Linear-scaling tight binding from a truncated-moment approach

A. F. Voter, J. D. Kress, and R. N. Silver
Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
(Received 28 August 1995; revised manuscript received 23 January 1996)

We present an approximation to the total-energy tight-binding (TB) method based on use of the kernel polynomial method within a truncated subspace. Chebyshev polynomial moments of the Hamiltonian matrix are generated in a stable and efficient manner through recursive matrix-vector multiples. To compute the energy, either the electronic density of states (DOS) or the zero-temperature Fermi function is smeared by convolution with the kernel polynomial, with Jackson damping to minimize Gibbs oscillations while maintaining the positivity of the DOS. These are shown to give approximate lower and upper bounds, respectively, on the exact TB energy, and are averaged to obtain an improved energy estimate. The scaling of the computational work is made linear in the number of atoms by truncating the moment computation at a certain range about each atom. Energy derivatives necessary for molecular dynamics are obtained via a matrix-polynomial derivative relation. The method converges to exact TB as the number of moments and the truncation range are increased. We demonstrate the convergence properties and viability of the method for materials simulations in an examination of defects in silicon. We also discuss the relative importance of truncation range versus number of moments. [S0163-1829(96)07719-3]

I. INTRODUCTION

The total-energy tight-binding (TB) method is increasingly viewed as a possible route to obtaining accurate interactions for atomistic simulation of materials. While TB is much less computationally demanding than first-principles calculations, it nonetheless suffers from a computational bottleneck in the Hamiltonian diagonalization step, which scales as the number of atoms (or number of basis functions, N) to the third power. Recently, considerable attention has been given to developing approximations that reduce this scaling from N^3 to N.[1–9] For example, this can be achieved by iterative approximation to the eigenvectors[1,4,5] or density matrix,[2,3] or via statistically based approximations.[6] These approaches offer substantial speed improvements as the number of atoms increases beyond 10^3. Typical break-even sizes, for which the N-scaling calculation becomes less expensive than the direct N^3 calculation, range from 50 to a few hundred atoms. Because any approach of this type requires a compromise, a given N-scaling method can have advantages over others, depending on the nature of the application. In the present work, we develop a method that may be suitable as a short-ranged, many-body interatomic potential. It is explicitly local in character, with a deterministic energy and exact energy derivatives. The N-scaling nature is achieved by a local truncation of the environment around each atom. If the truncation range is increased to infinity, the method is equivalent to the N^2-scaling form presented previously.[10,11] Although the present work deals exclusively with orthogonal TB, it may be possible to extend this approach to treat the more general case of nonorthogonal TB.[12]

This approach is closely related to the one presented by Goedecker and co-workers;[9,13,14] the main differences are in the use of a logical (rather than physical) local truncation procedure, the use of a Gibbs-damped kernel polynomial method (KPM), and the fact that exact energy derivatives are obtained in the present work, which is crucial if charge neutrality is imposed on each atom. The Gibbs-damped KPM, which has improved convergence with increasing number of moments, yields a unique Fermi energy (due to the Jackson form which gives a non-negative electronic density of states), and allows one to generate expressions for approximate upper and lower bounds to the TB energy which can be averaged to obtain an improved energy estimate.

There is also a loose correspondence with the spatially projected local-density-approximation method of Yang,[15] although the fundamentally different nature of the one-electron TB Hamiltonian prevents a direct comparison. In the present approach the Hamiltonian, and consequently the basis set, are truncated to a subspace surrounding one atom. The TB solution in this subspace is then projected onto one atom to eliminate surface effects. In the Yang approach, the basis truncation and projection are one and the same, while the Hamiltonian is not altered.

The paper is organized as follows. After a brief review of the total-energy TB method, we discuss the local-subspace truncation procedure. This is first implemented as an exact diagonalization within each subspace for the purpose of directly testing the local truncation approximation. We then introduce the second approximation, the Gibbs-damped kernel polynomial method (KPM). This replaces the direct diagonalization, thus allowing the analytic computation of energy derivatives. The locally truncated KPM approach is then tested on defects in silicon using the TB parameters of Goodwin, Skinner, and Pettifor.[16]

II. DESCRIPTION OF THE METHOD

A. Total-energy tight-binding method

We first briefly review the orthogonal tight-binding method[17] as it is currently used in atomistic simulations. Given a system of N_{atom} atoms at a specified geometry, one constructs the one-electron Hamiltonian matrix (**H**) over the N-dimensional basis consisting of a valence basis set on each

atom. The off-site matrix elements of **H** generally have a simple, two-center form, with a parametrized radial shape for each possible angular momentum combination [e.g., $h_{ss\sigma}(r)$, $h_{sp\sigma}(r)$, $h_{pp\sigma}(r)$, and $h_{pp\pi}(r)$ for an sp^3 basis]. Angular dependencies are determined using Slater-Koster relations.[18] Constructing **H** is usually a fast step.

Diagonalizing **H** gives a set of eigenvalues $\{\epsilon_i\}$, which define the electronic density of states (DOS)

$$n(\epsilon)=\sum_i^N \delta(\epsilon-\epsilon_i). \tag{1}$$

The electronic energy ($E_{\rm elec}$) is defined as the energy integral over the occupied states of this DOS,

$$E_{\rm elec}=2\int_{-\infty}^{\infty}\epsilon\theta(\epsilon-E_F)n(\epsilon)d\epsilon, \tag{2}$$

where $\theta(\epsilon)$ is a zero-temperature Fermi-Dirac distribution

$$\theta(\epsilon)=\begin{cases}1 & \text{if } \epsilon<0\\ 0 & \text{if } \epsilon>0,\end{cases} \tag{3}$$

and the factor of two accounts for the closed-shell spin state (two electrons per orbital). The Fermi level (E_F) is either prespecified or is determined by requiring that the system have the correct number of valence electrons,

$$N_{\rm val}=2\int_{-\infty}^{\infty}\theta(\epsilon-E_F)n(\epsilon)d\epsilon. \tag{4}$$

The total energy of the system is then obtained by augmenting $E_{\rm elec}$ with a repulsive potential

$$E_{\rm tot}=E_{\rm elec}+E_{\rm rep} \tag{5}$$

representing core-core interactions and neglected contributions to the true electronic energy, such as double-counting terms. This usually takes the form of a simple pair potential

$$E_{\rm rep}=\tfrac{1}{2}\sum_{i\neq j}\phi(r_{ij}), \tag{6}$$

where $\phi(r_{ij})$ is a function of the internuclear distance $r_{ij}=|\mathbf{r}_{ij}|$. More recently, some parametrizations have employed a many-body term for $E_{\rm rep}$, such as an embedded-atom form,[19] to improve the overall accuracy of the method (with no significant increase in computational cost).

B. Local truncation

There are two main parts to the method: the use of a locally truncated subspace for each atom, and an approximation of the DOS using the kernel polynomial method. Local subspace truncation, which is responsible for the N-scaling computational dependence, is described and tested in this section. For simplicity, we discuss a system with only one basis function per atom, so that we can refer to basis functions and atoms interchangeably. The generalization to more complicated basis sets is straightforward. We also momentarily ignore contributions to the energy from $E_{\rm rep}$.

For each atom i, a local subspace of the full Hamiltonian is defined, containing N_i basis functions. This can be accomplished by retaining only those atoms within a certain physical distance from atom i, the procedure used by Goedecker and Teter,[13] which we will refer to as ''physical'' truncation. Alternatively, the truncation can be based on the atom connectivity defined in the Hamiltonian. Defining two atoms as ''H linked'' if they have a nonzero matrix element between them, one can retain all atoms within a given number (L) of Hamiltonian links from atom i. We refer to this as ''logical'' truncation. For example, $L=1$ truncation retains all atoms within $r_{\rm cut}$ of atom i, where $r_{\rm cut}$ is the cutoff distance for $h(r)$, while $L=2$ truncation retains all H-linked neighbors of i and all the H-linked neighbors of those neighbors.

Within this subspace (e.g., with L links), we solve the TB electronic structure problem, projecting out the DOS belonging to atom i [$n_i(\epsilon)$]. From the eigenvalues ($\{\epsilon_j\}$; $j=1,N_i$) and eigenvectors ($\{C_{jk}\}$; $j,k=1,N_i$) obtained from diagonalization of the subspace Hamiltonian, the projected DOS is given by

$$n_i^{(L)}(\epsilon)=\sum_j^{N_i} C_{ij}^2\delta(\epsilon-\epsilon_j). \tag{7}$$

An L-link approximation to the total DOS for the system is then obtained by summing over all i,

$$n^{(L)}(\epsilon)=\sum_i^N n_i^{(L)}(\epsilon). \tag{8}$$

An important point is that the atom-projected DOS is a well-defined quantity; the sum in (8) recovers the exact DOS for the system if each $n_i(\epsilon)$ is projected out of the exact electronic solution for the whole system, i.e., $n^{(L)}(\epsilon)\to n(\epsilon)$ as $L\to\infty$ or $N_i\to N$.

Note that once the Fermi level has been found from Eq. (4), one can compute the energy of each atom individually, by replacing $n(\epsilon)$ with $n_i(\epsilon)$ in Eq. (2). Also, if desired, one can enforce charge neutrality [via Eq. (4)] atom by atom.[21] A feature of this ''atom-neutral'' approach is that it offers a completely local definition of the energy of each atom, which should be advantageous if the goal is to adapt this method for a fast, short-ranged empirical potential.

Because the computational work required to compute each atom DOS depends only on the size of the subspace for that atom, the work required for the whole system scales as N, whether or not charge neutrality is enforced atom by atom. More specifically, if each of the diagonalized subspaces are equal in size, the work will scale as $(N_i)^3N$. The method as outlined so far meets most of our requirements; it scales as N and converges to exact TB as the truncation range is increased. However, it is not a viable approach for many applications, because calculation of energy derivatives would require differentiation of the eigenvector matrix elements, which is difficult to perform analytically. (In contrast, derivatives of the eigenvalues can be obtained easily, by exploiting the Hellman-Feynman theorem.) As will be discussed in Sec. II C, this derivative problem can be sidestepped by approximating $n_i(\epsilon)$ with a Chebyshev polynomial moment expansion. However, it is useful first to examine the direct-diagonalization approach, because it gives the exact $n_i(\epsilon)$ for a given subspace. We can compare the physical versus logical truncation procedures, and directly study the convergence to exact TB as N_i is increased,

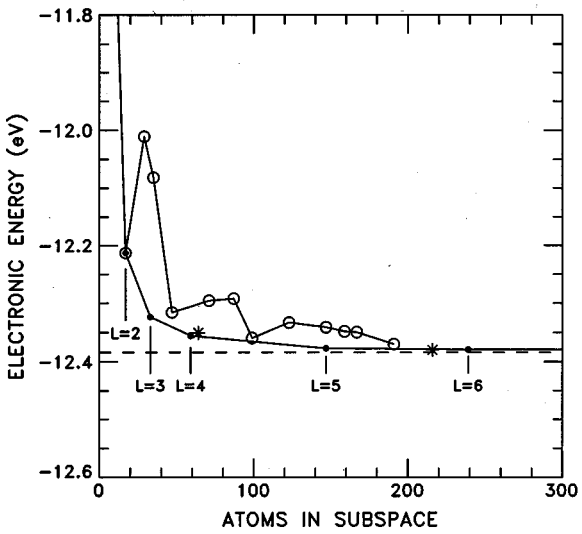

FIG. 1. E_{elec} from subspace diagonalization for a bulk Si atom using physical truncation (○), or logical truncation (●) to define the subspace. The $L=1$ point, which coincides with the first physical truncation point, is off the scale at -11.205 eV. Comparison is made to 64 ($2\times2\times2$ unit cells) and 216-atom ($3\times3\times3$) TB supercell calculations (*). The 1000-atom ($5\times5\times5$) TB supercell result is shown as a dotted line.

without any coupling to the approximations in the moment description. Also, the diagonalization corresponds to the infinite-moment limit of the Chebyshev moment approach, simplifying the study of the convergence with respect to the number of moments.

The calculations presented here employ the silicon tight-binding parametrization of Goodwin, Skinner, and Pettifor,[16] which has an sp^3 basis set. In all calculations, the diamond crystal lattice constant is set to 5.4291 Å, the result obtained from a 1000-atom supercell calculation. An abrupt cutoff is imposed at 3.6 Å (between first- and second-neighbor shells) for compatibility with previous studies.[20,21] The supercell calculations employ only the Γ point in the Brillouin zone.

Figure 1 shows the convergence of the electronic energy with subspace size, using both the logical truncation (up to $L=6$) and the physical truncation procedures. For this perfect diamond lattice, the physical truncation procedure picks up complete neighbor shells as the truncation radius is increased. Because of the first-nearest-neighbor cutoff in $h(r)$, the first two points ($L=1$ and 2) in the logical truncation curve correspond to exactly the same subspaces as for the physical truncation procedure, so the energies coincide. For larger subspaces, the logical truncation continues to converge monotonically toward the 1000-atom periodic supercell value (an approximation to the infinite-system result), passing very close to the energies from 64- and 216-atom supercell calculations. In contrast, the physical truncation shows less satisfactory, nonmonotonic behavior. We infer that logical truncation is preferable to physical truncation, although this assumption should be more thoroughly tested.

In tests of their local-orbital, N-scaling electronic structure method, Mauri and Galli[7] also observed that logical truncation was superior to physical truncation, judged by the residual error for a given subspace size. They did not see the nonmonotonic behavior observed here because of the variational nature of their solution.

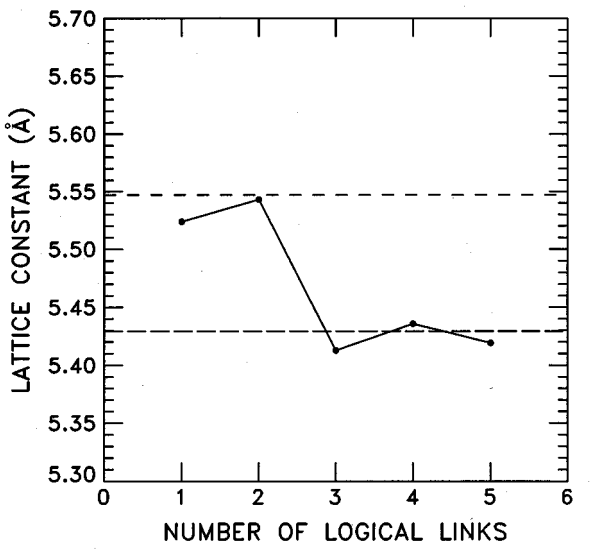

FIG. 2. Si lattice constant predicted using subspace diagonalization vs L. The dotted (upper) and dashed lines (lower) are the 216- and 1000-atom supercell results, respectively.

In Fig. 2, the Si lattice constant is plotted versus the number of logical truncation shells, L. The convergence is very rapid, with an error of less than 0.5% by $L=3$. Even for $L=1$, the error (relative to the 1000-atom supercell) is no larger than the error due to using a 216-atom supercell.

Figure 3 shows the convergence of the unrelaxed Si vacancy formation energy (E_{vac}) using the atom-neutral approach. These values decrease slightly if atom neutrality is not enforced. For comparison, the exact TB supercell results are shown for 64, 216, and 1000 atoms (enforcing charge neutrality has virtually no effect on the exact TB results). The subspace diagonalization values are in reasonable agreement with the exact TB values for $L=3$ and above. The predicted E_{vac} for a given L is similar to the value from the supercell whose size matches the size of the subspace.

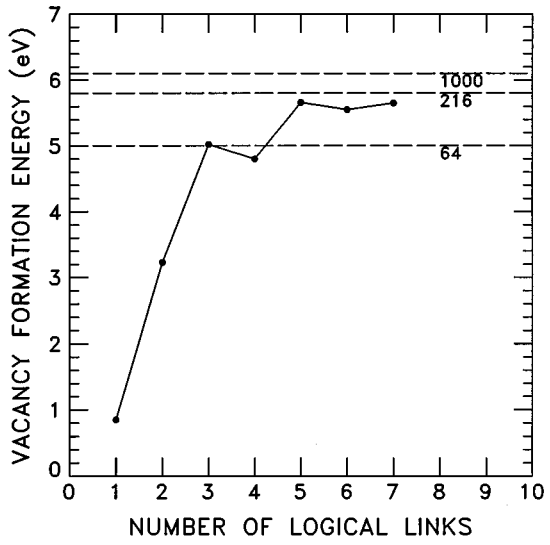

FIG. 3. Si unrelaxed vacancy formation energy using subspace diagonalization vs L. The dashed lines are the exact TB results for 64-, 216-, and 1000-atom supercells.

In a previous study using a low-order moment approximation to TB,[21] $E_{\rm vac}$ was found to be poorly described, converging slowly as the number of moments was increased. Both that method and the present method represent explicitly local approximations to TB. If one equates the number of moments in the low-order method to $2L$ in the present method, the local subspaces are equivalent. In both cases the convergence is from below, but the errors in the moment method are much larger. For example, $E_{\rm vac}$ is 2.6 eV lower with six moments than the $L=3$ subspace diagonalization result, and the 14-moment $E_{\rm vac}$ is 1.3 eV lower than the $L=7$ subspace diagonalization result.

C. Kernel polynomial method

The kernel polynomial method (KPM), implemented with Chebyshev polynomials, offers a controlled approximation to the bond integral in (2), while retaining facile differentiability. This section provides an introduction to the KPM suitable for the task at hand; the method has been presented in more formal detail elsewhere.[22,10,11]

Chebyshev polynomials of the first kind, $T_m(x)$, are defined by $T_0(x)=1$, $T_1(x)=x$, and the recurrence relation

$$T_m(x)=2xT_{m-1}(x)-T_{m-2}(x), \tag{9}$$

where the range of x is limited to $-1\leqslant x\leqslant 1$. They obey the orthogonality relation

$$\int_{-1}^{1} w(x)T_m(x)T_n(x)dx=q_m\delta_{m,n}, \tag{10}$$

where the weight function is

$$w(x)=\frac{1}{\sqrt{1-x^2}}, \tag{11}$$

and $q_m=\pi$ when $m=0$, and $\pi/2$ otherwise.

We can express an arbitrary function on the interval $\{-1{:}1\}$ as an expansion

$$f(x)=\sum_{m=0}^{\infty}\frac{a_m}{q_m}w(x)T_m(x). \tag{12}$$

Multiplying Eq. (12) by $T_n(x)$, integrating, and comparing with Eq. (10) shows that the expansion coefficients are just the Chebyshev moments of the function,

$$a_m=\mu_m=\int_{-1}^{1}f(x)T_m(x)dx. \tag{13}$$

While (12) is exact, we now consider truncating the expansion at M moments to obtain a usable approximation. We first shift and scale the energy to obtain $\mathbf{H}$, whose eigenvalues are in the range $\{-1{:}1\}$

$$\overline{\mathbf{H}}=(\mathbf{H}-b)/a, \tag{14a}$$

$$x=(\epsilon-b)/a. \tag{14b}$$

[This requires a determination of the range of the eigenvalues of $\mathbf{H}$, which can be accomplished in $O(N)$ work using a Lanczos approach[23] to find the lowest and highest eigenvalues.] We can then express the projected DOS for the ith basis function as

$$n_i(\epsilon)\simeq\sum_{m=0}^{M}\frac{\mu_{mi}}{q_m}g_m w(x)T_m(x), \tag{15}$$

where $\{\mu_{mi}\}$ are the Chebyshev moments of $\overline{\mathbf{H}}$,

$$\mu_{mi}=\langle i|T_m(\overline{\mathbf{H}})|i\rangle. \tag{16}$$

The Gibbs factors $\{g_m\}$ in (15) are designed to reduce the Gibbs oscillations resulting from the finite-M truncation of the series. We choose the form derived by Jackson,[24,11] which maintains the desired positive-definite nature of the DOS, while optimizing the energy resolution. For a DOS with continuity properties, this choice minimizes the uniform norm, which is defined as the maximum absolute error between the exact and approximate DOS.

The heart of the KPM is the relationship between the approximate DOS in (15) and the exact DOS. This connection can be more clearly stated after transforming to ϕ space, where $x=\cos(\phi)$ and the mth Chebyshev polynomial is simply $\cos(m\phi)$. In ϕ space, the approximate DOS is related to the exact DOS by a convolution with the *kernel polynomial*, an approximation to a Dirac delta function (normalized on $[0,2\pi]$)

$$\delta_K(\phi)=\sum_{m=0}^{M}\frac{1}{2q_m}g_m\cos(m\phi), \tag{17}$$

whose width is proportional to $1/M$. Figure 4 shows $\delta_K(\phi)$ for $M=100$, with and without Gibbs damping. Thus (15) represents a smearing of the exact DOS with a known resolution function. (In energy space, the width of the smearing function is not constant, reaching a maximum at $\epsilon=0$.) The generality of the KPM allows, as an alternative, convolution

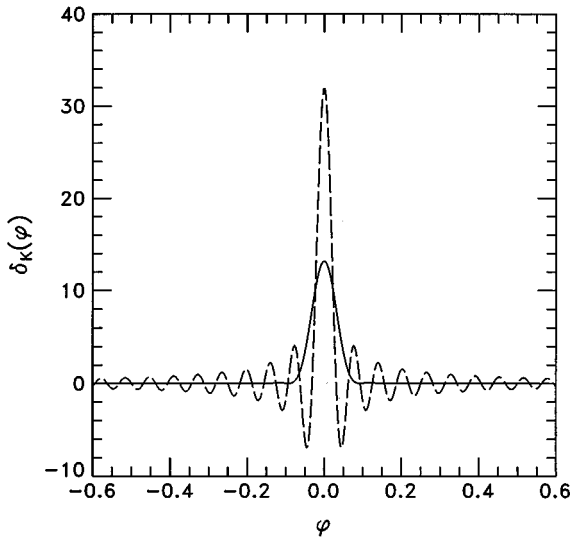

FIG. 4. The kernel polynomial for $M=100$ plotted vs the angular variable $\phi=\cos^{-1}(x)$. The raw polynomial (dotted line) exhibits Gibbs oscillations. Application of the Jackson damping factors substantially reduces these oscillations and enforces positivity (solid line).

of the zero-temperature Fermi function with (17), giving a smeared Fermi level instead of a smeared DOS. The resulting energy expressions, given below, can be averaged to obtain an improved energy estimate due to the approximate bounding properties presented in Sec. II E.

The Chebyshev moments in (16) are computed by defining the moment vector for a basis function i,

$$\psi_m = T_m(\overline{\mathbf{H}})|i\rangle, \tag{18}$$

such that

$$\langle i|T_m(\overline{\mathbf{H}})|i\rangle = \psi_0^\dagger \psi_m . \tag{19}$$

The time-consuming part of the calculation is then repeated matrix-vector multiples using the matrix version of the recurrence relation in (9),

$$\psi_m = 2\overline{\mathbf{H}}\psi_{m-1} - \psi_{m-2}. \tag{20}$$

Due to the sparse nature of **H**, each matrix-vector multiply requires computational work proportional to the subspace size N_i. Because this is then repeated for each basis function in the system, the overall work scales as MN_iN, or MN^2 if there is no local truncation.

Once the Chebyshev moments have been computed, calculation of $E_{\rm elec}$ is performed in ϕ space. The ϕ-space equivalent of Eq. (4) is

$$N_{\rm val} = 2\int_0^{2\pi} \overline{\theta}(\phi,\phi_F) n(\phi) d\phi, \tag{21}$$

where $n(\phi)$ is the ϕ-space DOS, ϕ_F is the Fermi angle [related to the Fermi energy by $E_F = a\cos(\phi_F) + b$], and $\overline{\theta}(\phi,\phi_F)$ is a 2π-periodic function that equals $\frac{1}{2}$ for $2\pi - \phi_F \geq \phi \geq \phi_F$ and is zero elsewhere. (The $\frac{1}{2}$ is a consequence of continuing ϕ onto the interval $[0,2\pi]$.) θ is related to θ in energy space [Eq. (3)]. Smearing the DOS by applying the kernel approximation to $n(\phi)$,

$$n_K(\phi) = \int_0^{2\pi} \delta_K(\phi-\phi') n(\phi') d\phi'. \tag{22}$$

and substituting $n_K(\phi)$ for $n(\phi)$ in Eq. (21), yields

$$N_{\rm val} = \mu_0 g_0 \left(1 - \frac{\phi_F}{\pi}\right) - \sum_{m=1}^{M} \frac{2 g_m \mu_m \sin(m\phi_F)}{m\pi}. \tag{23}$$

Because the choice for the Gibbs damping factors g_m guarantees that $n_K(\phi)$ is non-negative, Eq. (23) defines a unique solution for ϕ_F; this can be found, for example, by bisection. Taking a similar approach to the energy expression,

$$E_{\rm elec} = 2\int_0^{2\pi} \overline{\theta}(\phi,\phi_F)\cos(\phi) n(\phi) d\phi, \tag{24}$$

yields the smeared-DOS (SD) approximation to the energy,

$$E_{\rm elec}^{\rm SD} = -\frac{\mu_0 g_0 \sin(\phi_F)}{\pi} - \mu_1 g_1 \left(\frac{\phi_F}{\pi} - 1 + \frac{\sin(2\phi_F)}{2\pi}\right) - \sum_{m=3}^{M+1} \mu_{m-1} g_{m-1} \left(\frac{\sin(m\phi_F)}{m\pi} + \frac{\sin[(m-2)\phi_F]}{(m-2)\pi}\right). \tag{25}$$

For the smeared-Fermi (SF) approximation, we convolute $\overline{\theta}(\phi,\phi_F)$ with $\delta_K(\phi)$,

$$\overline{\theta_K}(\phi,\phi_F) = \int_0^{2\pi} \delta_K(\phi-\phi')\overline{\theta_K}(\phi',\phi_F) d\phi'. \tag{26}$$

Substitution of $\overline{\theta_K}$ for $\overline{\theta}$ in Eq. (21) yields the same expression for $N_{\rm val}$ as in (23). Thus ϕ_F is the same for both the SF and SD approximations. The SF energy expression is found by substituting $\theta_K(\phi,\phi_F)$ for $\theta(\phi,\phi_F)$ in Eq. (24),

$$E_{\rm elec}^{\rm SF} = \mu_1 g_0 \left(1 - \frac{\phi_F}{\pi}\right) - \sum_{m=1}^{M} (\mu_{m-1} + \mu_{m+1}) g_m \frac{\sin(m\phi_F)}{m\pi}. \tag{27}$$

As discussed below, the average of $E_{\rm elec}^{\rm SF}$ and $E_{\rm elec}^{\rm SD}$ offers a better approximation to the true energy than either one alone.

As with the exact-diagonalization case, charge neutrality can be imposed on each atom (atom-neutral method); $n_i(\epsilon)$ is found by summing the moments over the basis functions on atom i. Alternatively, the DOS for the whole system can be obtained by summing over all N basis functions. Note that due to the linearity of the KPM, determining $n_i(\epsilon)$ for each atom, and then summing to obtain $n(\epsilon)$, yields a total DOS that is equivalent to that obtained by summing the moments over all atoms before applying KPM once to determine $n(\epsilon)$. This is in contrast to nonlinear methods such as maximum entropy, where the total DOS contains more information if it is obtained by summing the maximum-entropy DOS for each atom (or rotationally invariant subspace) rather than determining it directly from the total system moments.[21]

D. Energy derivatives

Differentiating the electronic energy with respect to an atom coordinate (y_j) can be accomplished with the following decomposition:

$$\frac{\partial E}{\partial y_j} = \sum_{i=0}^{N}\sum_{m=0}^{M} \frac{\partial E}{\partial \mu_{mi}} \frac{\partial \mu_{mi}}{\partial y_j}. \tag{28}$$

The best route for evaluating $\partial E/\partial \mu_{mi}$ depends on whether there is a fixed Fermi level (the easiest case), a floating Fermi level [solving Eq. (4)], or an individual Fermi level for each atom (atom-neutral method). Analytical expressions can be derived for each case.

The more formidable task, differentiating the moments, merits discussion. If no local truncation or charge neutrality is applied, the total moment

$$\mu_m = \sum_i^N \langle i|T_m(\overline{\mathbf{H}})|i\rangle \tag{29}$$

can be differentiated in a straightforward fashion. The trace over individual moments allows cyclic permutation of the matrices in each polynomial term, so that the scalar equivalent can be employed:

$$\frac{\partial \mu_m}{\partial y_j} = \sum_i m\langle i| \frac{\partial \overline{\mathbf{H}}}{\partial y_j} U_{m-1}(\overline{\mathbf{H}})|i\rangle, \tag{30}$$

where $U_m(x)$ is the mth-order Chebyshev polynomial of the second kind.

In the case of local truncation, which breaks the trace in (29), this approach is only approximate (contrary to the assertion in Ref. 13). To obtain the exact derivative requires a full matrix treatment of the polynomial expansion, which can be accomplished in a fashion that retains the N scaling. We exploit the derivative recurrence relation

$$\frac{\partial T_m(\overline{\mathbf{H}})}{\partial y_j} = \frac{\partial T_{m-2}(\overline{\mathbf{H}})}{\partial y_j} + \sum_{\alpha=0}^{m-1} (1+k_\alpha)(1+k_{m-1-\alpha}) T_\alpha(\overline{\mathbf{H}}) \times \frac{\partial \overline{\mathbf{H}}}{\partial y_j} T_{m-1-\alpha}(\overline{\mathbf{H}}), \tag{31}$$

initialized with $\partial T_0(\overline{\mathbf{H}})/\partial y_j = 0$ and $\partial T_1(\overline{\mathbf{H}})/\partial y_j = \partial \overline{\mathbf{H}}/\partial y_j$ (here $k_\alpha = 0$ if $\alpha \leqslant 0$, and 1 otherwise). Equation (31), which is derived in the Appendix, allows the derivatives to be computed using the ψ_m vectors of Eq. (18). Although the derivatives can be approximated using Eq. (30) (an approximation that becomes very good for large L), Eq. (31) is necessary for the atom-neutral case, because the moment for each atom must be differentiated.

Using either Eqs. (30) or (31), the energy range spanned by the DOS, as specified by a and b in Eq. (12), must be kept fixed, to avoid terms of the form $(\partial E/\partial a)(\partial a/\partial \mu_m)$. For the Si calculations presented below, the end points of the DOS energy range were fixed at -16.0 and $+10.0$ eV (the self-energies ϵ_s and ϵ_p were set according to $\epsilon_s = -3\epsilon_p = -6.221\,25$ eV, zeroing the first moment of the DOS).

E. Approximate energy bounds

Applying the KPM to smear the DOS leads to an approximate lower bound on the true TB energy, while smearing the Fermi function gives an approximate upper lower bound. These bounding relationships are derived in this section. For simplicity, we first shift the energy scale by $-E_F^{(0)}$ to place the Fermi level at $E=0$. The electronic energy in the unshifted energy system ($E_{\text{elec}}^{(0)}$) can be recovered from the electronic energy in the shifted system (E_{elec}) using

$$E_{\text{elec}}^{(0)} = E_{\text{elec}} + E_F^{(0)} N_{\text{val}}, \tag{32}$$

where N_{val} is given by Eq. (4). The energy expression now becomes

$$E_{\text{elec}} = 2\int_{-\infty}^{\infty} \theta(\epsilon)\,\epsilon\, n(\epsilon)\, d\epsilon. \tag{33}$$

The smeared-DOS (SD) approximation to (33) is

$$E_{\text{elec}}^{\text{SD}} = 2\int_{-\infty}^{\infty} \theta(\epsilon)\,\epsilon\; n_s(\epsilon)\, d\epsilon. \tag{34}$$

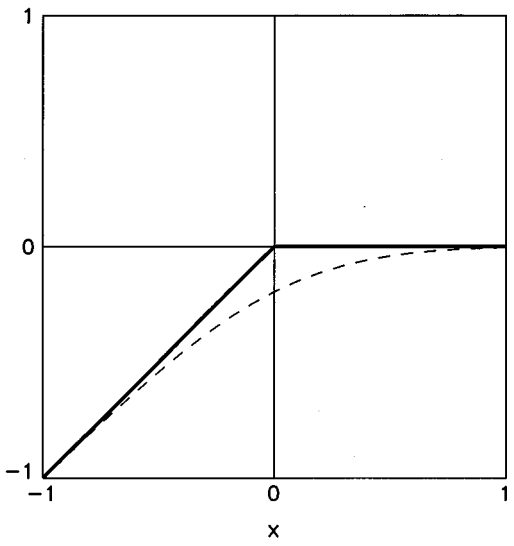

FIG. 5. Illustration that the smeared product function $[\epsilon\theta(\epsilon)]_s$ is a lower bound on the unsmeared form. A Gaussian with $\sigma = 0.5$ was used as the smearing function for this example.

Here an s subscript will indicate energy smearing via convolution with a symmetric, non-negative, normalized smearing function:

$$f_s(\epsilon) = \int_{-\infty}^{\infty} f(\epsilon') S(\epsilon' - \epsilon)\, d\epsilon,$$
$$S(-\epsilon) = S(\epsilon), \tag{35}$$
$$\int_{-\infty}^{\infty} S(\epsilon)\, d\epsilon = 1.$$

By swapping the convolution, Eq. (34) can be rewritten as an integral over the exact DOS times a smearing of $\epsilon\theta(\epsilon)$,

$$E_{\text{elec}}^{\text{SD}} = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \epsilon\theta(\epsilon) n(\epsilon') S(\epsilon' - \epsilon)\, d\epsilon'\, d\epsilon$$
$$= 2\int_{-\infty}^{\infty} [\epsilon\theta(\epsilon)]_s n(\epsilon)\, d\epsilon. \tag{36}$$

Convolution of the function $\epsilon\theta(\epsilon)$ with S yields a function that equals or lies below $\epsilon\theta(\epsilon)$ at every point, as indicated in Fig. 5. Recalling that $n(\epsilon)$ is non-negative, the bounding property is easily shown:

$$E_{\text{elec}}^{\text{SD}} = E_{\text{elec}} + 2\int_{-\infty}^{\infty} \{[\epsilon\theta(\epsilon)]_s - [\epsilon\theta(\epsilon)]\} n(\epsilon)\, d\epsilon$$
$$\leqslant E_{\text{elec}}. \tag{37}$$

Within the KPM, $E_{\text{elec}}^{\text{SD}}$ is only an approximate lower bound to E_{elec}, because the smearing is in fact a true convolution only in ϕ variables, not in ϵ (as discussed in Sec. II C), and the Fermi level may not be exact.

Turning now to the smeared Fermi-level (SF) approximation, we have

$$E_{\text{elec}}^{\text{SF}} = 2\int_{-\infty}^{\infty} \theta_s(\epsilon)\,\epsilon\; n(\epsilon)\, d\epsilon, \tag{38}$$

which we rewrite as

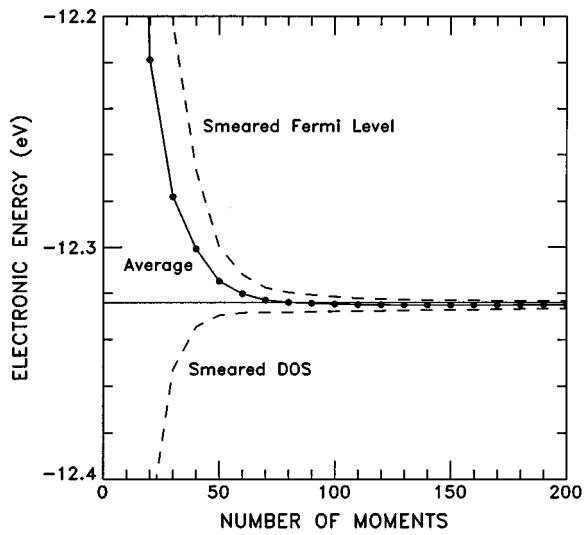

FIG. 6. Si $E_{\rm elec}$ using KPM with $L=3$ truncation vs the number of moments, M; smeared Fermi level (upper dashed line); smeared DOS (lower dashed line); and the average of the two (solid). The straight line is the exact $L=3$ result (subspace diagonalization).

$$E_{\rm elec}^{\rm SF}=E_{\rm elec}+2\int_{-\infty}^{0}[\theta_s(\epsilon)-\theta(\epsilon)]\epsilon\, n(\epsilon)d\epsilon$$
$$+2\int_{0}^{\infty}[\theta_s(\epsilon)-\theta(\epsilon)]\epsilon\, n(\epsilon)d\epsilon$$
$$=E_{\rm elec}+2\int_{-\infty}^{0}[\theta_s(\epsilon)-1]\epsilon\, n(\epsilon)d\epsilon$$
$$+2\int_{0}^{\infty}[\theta_s(\epsilon)]\epsilon\, n(\epsilon)d\epsilon. \quad (39)$$

The first integral is non-negative, because over its range $\theta_s(\epsilon)\leqslant 1$, $\epsilon\leqslant 0$, and $n(\epsilon)\geqslant 0$. The second integral is also non-negative, because over its range $\theta_s(\epsilon)\geqslant 0$, $\epsilon\geqslant 0$, and $n(\epsilon)\geqslant 0$. Thus $E_{\rm elec}^{\rm SF}$ is an upper bound on $E_{\rm elec}$. This proof does not require that $\theta_s(\epsilon)$ be a true convolution, only that $0\leqslant\theta_s(\epsilon)\leqslant 1$ everywhere. In the KPM, the Jackson form of Gibbs damping ensures this property, but the overall bound is approximate because the position of E_F is not necessarily exact.

Because the errors in $E_{\rm elec}^{\rm SD}$ and $E_{\rm elec}^{\rm SF}$ are caused by the same smearing function, the absolute errors are expected to be similar in magnitude. This raises the possibility that the average of these two approximate bounds gives a less biased estimate of the exact TB energy. Empirically, we have found this to be true.

III. SILICON TEST CALCULATIONS

Figure 6 shows the moment convergence of the electronic energy (per atom) for the $L=3$ truncated Si system. The upper dashed curve results from applying the KPM to smear the Fermi function, while the lower dashed curve results from using the KPM to smear the DOS. Using the average of the smeared DOS and smeared Fermi methods, convergence to the exact $L=3$ result is seen to be more rapid, deviating by 0.08% at $M=50$ and 0.005% (2.4×10^{-3} eV) at $M=100$. We use this average energy form hereafter.

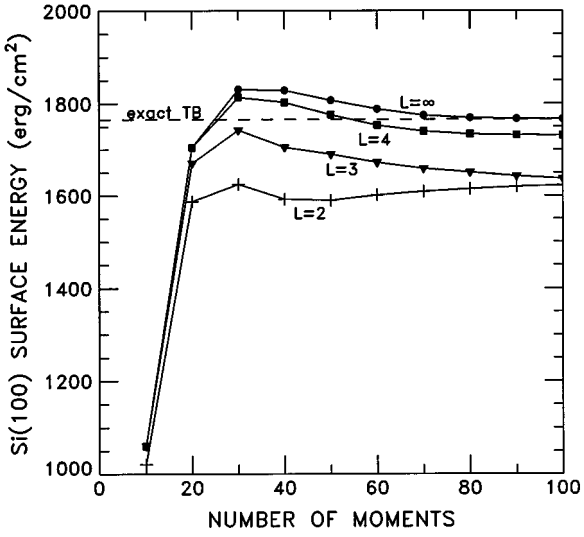

FIG. 7. Si(100) unrelaxed surface energy vs M for the 216-atom system, using KPM at various levels of logical truncation; $L=2(+)$, $L=3$ (triangle), $L=4$ (square), and $L=\infty$ (●). The straight line is the full TB diagonalization.

Figure 7 shows the moment convergence of the unrelaxed (100) surface energy for $L=2$, $L=3$, $L=4$, and $L=\infty$. These calculations employed a 216-atom supercell with a global Fermi level. While the $L=1$ case gives a very poor description (negative surface energy, not shown), $L=2$ through $L=4$ each appear well converged by $M=100$, and the sequence of their asymptotes converges toward the exact TB result.

Figure 8 shows $E_{\rm vac}$ for a 216-atom system. The upper curve is the unrelaxed result, using the atom-neutral KPM, showing rapid convergence to the exact $L=3$ result of 5.02 eV (dashed line). Also shown is $E_{\rm vac}$ after a full relaxation of the atomic coordinates at each moment level. It appears con-

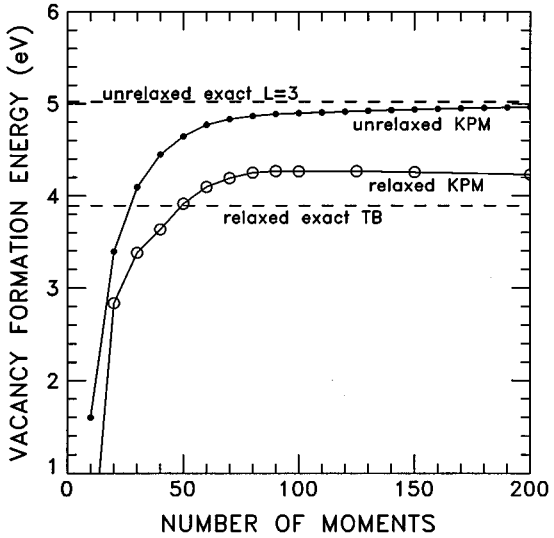

FIG. 8. Si vacancy formation energy for the 216-atom system using KPM with $L=3$ truncation; unrelaxed geometry (●) and relaxed geometry (○). The upper dotted line is the exact, unrelaxed $L=3$ result (from subspace diagonalization), while the lower dotted line is the relaxed, untruncated, exact TB result.

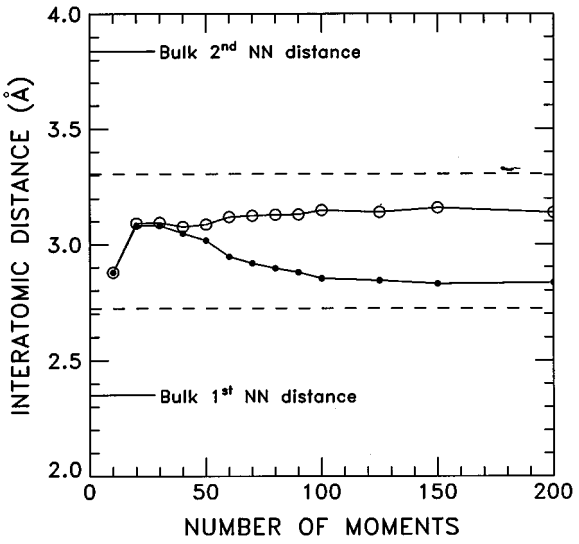

FIG. 9. Distances between atoms adjacent to relaxed vacancy, using atom-neutral, $L=3$ KPM. Jahn-Teller distortion breaks $R_1=R_2$ symmetry, leading to short ''bonds'' [R_1(●)] and long ''bonds'' [R_2(○)]. Dotted lines show R_1 and R_2 for untruncated, exact TB.

verged by $M=100$, offering a good approximation to the untruncated, relaxed, exact TB result of 3.89 eV. (Note that because energy derivatives are not easily determined for the truncated-subspace diagonalization, the infinite-moment limit for the relaxed $L=3$ case is not available.)

First-principles calculations[25] have shown the relaxed Si vacancy exhibits a D_{2d} Jahn-Teller distortion in which the four neighbors of the vacancy site, initially second-nearest neighbors to each other, pair up to form two shorter bonds. This was also observed in TB studies by Wang, Chan, and Ho,[20] as well as in the present study. This behavior is properly predicted by the L-truncated KPM approximation if enough moments are used. Figure 9 shows the relaxed distance between the two unique pairs of vacancy-neighbor atoms as a function of M using the atom-neutral method. Up to $M\sim30$, the two distances are equivalent ($R_1=R_2$), but reduced from the unrelaxed bulk value ($R_1=R_2=3.84$ Å), indicating a symmetric relaxation of the atoms toward the vacancy. At $M\sim50$, there is a bifurcation as the Jahn-Teller distortion occurs, with R_1, the distance between the atoms making the new bond, becoming shorter than R_2, the distance between the pairs of atoms not involved in a new bond. (Note that this new ''bond'' is still much longer than the bulk nearest-neighbor distance, 2.35 Å.) For $M>100$, there is qualitative agreement with the exact, untruncated TB geometry (dotted lines). This transition at $M\sim50$ is a consequence of the M-dependent energy resolution of the DOS. Below a critical value of M, the smearing obscures the information that there is a partially filled state at the Fermi level.

If the atom-neutral restriction is eliminated without changing the atomic geometry, the vacancy formation energy at $M=100$ drops from 4.26 to 3.69 eV. Allowing the atomic geometry to relax lowers this further, to 3.54 eV. Interestingly, this relaxed global Fermi-level system loses nearly all of the Jahn-Teller distortion, with $R_1=2.93$ and $R_2=2.94$. The atom-neutral description thus appears to offer a more accurate approximation to the true tight binding.

We have also studied the vacancy system for $L=2$ KPM. At $M=90$, the unrelaxed vacancy formation energy of 2.96 eV is similar to the subspace diagonalization result plotted in Fig. 3 (about 3.2 eV), but, upon relaxation, drops to 0.04 eV. This nearly 3-eV relaxation is strikingly different from the 0.7-eV relaxation effect for $L=3$. We conclude that artificial behavior can result when the overall truncation range is too short.

IV. CONCLUSIONS

The locally truncated N-scaling approximation to TB presented here should be useful for atomistic simulation of materials. The kernel polynomial method gives a well-controlled approximation to the exact TB energy. Applying Gibbs damping in the Jackson form eliminates unphysical behavior such as negative DOS values and multiple solutions for the Fermi energy. Approximate upper and lower bounds to the exact energy result from smearing either the DOS or the Fermi function, respectively, and averaging these approximate bounds gives an energy estimate that converges rapidly with the number of moments. The exact energy derivatives derived here are crucial for short truncation lengths or when the atom-neutral method is used. A logical truncation range of $L=3$–5, with $M\simeq100$ moments, appears adequate for investigating defect energetics and geometries.

In comparing the present method to an untruncated moment approach, we note that there is a significant disparity between the necessary number of moments ($M\sim50$–100) and the necessary subspace size ($L\sim3$–5). The use of 50 moments without truncation would define a very large subspace corresponding to $L=25$, which is about six lattice constants in radius for the silicon system with a nearest-neighbor cutoff. Conversely, limiting the calculation to a subspace size of $L=5$ with untruncated moments would set the number of moments at $M=10$, which has been shown here and elsewhere[21] to be inadequate for Si (although it appears to be adequate for metals[21,26–28]).

For use as a short-ranged interatomic potential form with exact energy derivatives, improved accuracy could be obtained by refitting the tight-binding parameters specifically for the desired values of M and L. However, use of a truncation range that is too short may cause unphysical behavior that cannot by mitigated by adjusting the TB parameters.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Department of Energy, Office of Basic Energy Sciences, for funding this work, and thank S. Goedecker, H. Röder, I. Kwon, and T. Lenosky for illuminating discussions.

APPENDIX

Here we derive Eq. (31), the derivative of a Chebyshev polynomial of a matrix. We write the Chebyshev recurrence relation as

$$T_{m+1}=(1+k_m)\mathbf{X}T_m-T_{m-1}. \tag{A1}$$

Here T_α is understood to mean $T_\alpha(\mathbf{X})$, where $\mathbf{X}$ is a matrix, and

$$k_\alpha = \begin{cases} 1 & \text{if } \alpha \geqslant 1 \\ 0 & \text{if } \alpha \leqslant 0. \end{cases} \tag{A2}$$

Differentiating (A1) gives

$$T'_{m+1} = (1+k_m)\mathbf{X}'T_m + (1+k_m)\mathbf{X}T'_m - T'_{m-1}. \tag{A3}$$

Replacing index m with $m-\alpha$, multiplying by $(1+k_\alpha)T_\alpha$, summing over α from zero to m, and reordering terms leads to

$$\sum_{\alpha=0}^{m} (1+k_\alpha)(1+k_{m-\alpha})T_\alpha \mathbf{X}' T_{m-\alpha}$$

$$= \sum_{\alpha=0}^{m} (1+k_\alpha)T_\alpha T'_{m+1-\alpha} + \sum_{\alpha=0}^{m} (1+k_\alpha)T_\alpha T'_{m-1-\alpha}$$

$$- \sum_{\alpha=0}^{m} (1+k_\alpha)(1+k_{m-\alpha})T_\alpha \mathbf{X}T'_{m-\alpha}. \tag{A4}$$

Using the recurrence relation in the form $(1+k_\alpha)T_\alpha\mathbf{X} = T_{\alpha+1} + T_{\alpha-1}$ to modify the terms in the negative sum, the right-hand side of (A4) becomes

$$\chi_{\mathrm{RHS(A4)}} = \sum_{\alpha=0}^{m} (1+k_\alpha)T_\alpha T'_{m+1-\alpha} + \sum_{\alpha=0}^{m} (1+k_\alpha)T_\alpha T'_{m-1-\alpha}$$

$$- \sum_{\alpha=0}^{m} (1+k_{m-\alpha})T_{\alpha+1} T'_{m-\alpha}$$

$$- \sum_{\alpha=0}^{m} (1+k_{m-\alpha})T_{\alpha-1} T'_{m-\alpha}. \tag{A5}$$

Shifting the indices on the negative terms (so that T always appears with an α subscript),

$$\chi_{\mathrm{RHS(A4)}} = \sum_{\alpha=0}^{m} (1+k_\alpha)T_\alpha T'_{m+1-\alpha} + \sum_{\alpha=0}^{m} (1+k_\alpha)T_\alpha T'_{m-1-\alpha}$$

$$- \sum_{\alpha=1}^{m+1} (1+k_{m+1-\alpha})T_\alpha T'_{m+1-\alpha}$$

$$- \sum_{\alpha=-1}^{m-1} (1+k_{m-1-\alpha})T_\alpha T'_{m-1-\alpha}, \tag{A6}$$

and noting that $T_0=1$, $k_0=0$, $T'_0=0$ and $T_{-1}=0$, and inserting (A2),

$$\chi_{\mathrm{RHS(A4)}} = T'_{m+1} + \sum_{\alpha=1}^{m} 2T_\alpha T'_{m+1-\alpha} + T'_{m-1}$$

$$+ \sum_{\alpha=1}^{m-2} 2T_\alpha T'_{m-1-\alpha} - \sum_{\alpha=1}^{m} 2T_\alpha T'_{m+1-\alpha} - 2T'_{m-1}$$

$$- \sum_{\alpha=1}^{m-2} 2T_\alpha T'_{m-1-\alpha}$$

$$= T'_{m+1} - T'_{m-1}. \tag{A7}$$

Thus, from (A7) and (A4),

$$T'_{m+1} = T'_{m-1} + \sum_{\alpha=0}^{m} (1+k_\alpha)(1+k_{m-\alpha})T_\alpha \mathbf{X}' T_{m-\alpha}, \tag{A8}$$

which matches Eq. (31) if the index m is replaced with $m-1$.

[1] G. Galli, and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992).
[2] X.-P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10 891 (1993).
[3] M. S. Daw, Phys. Rev. B **47**, 10 895 (1993).
[4] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993).
[5] P. Ordejon, D. A. Drabold, M. Grumbach, and R. Martin, Phys. Rev. B **48**, 14 646 (1993).
[6] D. A. Drabold and O. F. Sankey, Phys. Rev. Lett. **70**, 3631 (1993).
[7] F. Mauri and G. Galli, Phys. Rev. B **50**, 4316 (1994).
[8] E. B. Stechel, A. R. Williams, and P. J. Feibelman, Phys. Rev. B **49**, 10 088 (1994).
[9] S. Goedecker and L. Colombo, Phys. Rev. Lett. **73**, 122 (1994).
[10] R. N. Silver, H. Röder, A. F. Voter, and J. D. Kress, in *Simulation MultiConference '95 Proceedings, High Performance Computing '95*, edited by A. Tentner (Society for Computer Simulation, San Diego, 1995), p. 200.
[11] R. N. Silver, H. Röder, A. F. Voter, and J. D. Kress, J. Comput. Phys. **124**, 115 (1996).
[12] H. Röder, R. N. Silver, H. D. Kress, G. A. Landrum, D. Drabold, and J. J. Dong, in Simulation MultiConference '96 Proceedings, High Peformance Computing '96, edited by A. Tentner (Society for Computer Simulation, San Diego, in press).
[13] S. Goedecker and M. Teter, Phys. Rev. B **51**, 9455 (1995).
[14] S. Goedecker, J. Comput. Phys. **118**, 261 (1995).
[15] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).
[16] L. Goodwin, A. J. Skinner, and D. G. Pettifor, Europhys. Lett. **9**, 701 (1989).
[17] A. P. Sutton, M. W. Finnis, D. G. Pettifor, and Y. Ohta, J. Phys. C **21**, 35 (1988).
[18] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).
[19] I. Kwon, R. Biswas, C. Z. Wang, K. M. Ho, and C. M. Soukoulis, Phys. Rev. B **49**, 7242 (1994).
[20] C. Z. Wang, C. T. Chan, and K. M. Ho, Phys. Rev. Lett. **66**, 189 (1991).
[21] J. D. Kress and A. F. Voter, Phys. Rev. B **52**, 8766 (1995).
[22] R. N. Silver and H. Röder, Int. J. Mod. Phys. C **5**, 735 (1994).
[23] B. N. Parlett, *The Symmetric Eigenvalue Problem* (Prentice-Hall, Englewood Cliffs, NJ, 1980).
[24] D. Jackson, *Theory of Approximations* (American Mathematical Society Colloquium Publications, Providence, RI, 1930), Vol. XI.
[25] G. A. Baraff and M. Schluter, Phys. Rev. B **30**, 3460 (1984).
[26] A. E. Carlsson, Phys. Rev. B **44**, 6590 (1991).
[27] J. A. Moriarty and R. Phillips, Phys. Rev. Lett. **66**, 3036 (1991).
[28] S. M. Foiles, Phys. Rev. B **48**, 4287 (1993).