some well-studied examples, and conclude that the present method may prove superior to the recursion method and to many-atom potentials.

We start with a single-particle Hamiltonian \underline{H} connecting various sites (we consider here only s states, but the extension to p and d states is straightforward). The density matrix $\underline{\rho} = \theta(\mu \underline{1} - \underline{H})$ commutes with the Hamiltonian and has the same eigenstates. The eigenvalues of $\underline{\rho}$ are the occupation numbers for the states. The chemical potential μ is defined by fixing the total number of electrons, $N = \text{Tr}[\underline{\rho}]$, and the band energy¹³ is given by $E = \text{Tr}[\underline{H}\underline{\rho}]$. An interesting way of expressing the well-known variational theorem¹⁴ is to consider any symmetric matrix $\underline{\tilde{\rho}}$ and the associated energy function $\tilde{E} = \text{Tr}[\underline{H}\underline{\tilde{\rho}}]$. Then the true ground-state density matrix is given by minimizing \tilde{E} with respect to $\underline{\tilde{\rho}}$ with two constraints: (1) $N = \text{Tr}\underline{\tilde{\rho}}$ and (2) the eigenvalues of $\underline{\tilde{\rho}}$ must be bounded between 0

and 1. Of course, the nonholonomic constraint on the eigenvalues causes difficulty in implementing this variational theorem.

A solution to this difficulty can be obtained by considering a property of the smeared step function $\theta(\beta, z) = 1/(1 + \exp(-\beta z))$. Here, β is like an inverse temperature parameter, but we attach no physical significance to its value; we are, instead, trying to approach the asymptotic solution for infinite β . The first derivative with respect to β can be written in terms of θ itself: $\partial\theta/\partial\beta = z\theta_\beta(1 - \theta_\beta)$. We could solve numerically for $\theta(\beta, z)$ by starting with $\theta(0, z) = \frac{1}{2}$. Then we integrate to some finite β using the differential equation.

We can do the same for $\underline{\rho}$. In general, the solution to the variational problem is $\underline{\rho} = \lim(\beta \rightarrow \infty)\theta(\beta, \underline{z})$, with $\underline{z} = \mu \underline{1} - \underline{H}$. Thus, we start out at $\beta = 0$ with $\underline{\rho} = f \underline{1}$ (f is the fractional site occupancy) and solve $\partial\underline{\rho}/\partial\beta = \underline{z}\underline{\rho}(1 - \underline{\rho})$. As will be explained in the following, it is more generally useful to symmetrize the equation of motion as

$$\frac{\partial\underline{\rho}}{\partial\beta} = \left(\frac{1}{2}\right)\{\underline{z}, \underline{\rho}\} - \frac{(\underline{\rho}\underline{\rho}\underline{z} + \underline{\rho}\underline{z}\underline{\rho} + \underline{z}\underline{\rho}\underline{\rho})}{3}, \quad (1)$$

which is the basis for what follows.

The physical significance of the equation is given by considering the basis where $\underline{\rho}$ is diagonal with eigenvalues n_α . The eigenvectors of $\underline{\rho}$ depend on β , but because $\underline{\rho}$ is Hermitian, $d(\rho_{\alpha\alpha})/d\beta = d(n_\alpha)/d\beta$. Then Eq. (1) becomes

$$d(n_\alpha)/d\beta = (\mu - H_{\alpha\alpha})n_\alpha(1 - n_\alpha), \quad (2)$$

where $H_{\alpha\alpha} = \langle \alpha | H | \alpha \rangle$. Here it is clear that if $\mu > H_{\alpha\alpha}$, n_α will tend to 1 (filled states below the Fermi level), and if $\mu < H_{\alpha\alpha}$, n_α will tend to 0 (empty states above). At the fixed point, when the derivative vanishes, we have the condition that either $n_\alpha(1 - n_\alpha) = 0$ (in which case $n_\alpha = 0$ or 1) or $(\mu - H_{\alpha\alpha}) = 0$ (in which case we are at the Fermi level, where the occupation can be fractional). One can prove some useful properties of the equation of motion.

(1) The true density matrix is the asymptotic solution to Eq. (1) [that is, $\underline{\rho} = \theta(\infty, \underline{z})$ satisfies $\partial\underline{\rho}/\partial\beta = 0$]. This follows easily from $[\underline{\rho}, \underline{H}] = 0$ and $\underline{\rho}\underline{\rho} = \underline{\rho}$.

(2) Requiring that N is constant fixes $\mu = \text{Tr}[\underline{H}\underline{\rho}(1 - \underline{\rho})]/\text{Tr}[\underline{\rho}(1 - \underline{\rho})]$. This equation may look rather strange at first, because as $\beta \rightarrow \infty$, $\underline{\rho}\underline{\rho} \rightarrow \underline{\rho}$ and μ is the ratio of two numbers of going to zero. However, at finite β , $\underline{\rho}$ is not idempotent, and μ is well defined. Actually, this equation for μ is quite reasonable. Define the broadened delta function $\delta(\beta, z) = \partial\theta(\beta, z)/\partial z$. We can

show that $\delta(\beta, z) = \beta\theta_\beta(1 - \theta_\beta)$. Thus, $\rho(1 - \rho)$ is related to the density of states at the Fermi level, and μ is equal to the Fermi energy. (Of course, another way to go is to fix μ from the beginning and let the number N vary.)

(3) One can demonstrate readily from Eq. 2 that if the eigenvalues of $\rho(0, z)$ are bounded between 0 and 1, then they will remain similarly bounded for finite β .

(4) $dE/d\beta = 0$. Using Eq. (1),

$$\frac{\partial E}{\partial \beta} = \text{Tr} \left[\underline{H} \frac{\partial \rho}{\partial \beta} \right] = -\text{Tr} \left[\underline{z} \frac{\partial \rho}{\partial \beta} \right],$$

and working in the basis which diagonalizes $\underline{\rho}$, it is simple to rewrite $dE/d\beta$ as

$$-\frac{1}{6} \sum_{\alpha, \gamma} [3(n_\alpha + n_\gamma) - 2(n_\alpha^2 + n_\gamma^2 + n_\alpha n_\gamma)] (z_{\alpha\gamma}^2),$$

which is negative definite for n_α bounded by 0 and 1.

(5) In using the equation of motion for the density matrix, it may be useful to form an initial guess as to which $\rho(0, z)$ which does not commute with \underline{H} . This would be the case, for example, if one has the solution for one set of atomic positions and then allows the atoms to move slightly, as in a molecular-dynamics calculation. In that case, starting with the previous solution for the density matrix may be more efficient than starting over from scratch. However, we must now show that Eq. (1) provides a reasonable solution when the initial guess does not commute with \underline{H} . In this case, one can show that the equation of motion removes the part of $\underline{\rho}$ that does not commute. That is for the solution of $\partial \rho / \partial \beta = 0$, $[\underline{H}, \underline{\rho}] = 0$. Again working in the basis which diagonalizes $\underline{\rho}$, the fixed point gives $0 = x_{\alpha\gamma} z_{\alpha\gamma}$, with $x_{\alpha\gamma} = 3(n_\alpha + n_\gamma) - 2(n_\alpha^2 + n_\gamma^2 + n_\alpha n_\gamma)$. The commutator $C = [\underline{\rho}, \underline{H}]$ has elements $C_{\alpha\gamma} = (n_\alpha - n_\gamma) z_{\alpha\gamma}$. Now $x_{\alpha\gamma} = 0$ if either $n_\alpha = n_\gamma = 0$ or 1, in which case $C_{\alpha\gamma}$ is also zero. If $x_{\alpha\gamma} \neq 0$, then $z_{\alpha\gamma} = 0$ and $C_{\alpha\gamma}$ is zero anyway. This requirement dictates the symmetrized form of Eq. (1); some other choices of the ordering do not force the commutator to vanish at the fixed point.

The equation of motion can be restated as a steepest-descent solution to a modified variational theorem. Consider minimization of $G = \text{Tr}[\underline{H}(3\underline{\rho}\underline{\rho} - 2\underline{\rho}\underline{\rho}\underline{\rho})]$ with respect to variations in $\underline{\rho}$, and one recovers Eq. (1). In this way, we have turned the variation of a functional linear in $\underline{\rho}$ with nonholonomic constraints into the variation of a functional which is nonlinear in $\underline{\rho}$ without constraints.¹⁵ In fact, many of the properties proven above are easier to prove using this G function. In general, it may be easier to deal with minimization of G than with the equation of motion [Eq. (1)].

At this point one can make comparisons to the previous work of McWeeney.^{16,17} McWeeney demonstrated a method for varying $\underline{\rho}$ within the restricted space of idempotency ($\underline{\rho} = \underline{\rho}\underline{\rho}$). This restricts the eigenvalues of $\underline{\rho}$ to be 0 or 1 during the entire variational process. Using steepest descents within this subspace gives $\Delta \underline{\rho} = \Delta \beta (\frac{1}{2})(\underline{z}\underline{\rho} + \underline{\rho}\underline{z} - 2\underline{\rho}\underline{z}\underline{\rho})$. All of the properties listed above apply to this equation of motion as well. McWeeney applied this to calculations of molecular systems. The difference in the present work is in allowing $\underline{\rho}$ to be nonidempotent, and also in developing the approximation

for extended systems that will be discussed in the following.

In practice, one cannot solve for the *exact* density matrix of an extended system. This is because ρ_{ij} has infinite range and an infinite number of elements. However, there are reasons to believe that to a good approximation one can consider ρ_{ij} to be cut off beyond some range (i.e., banded in a localized basis). For a free electron metal in N_d dimensions, the density matrix between sites separated by a distance r falls off like r^{-p} , where $p = (N_d + 1)/2$. For insulators, the density matrix decays exponentially.

The idea of limiting the range of the density matrix is supported by the successes of the EAM (Finnis-Sinclair), many-body potentials, the cluster functionals, and bond-order potentials. These schemes represent approximations where the density matrix is manifestly banded. For example, the band energy in the Finnis-Sinclair⁵ [EAM (Ref. 4)] is assumed to scale with the second moment on each site: $E = -\alpha(\sum_m \sqrt{\mu_{2,m}})$, where $\mu_{2,m} = \sum_{n \neq m} (H_{nm})^2$ and α is an empirical constant. This can be rewritten in terms of a density matrix as $E = \text{Tr}[\underline{H}\underline{\rho}]$, with the off-diagonal elements of $\underline{\rho}$ taken to be

$$\rho_{mn} = \left[-\frac{\alpha}{2} \right] \left[\frac{1}{\sqrt{\mu_{2,m}}} + \frac{1}{\sqrt{\mu_{2,n}}} \right] H_{mn}, \quad (3)$$

which has the same range as the Hamiltonian. The EAM accounts locally for coordination effects and has been very successful in application to noble metals. But it is well known³ that the EAM neglects directional-bonding contributions that are important for metals from the center of the transition series and for semiconductors. An improvement that helps incorporate some of the directional bonding is to include higher powers of the Hamiltonian, which in effect includes many-body clusters in the energy. The range of the density matrix is then some small multiple of the range of the Hamiltonian. Moriarty's many-body potential,¹⁰ Pettifor's bond-order potentials,¹¹ and Carlsson's fourth-moment cluster functional¹² can be expressed in just such a way, and do account for directional-bonding and band-filling effects. In all of these schemes, the approximation to the density matrix is justified by physical argument and comparison to reference systems. With this motivation in mind, it seems reasonable to approximate the density matrix by a banded matrix.

We can rank the quality of solutions then by the range of $\underline{\rho}$. This can be compared with the cluster functionals, where the range is determined by the powers of \underline{H} . The lowest (nontrivial) level is to set the range of $\underline{\rho}$ equal to the range of \underline{H} . This would be compared with the EAM (second-moment treatment). However, even for a cutoff at nearest neighbors, the present scheme incorporates more of the essential physics than the EAM. (This will be demonstrated in examples near the end of the paper.)

We therefore propose to solve the equation of motion [Eq. (1)] for a "pruned" density matrix. At each iteration, the change in $\underline{\rho}$ is computed only for elements inside the cutoff; elements outside the cutoff are always held at zero. For the pruned case, the properties [(1)–(5)] listed previously must be modified. For example, the final solu-

tion for the pruned density matrix may not commute with the Hamiltonian. Nonetheless, the algorithm gives in practice a very reasonable approximation to the correct density matrix, and in the limit of large range we recover the exact solution.

The practical implementation of this algorithm involves the streamlining of the banded matrix multiplications in Eq. (1). The number of nonzero elements of $\underline{\rho}$ scales linearly with the number of sites N_a in the supercell. In ordinary matrix multiplication, many of the multiplications involve elements with value 0. Instead, it is more efficient for large systems to construct multiplication maps whereby only nonzero multiplications are considered. Doing this reduces the computer time to $O(N_a)$.

Having implemented this scheme we can now display the behavior of solutions for three well-studied examples: one-dimensional (1D) infinite chain, the semi-infinite chain, and a vacancy in a simple cubic lattice. In all cases we consider a nearest-neighbor distance of 1, and a Hamiltonian with a nearest-neighbor interaction of -1 and no other interactions. We have picked fairly simple examples for ease of analysis; in the future we will publish results of work on more complicated systems.

In the first example, we consider an infinite, linear chain. For this case, $\rho_{mn} = \sin[\pi f(m-n)]/[\pi(m-n)]$, where f is the band filling (number of electrons per site). The envelope of this function decays like $1/\text{distance}$, which is not very fast. But because the total energy is $\sum_{m,n} \rho_{mn} H_{nm}$, what we really need to do is calculate $\underline{\rho}$ at the nearest neighbor only, and it will be demonstrated here that ρ_{01} converges well with the cutoff distance. In Fig. 1, we show the density matrix as calculated from Eq. (1) using different cutoff distances, and compare it to the exact solution. The convergence in the density matrix is quickest for half-filled bands. In Fig. 2, we plot the energy for different fillings and different cutoffs. Using the cutoff at the nearest neighbors gains better than 90% of the total energy in the half-filled case. It is clear that the convergence is faster for half-filled bands.

The second example involves a semi-infinite chain. In the full solution, the Fermi distribution function introduces a long-range effect of the termination of the chain: the density matrix develops long-range oscillations. To illustrate the oscillations, we look at the density matrix between a site (n) and its neighboring site ($n+1$), as a function of n (the end of the chain is at $n=1$). Figure 3 shows the exact and approximate solutions. At $f = \frac{1}{2}$, the exact solution oscillates with a wavelength of two lattice spacings and has a decay envelope of $1/n$. Solving Eq. (1) under these conditions shows a similar oscillation even when we prune the density matrix to nearest neighbors, although the oscillation is damped out by $n=6$.

In the third example we consider something more complicated: the change in the band energy for creating a vacancy in a simple cubic supercell of 1000 atoms. The traditional band-structure calculation gives 0.42. The current method gives 0.21 for a density matrix cutoff after nearest neighbors and 0.40 for a cutoff after third-nearest neighbors.

Compared to many-atom potentials, the current scheme has the advantage that it produces a variational

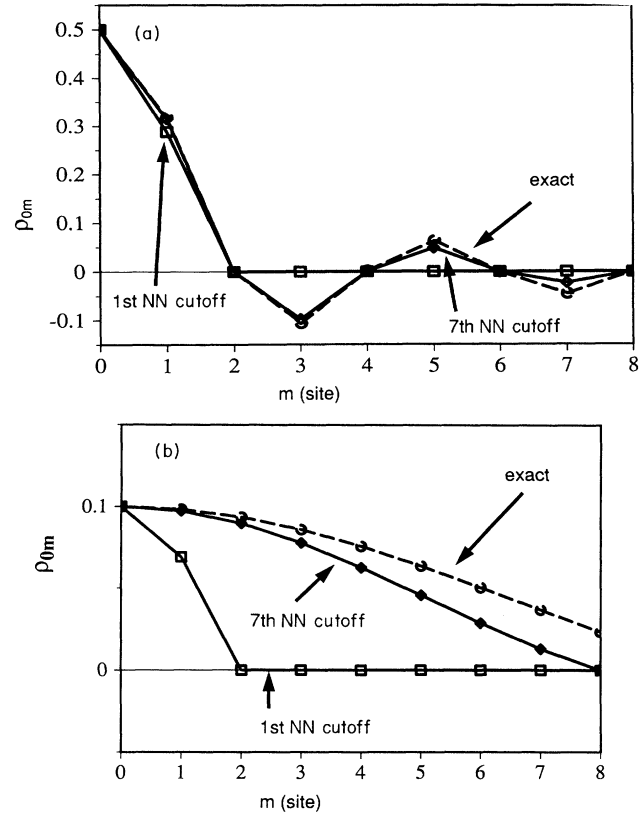


FIG. 1. Elements of the density matrix for an infinite 1D chain of sites with s states and nearest-neighbor interactions. The plot shows ρ_{0m} vs m . Comparison is made between approximate solutions (1-NN cutoff is labeled by squares, 7-NN cutoff is labeled by diamonds) with the exact solution (labeled by circles). (a) $f = \frac{1}{2}$ (half-filled). (b) $f = \frac{1}{10}$.

density matrix that obeys the eigenvalue constraints. It can be demonstrated easily that the density matrix corresponding to the EAM [Eq. (3)] violates the eigenvalue constraints (as is also true for the many-atom potentials). The current method produces the best physically reason-

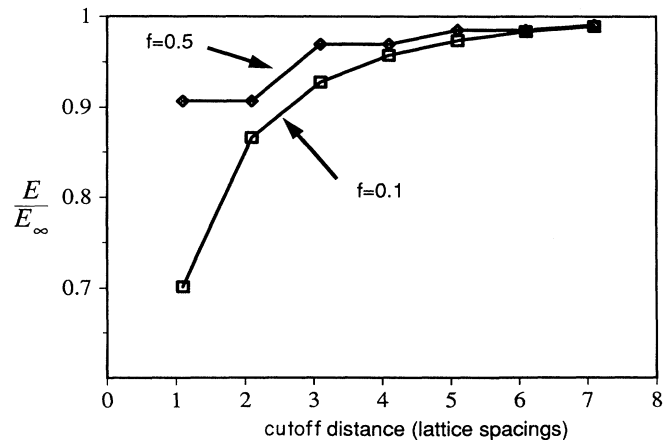


FIG. 2. Energy evaluated for various cutoffs in the density matrix. The energy is normalized by the exact result. Results for $f = \frac{1}{2}$ (diamonds) and $f = \frac{1}{10}$ (squares) are shown.

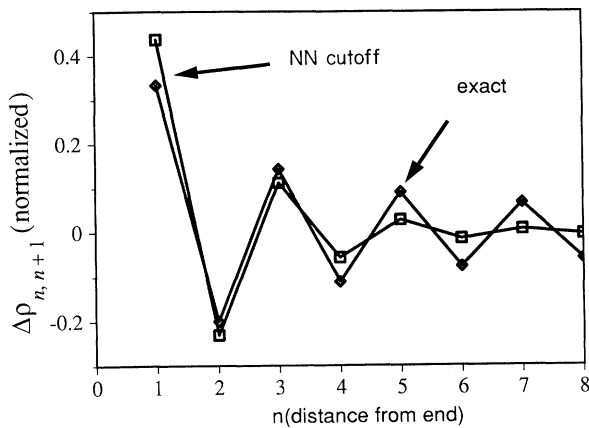


FIG. 3. Oscillations in the density matrix for a semi-infinite, 1D chain of sites. The quantity plotted is the nearest-neighbor element $\rho_{n,n+1}$ vs the site index n . The end of the chain is at $n = 1$. The band is half-filled ($f = \frac{1}{2}$). The approximate solution is for a nearest-neighbor cutoff in the density matrix. In both solutions, the asymptotic (bulk) value has been subtracted off and also used as normalization.

able density matrix within the constraint of the cutoff distance.

The recursion method approximates the DOS, which involves an additional integral to produce a total energy; the density matrix, on the other hand, is already an integrated quantity. During a molecular-dynamics run (or energy minimization), the recursion coefficients must be recalculated at each time step; knowledge of the recursion coefficients from previous steps is of no particular help. However, the density matrix from a previous step can be used as the starting point for finding the current density matrix and will decrease the number of iterations.

In terms of computational costs, we find that the densi-

ty matrix with a range twice that of the Hamiltonian is comparable in speed to second-level recursion and to four-atom potentials.

Finally although we have framed the discussion of the current method in terms of tight-binding theory, the same arguments follow for using a self-consistent Hamiltonian expressed in a localized basis. For example, a calculation of the properties of a supercell of Si atoms using local-density-approximation (LDA) Hamiltonian with Gaussian basis would be entirely feasible. Another application would be the study of the Hubbard Hamiltonian (or any model that can be cast in tight-binding form), where a density matrix for each spin would be coupled by a local electrostatic interaction.

In conclusion, we have proposed a means of calculating energetics from a one-electron Hamiltonian, based on the density matrix. Previous models of the energetics can be expressed as *Ansätze* for the density matrix. The current scheme calculates the density matrix in real space with the approximation that the elements are cutoff at some range, and scales linearly with the size of the system. Three examples illustrate the utility of this approach. It is straightforward to extend the calculations to more complicated Hamiltonians (such as Gaussian-basis LDA or tight-binding calculations for *s-p* semiconductors or *d*-band metals) and to nonorthogonal basis functions.

The author would like to thank Professor David Vanderbilt for discussing his closely related work prior to publication, and also Professor Anders Carlsson, Dr. Stephen Foiles, Dr. Daryl Chrzan, and Professor Talat Rahman for useful conversations and comments on this manuscript. This work has been supported by the Office of Basic Energy Sciences, U.S. Department of Energy.

¹See, e.g., A. P. Sutton, M. W. Finnis, D. G. Pettifor, and Y. Ohta, *J. Phys. C* **21**, 35 (1988).

²See the articles by V. Heine, R. Haydock, and M. J. Kelly in *Solid State Physics: Advances in Research and Applications*, edited by H. Ehrenreich and D. Turnbull (Academic, New York, 1980), Vol. 35.

³A. E. Carlsson, in *Solid State Physics: Advances in Research and Applications* (Ref. 2), Vol 43.

⁴M. S. Daw and M. I. Baskes, *Phys. Rev. Lett.* **50**, 1285 (1983); *Phys. Rev. B* **29**, 6443 (1984).

⁵M. W. Finnis and J. E. Sinclair, *Philos. Mag. A* **50**, 45 (1984).

⁶K. W. Jacobsen, J. K. Nørskov, and M. J. Puska, *Phys. Rev. B* **35**, 7423 (1987).

⁷J. Tersoff, *Phys. Rev. Lett.* **56**, 632 (1986).

⁸J. R. Chelikowsky, J. C. Phillips, M. Kamal, and M. Strauss, *Phys. Rev. Lett.* **62**, 292 (1989).

⁹A. E. Carlsson, P. A. Fedders, and C. W. Myles, *Phys. Rev. B* **41**, 1247 (1990).

¹⁰J. A. Moriarty, *Phys. Rev. B* **38**, 3199 (1988); J. A. Moriarty and R. Phillips, *Phys. Rev. Lett.* **66**, 3036 (1991).

¹¹D. G. Pettifor, *Phys. Rev. Lett.* **63**, 2480 (1989).

¹²A. E. Carlsson, *Phys. Rev. B* **44**, 6590 (1991).

¹³The total binding energy of a collection of atoms may be written in the form $U = E_b + \frac{1}{2} \sum_{mn} \Phi(R_{mn})$, where E_b is the band energy and Φ is a pair interaction. [See A. P. Sutton, M. W. Finnis, D. G. Pettifor, and Y. Ohta, *J. Phys. C* **21**, 35 (1988)]. In this paper we will only be concerned with calculating the band energy.

¹⁴A. E. Carlsson and N. W. Ashcroft, *Phys. Rev. B* **27**, 2101 (1983).

¹⁵Strictly speaking, the minimum we seek is only a local minimum; G can become unbounded if the eigenvalues of ρ are allowed to exceed the region $-\frac{1}{2}$ to $\frac{3}{2}$. However, the variational problem is still easier in this form.

¹⁶R. McWeeny, *Rev. Mod. Phys.* **32**, 335 (1960).

¹⁷McWeeny also showed how a nearly idempotent matrix could be made more idempotent ("purified"). A matrix ρ can be purified by forming $\rho' = (3\rho\rho - 2\rho\rho\rho)$, if the eigenvalues of ρ fall between $-\frac{1}{2}$ and $\frac{3}{2}$. The purification form is just the argument of the nonlinear functional G in the previous paragraph.