

## Order- $N$ method for a nonorthogonal tight-binding Hamiltonian

C. S. Jayanthi and S. Y. Wu

*Department of Physics, University of Louisville, Louisville, Kentucky 40292*

J. Cocks

*Information Technology, University of Louisville, Louisville, Kentucky 40292*

N. S. Luo and Z. L. Xie

*Department of Physics, University of Louisville, Louisville, Kentucky 40292*

M. Menon

*Department of Physics and Astronomy, University of Kentucky, Lexington, Kentucky 40506*

G. Yang

*Department of Physics, University of Louisville, Louisville, Kentucky 40292*

(Received 8 August 1997)

We have developed an order- $N$  method to calculate the total energy and atomic forces for a nonorthogonal tight-binding Hamiltonian. We have first confirmed the validity of our approach by comparing our results with the exact calculation for a small silicon cluster. The efficiency of our approach for a large system is demonstrated by determining the equilibrium configuration of a  $\text{Si}_{1000}$  cluster using quenching and annealing techniques based on molecular dynamics. [S0163-1829(98)02908-7]

Total energy calculations and molecular dynamics (MD) simulations with forces derived from first-principles or tight-binding Hamiltonians are the central techniques of many materials science problems.<sup>1</sup> However, the application of these methods to study properties of complex systems of realistic sizes ( $N \sim 10^3 - 10^6$ ) is presently restricted by the  $N^3$  scaling in the computational time of total energy and atomic forces. To overcome this bottleneck, several linear scaling methods or order- $N$  procedures [ $O(N)$ ] have been proposed in recent years.<sup>2-13</sup> Most of the implementations of the  $O(N)$  procedure have been developed for the orthogonal tight-binding Hamiltonian. The  $O(N)$  techniques may be roughly grouped into two categories: Approaches based on (i) the density matrix<sup>2,3,7,8,10-13</sup> and (ii) localized wave functions (LWF).<sup>4-6,9</sup> The  $O(N)$  scaling, in these approaches, arises from the decay and/or truncation<sup>14</sup> of these respective quantities in real space. These methodologies, in turn, vary a great deal in their  $O(N)$  approximations and approaches to calculate the density matrix or the choice of local wave function. Some approaches<sup>2,3,12</sup> use the recursion method<sup>15</sup> to directly calculate the density matrix. Others expand the density matrix in terms of powers of the Hamiltonian matrix.<sup>10,13</sup> Variational schemes are also available which use either the density matrix elements<sup>7</sup> or the coefficients of expansion of the wave functions<sup>9</sup> as variational parameters. In the implementation of the variational schemes, further constraints such as the purification transformation<sup>16</sup> in the case of the density-matrix-based approach or the truncation of the expansion of  $S^{-1}$  (with  $S$  being the overlap matrix) in the LWF approach are also invoked. These  $O(N)$  methods have been used to carry out a restricted-geometry optimization of systems of large sizes.<sup>17</sup>

Recently, Mckinnon and Choy<sup>18</sup> have shown that, starting from a nonorthogonal basis with only nearest-neighbor overlaps, the process of orthogonalization gives rise to terms be-

yond the nearest-neighbor hopping terms in the tight-binding Hamiltonian. Based on the results of their work, they conclude that a more appropriate starting point for tight-binding calculations is a nonorthogonal localized basis set. More importantly, in a tight-binding molecular-dynamics simulation, it is impossible to maintain the orthogonality of the basis set at every time step. This is because the condition of orthogonality is environment-dependent. Although the orthogonality condition for a given configuration may be achieved via the Löwdin transformation, the orthogonality of the basis set corresponding to a given configuration will no longer be operational for the next configuration in a MD simulation as the system relaxes under the action of its own atomic forces. Thus, to develop a transferable set of tight-binding parameters, which is applicable for different local environments as encountered in MD simulations, it is more advantageous to use a nonorthogonal tight-binding Hamiltonian. Reliable and transferable sets of nonorthogonal tight-binding parameters for carbon and silicon are now available in the literature.<sup>19</sup>

An  $O(N)$  scheme for the calculation of total energy using a nonorthogonal tight-binding Hamiltonian (NOTB) has been reported in the literature for the variational approach based on the density matrix.<sup>20</sup> In a nonorthogonal basis set the atomic forces derived from the Hellman-Feynman theorem depend on, in addition to the density-matrix elements, extra terms related to the nonorthogonality (overlap) of the basis set [see Eq. (4) below]. Since these terms cannot be evaluated from the density-matrix elements, the variational approach for the NOTB Hamiltonian cannot yield an  $O(N)$  scheme for the simultaneous calculations of atomic forces and the total energy.

In this communication, we present an  $O(N)$  approach for the calculation of both the total energy as well as atomic forces for a NOTB Hamiltonian.<sup>21</sup> Our  $O(N)$  scheme for energy relies on the fact that the pseudo-density-matrix ele-

ments [see Eq. (2)] decay in real space. The  $O(N)$  scheme for atomic forces has an additional requirement, as will be discussed later. To demonstrate the efficiency of our  $O(N)$  scheme, we study the equilibrium configuration of a large silicon cluster by carrying out a full geometry search.

In the empirical NOTB scheme, the Hamiltonian is defined in terms of parametrized matrix elements  $H_{i\alpha,j\beta}(R_{ij})$  in some finite set of basis functions  $\{\varphi_{i\alpha}(\mathbf{r})\}$ , not explicitly stated, where  $i\alpha$  denotes the  $\alpha$  orbital at the site  $i$ . These parametrized matrix elements are usually constructed by fitting to an experimental and/or first-principles database of properties of both bulk systems as well as clusters. Expanding the wave function  $\Psi_\lambda$  of  $H$  in a finite set of basis functions, we have  $\Psi_\lambda = \sum_{i\alpha} c_{i\alpha\lambda} \varphi_{i\alpha}(\mathbf{r})$ , where the column vector of the coefficients,  $c_\lambda$ , satisfies the general eigenvalue equation  $Hc_\lambda = E_\lambda Sc_\lambda$  with  $c_\lambda$  taking the role of the wave function and  $c_\lambda^T Sc_\lambda = 1$ . The key to the empirical NOTB scheme is that  $H_{i\alpha,j\beta}(R_{ij}) = \int \varphi_{i\alpha}^*(\mathbf{r}) H \varphi_{j\beta}(\mathbf{r}) d\mathbf{r}$  and  $S_{i\alpha,j\beta}(R_{ij}) = \int \varphi_{i\alpha}^*(\mathbf{r}) \varphi_{j\beta}(\mathbf{r}) d\mathbf{r}$  are parametrized functions of  $R_{ij}$ , which are used as inputs to the general eigenvalue equation. The electronic energy of the system in the nonorthogonal basis can be written as

$$E_{\text{el}} = 2 \sum_{\lambda(\text{occ})} E_\lambda = 2 \sum_{i\alpha,j\beta} \gamma_{i\alpha,j\beta} H_{j\beta,i\alpha}, \quad (1)$$

where<sup>22</sup>

$$\gamma_{i\alpha,j\beta} = -\frac{1}{\pi} \lim_{\varepsilon \rightarrow 0} \int^{E_F} \text{Im} G_{i\alpha,j\beta}(E + i\varepsilon) dE \quad (2)$$

and  $G$  is the generalized Green's function defined by  $G \equiv (ES - H)^{-1}$ . The quantity  $\gamma_{i\alpha,j\beta}$  defined in Eq. (2) is referred as the “pseudo-density-matrix.” It is related to the density-matrix elements  $\rho_{i\alpha,j\beta}$  in the nonorthogonal basis set by  $\rho_{i\alpha,j\beta} = \sum_{k\gamma,l\delta} \gamma_{k\gamma,l\delta} S_{i\alpha,k\gamma} S_{l\delta,j\beta}$ . Thus,  $\gamma_{i\alpha,j\beta}$ 's become identical to the density-matrix elements for an orthogonal basis set. Furthermore,  $\gamma_{i\alpha,j\beta}$ 's are related to the total number of electrons in the system,  $N$ , via the following expression:

$$N = 2 \sum_{i\alpha,j\beta} \gamma_{i\alpha,j\beta} S_{j\beta,i\alpha}. \quad (3)$$

Therefore, in some sense,  $\gamma_{i\alpha,j\beta}$  is also characterized by its relationship to the Fermi energy,  $E_F$ . In fact, an accurate determination of the pseudo-density-matrix requires calculating  $\gamma_{i\alpha,j\beta}$ ,  $N$ , and  $E_F$  self-consistently [see Eqs. (2) and (3)].

In the NOTB approach, the electronic contribution to the force acting on the  $i$ th atom can be evaluated by the Hellman-Feynman theorem as

$$\mathbf{F}_{i,\text{el}} = -\frac{\partial E_{\text{el}}}{\partial \mathbf{R}_i} = -4 \sum_{j\beta} \left( \gamma_{i\alpha,j\beta} \frac{\partial H_{j\beta,i\alpha}(R_{ij})}{\partial \mathbf{R}_i} - \delta_{i\alpha,j\beta} \frac{\partial S_{j\beta,i\alpha}(R_{ij})}{\partial \mathbf{R}_i} \right), \quad (4)$$

where  $\delta_{i\alpha,j\beta} = -1/\pi \lim_{\varepsilon \rightarrow 0} \int^{E_F} E \text{Im} G_{i\alpha,j\beta}(E + i\varepsilon) dE$ .

For the tight-binding Hamiltonians, the total energy is given by  $E_{\text{tot}} = E_{\text{el}} + E_{\text{rep}}$ , where  $E_{\text{rep}}$  is the sum of pairwise

repulsive terms which are also obtained by fitting. Therefore, the total force acting on a given atom, say the atom at the site  $i$ , is given by  $\mathbf{F}_i = \mathbf{F}_{i,\text{el}} - \partial E_{\text{rep}} / \partial \mathbf{R}_i$ . In the NOTB scheme, Eqs. (1)–(4) are the building blocks for the calculation of the total energy and the atomic forces.

The feasibility of developing a linear scaling algorithm for the total energy and atomic forces calculated within the NOTB scheme depends on the decay of the “pseudo-density-matrix” in real space, namely,  $\gamma_{i\alpha,j\beta}(R_{ij}) \rightarrow 0$  as  $R_{ij} \rightarrow \infty$ . The  $O(N)$  procedure for the atomic force requires in addition that  $\delta_{i\alpha,j\beta}(R_{ij}) \rightarrow 0$  as  $R_{ij} \rightarrow \infty$ . Rewriting Eq. (1) in the form  $E_{\text{el}} = 2 \sum_{i\alpha} (\sum_{j\beta} \gamma_{i\alpha,j\beta} H_{j\beta,i\alpha})$ , the summation over  $j$  in the parentheses can be truncated to include only terms within a sphere of radius  $R_0$  about  $i$  if it can be established that  $\gamma_{i\alpha,j\beta}(R_{ij}) \approx 0$  for  $R_{ij} > R_0$ . With this condition satisfied, one can see from the rewritten expression for the electronic energy that the computational time for the calculation of the total energy will depend linearly on the size of the system. However, the  $E_{\text{el}}$  calculated using the truncated summation over  $j$  will be accurate only if  $\gamma_{i\alpha,j\beta}(R_{ij})$  for  $R_{ij} \leq R_0$  can be determined accurately.

Our experience with the real-space Green's function (RSGF) indicates that an accurate calculation of  $G_{i\alpha,j\beta}(R_{ij})$  can be accomplished using a convergence procedure<sup>23</sup> with the radius of convergence given by  $R_c = R_{ij} + R_{\text{bf}}$ , defines the size of the buffer zone so that  $G_{i\alpha,j\beta}(R_{ij})$  will have included sufficient influence from the region beyond the interior region defined by  $R_{ij}$  about the site  $i$ . Our experience also shows that  $R_{\text{bf}}$  includes on the average about two shells of atoms more than those included in the region defined by  $R_{ij}$ . Furthermore, because the calculation of  $\gamma_{i\alpha,j\beta}$  involves the “integration” of  $G_{i\alpha,j\beta}$ , it usually requires less stringent convergence on the Green's function in the calculation of  $\gamma_{i\alpha,j\beta}$ . Therefore, the  $O(N)$  procedure for the calculation of the electronic energy will be accurate if the summation over  $j$  is truncated to include terms within a sphere of radius  $R_c = R_0 + R_{\text{bf}}$ . Similarly, in the calculation of atomic forces, an accurate determination of  $\delta_{i\alpha,j\beta}(R_{ij})$  can be achieved within the radius  $R_c = R_0 + R_{\text{bf}}$ .

The method of RSGF shifts the computational effort of inverting large matrix to matrix multiplications and inver-

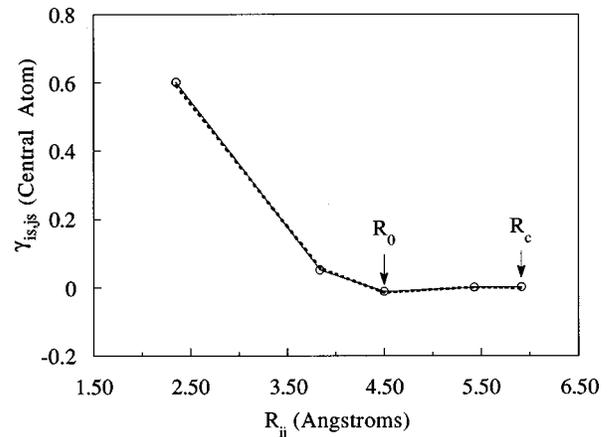


FIG. 1. The “pseudo-density” matrix element between the central atom of a  $\text{Si}_7$  cluster and its neighbors as a function of  $R_{ij}$  for  $s$ - $s$  orbitals calculated by direct diagonalization (open circles) and our  $O(N)$  calculation (dashed line).

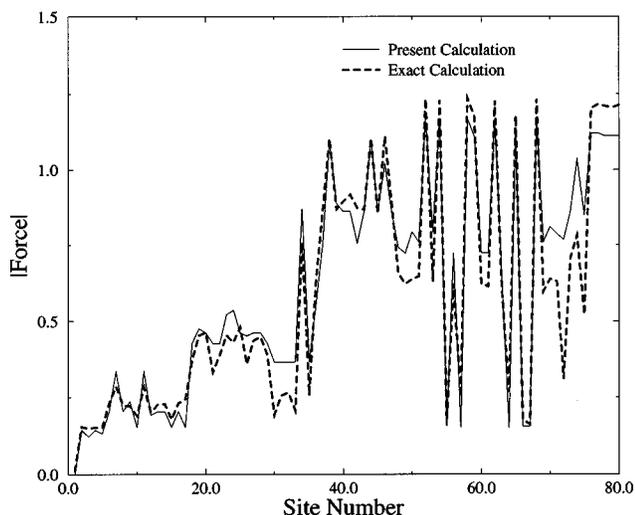


FIG. 2. A comparison of the magnitudes of forces acting at each site of an unstable  $\text{Si}_{80}$  cluster as obtained by our  $O(N)$  calculation (solid line) with the direct diagonalization (dashed line).

sions of a sequence of matrices of dimension smaller than the original matrix.<sup>24</sup> This feature allows the calculation of  $G_{i\alpha,j\beta}$  for  $R_{ij} \leq R_0$  without actually having to invert the entire matrix corresponding to  $R_{ij} = R_c$ . Hence this method leads to an efficient numerical calculation of  $\gamma_{i\alpha,j\beta}$ .<sup>25</sup>

Since our  $O(N)$  scheme hinges on the decay of the pseudo-density-matrix, we first demonstrate that the condition  $\gamma_{i\alpha,j\beta}(R_{ij}) \sim 0$  for  $R_{ij} > R_0$  is satisfied within the context of NOTB by considering the case of a  $\text{Si}_{87}$  cluster in an unstable tetrahedral network configuration. This cluster is small enough so that direct diagonalization can be used to obtain the exact solution and yet sufficiently large to demonstrate the condition stated above. We chose the unstable structure as the working example because we are also interested in comparing the forces acting on each atom between those obtained by the exact calculation and those by our NOTB/ $O(N)$  procedure. We used the NOTB parametrized matrix elements developed by Menon and Subbaswamy<sup>19</sup> to

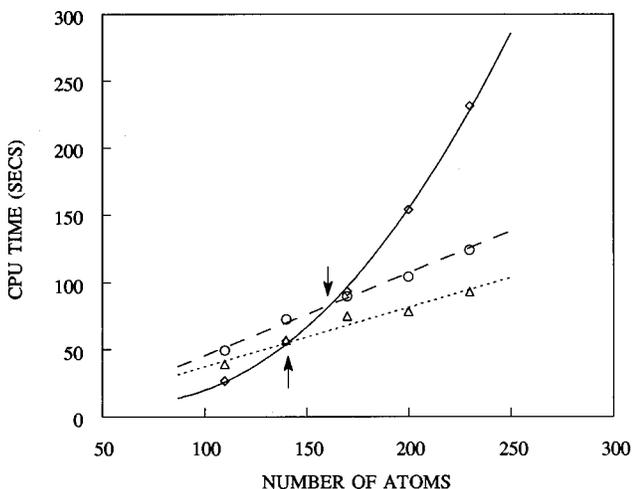
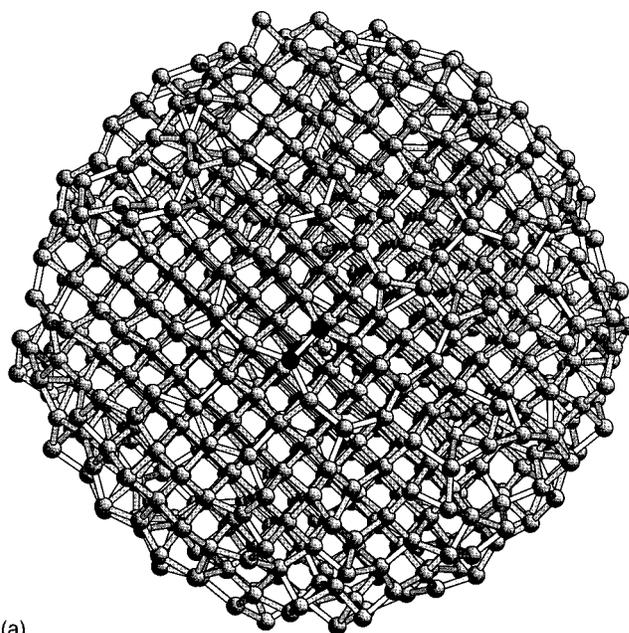
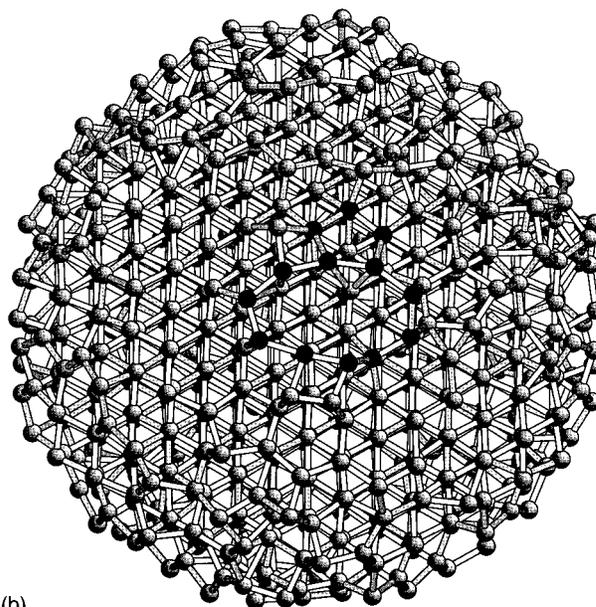


FIG. 3. The CPU time for total energy as obtained by direct diagonalization (solid line) is compared with our  $O(N)$  method calculated for  $R_c = 4.6 \text{ \AA}$  (dashed line with triangles) and  $R_c = 5.6 \text{ \AA}$  (dotted line with circles), respectively.



(a)



(b)

FIG. 4. (a) Top view of the  $\text{Si}_{1000}$  as viewed along the  $\langle 100 \rangle$  direction. The dimer associated with the reconstruction of  $\text{Si}(100)$  “surface” is highlighted. (b) Top view of the  $\text{Si}_{1000}$  as viewed along the  $\langle 111 \rangle$  direction. The dimer-atom stacking fault region associated with the reconstruction of  $\text{Si}(111)$  surface is highlighted.

model the  $\text{Si}_N$  cluster. In Fig. 1, the “pseudo-density-matrix” element between the central atom of a  $\text{Si}_{87}$  cluster and its neighbors, as obtained by direct diagonalization (open circles), is shown as a function of  $R_{ij}$  for  $s$ - $s$  orbitals. It can be seen that  $\gamma_{i\alpha,j\beta}(R_{ij}) \approx 0$  for  $R_{ij} \geq$  third nearest-neighbor distance, confirming that a NOTB/ $O(N)$  procedure is feasible. This calculation indicates that  $R_0$  for Si is of the order of magnitude  $4.50 \text{ \AA}$ . Our order ( $N$ ) procedure for the calculation of the total energy and the atomic forces starts with the selection of  $R_c$  such that  $\gamma_{i\alpha,j\beta}(R_{ij})$  needed for an accurate evaluation of the electronic energy and atomic forces can be determined reliably. Recalling that  $R_c = R_0 + R_{\text{bf}}$ , the exact result (open circles) shown in Fig. 1 suggests that  $R_c$

=fifth nearest-neighbor distance ( $R_c \approx 6.00 \text{ \AA}$ ) will be a good choice for Si. Using this value of  $R_c$  and the method of RSGF,<sup>23,24</sup> one can obtain an accurate and efficient determination of  $\gamma_{i\alpha,j\beta}(R_{ij})$  for  $R_{ij} \leq 4.50 \text{ \AA}$ . We then choose an  $E_F$  and compute all the  $\gamma_{i\alpha,j\beta}(R_{ij})$ 's according to Eq. (2). They are substituted into Eq. (3) to calculate the total number of electrons so as to check the consistency between the  $E_F$  and  $N$ . If the Fermi energy chosen does not yield the correct total number of electrons, we repeat the calculation of  $\gamma_{i\alpha,j\beta}$  with new value of  $E_F$  until the consistency between  $E_F$  and  $N$  is achieved. The resulting  $\gamma_{i\alpha,j\beta}(R_{ij})$ 's and  $\delta_{i\alpha,j\beta}$ 's (corresponding to the final  $E_F$ ) are used to calculate  $E_{el}$  and  $\mathbf{F}_{el}$  according to Eqs. (1) and (4), respectively. In Fig. 1, the dashed curve gives  $\gamma_{i\alpha,j\beta}(R_{ij})$  as a function of  $R_{ij}$  using our  $O(N)$  scheme for the same case as discussed above. The agreement between the solid curve and the dashed curve is outstanding up to the third nearest neighbor, and for  $R_{ij} >$  third nearest neighbor distance ( $4.50 \text{ \AA}$ ),  $\gamma_{i\alpha,j\beta}(R_{ij}) \approx 0$ , as in the exact calculation. This result confirms the reliability of our  $O(N)$  procedure and the validity of the approximations used such as the truncation of the summation over  $j$  within the sphere of radius  $R_c$ . We have also compared the accuracy of our results with the direct diagonalization (exact result) for  $n_{i\alpha} = \sum_{R_{ij} \leq R_0} \gamma_{i\alpha,j\beta} S_{j\beta,i\alpha}$ ,  $E_F$ , and the total energy for the  $\text{Si}_{87}$  cluster in the tetrahedral network structure. We found an excellent agreement between our  $O(N)$  calculation and the exact result (less than 1% error). In Fig. 2, we compare the magnitude of atomic forces acting on each atom of a  $\text{Si}_{80}$  cluster in the tetrahedral network structure with the exact result and again find excellent agreements. Finally, the efficiency of our order ( $N$ ) procedure is demonstrated in Fig. 3 where the CPU time for the calculation of the total energy is plotted against the number of atoms in the system for different  $R_c$ 's. Our order ( $N$ ) scheme indeed shows a linear scaling while the exact diagonalization shows an  $N^3$  dependence. Depending on the value of  $R_c$  chosen, the break-even size in the CPU time for total energy is different.

To demonstrate the efficiency of our  $O(N)$ /NOTB-MD scheme, we have determined the stable configuration of a  $\text{Si}_{1000}$  cluster by allowing it to relax under the action of atomic forces from an initial diamond configuration. The optimization is gauged by the magnitudes of the atomic forces and the total energy. A stable equilibrium configuration is considered to have been reached if the force acting on each atom is very small and the total energy of the configuration is at a minimum. Top views of the stable cluster corresponding to two orientations (along  $\langle 100 \rangle$  and  $\langle 111 \rangle$  directions) are shown in Figs. 4(a) and 4(b), respectively. Typically, the atomic forces for the structure shown are in the range of  $10^{-3} \text{ eV/\AA}$  with a few in the range of  $0.05 \text{ eV/\AA}$ . The configuration shown has a cohesive energy of  $4.80 \text{ eV}$ . In Fig. 4(a), it can be seen that the interior of the cluster already has the bulklike tetrahedral structure. On the "surface," one may discern the appearance of dimers associated with the Si(100) reconstruction. Considering the fact that the "surface" of a 1000-atom Si cluster is not large enough to accommodate the  $7 \times 7$  reconstruction of the Si(111) surface, it is interesting to note that features similar to the dimer-atom-stacking fault (DAS) associated with the  $7 \times 7$  reconstruction<sup>26</sup> of the Si(111) surface appear in Fig. 4(b). These features suggest that precursors to the most stable surfaces, namely (111) and (100), are the first to evolve for the  $\text{Si}_{1000}$  cluster. The calculation of atomic forces at 1000 sites takes approximately 1.35 min on a Convex/Exemplar with 16 processors (HP 735). Therefore, 1000 MD steps for a  $\text{Si}_{1000}$  cluster can be accomplished in less than a day.

In summary, we have developed an  $O(N)$  procedure for the nonorthogonal tight-binding (NOTB) Hamiltonian. The framework of our  $O(N)$  approach can be used in the context of the first-principles methods such as TB-LMTO in real space.<sup>22</sup>

This work was supported by the National Science Foundation (Grant Nos. OSR-9452895 and PHY94-07194) and the Research Corporation (Cottrell College of Science Grant). We acknowledge the usage of facilities at the supercomputing center of the University of Kentucky.

- 
- <sup>1</sup>R. Car, in *Quantum Theory of Real Materials*, edited by J. R. Chelikowsky and S. G. Louie (Kluwer Academic Publishers, Boston, 1996), p. 23.
- <sup>2</sup>S. Baroni and P. Giannozzi, *Europhys. Lett.* **17**, 547 (1992).
- <sup>3</sup>W. Zhang *et al.*, *Solid State Commun.* **86**, 607 (1993).
- <sup>4</sup>W. Yang, *Phys. Rev. Lett.* **66**, 1438 (1991).
- <sup>5</sup>G. Galli and M. Parrinello, *Phys. Rev. Lett.* **69**, 3547 (1992).
- <sup>6</sup>F. Mauri *et al.*, *Phys. Rev. B* **47**, 9973 (1993).
- <sup>7</sup>X.-P. Li *et al.*, *Phys. Rev. B* **47**, 10 891 (1993).
- <sup>8</sup>M. Daw, *Phys. Rev. B* **47**, 10 895 (1993).
- <sup>9</sup>P. Ordejon *et al.*, *Phys. Rev. B* **48**, 14 646 (1993).
- <sup>10</sup>S. Goedecker and L. Colombo, *Phys. Rev. Lett.* **73**, 122 (1994).
- <sup>11</sup>S. Goedecker, *J. Comput. Phys.* **118**, 261 (1995).
- <sup>12</sup>A. P. Horsfield *et al.*, *Phys. Rev. B* **53**, 12 694 (1996).
- <sup>13</sup>U. Stephan and D. A. Drabold, *Phys. Rev. B* (to be published).
- <sup>14</sup>The decay of the density matrix (exponential for insulators and semiconductors, and power law for metals) is related to the decay length of the Wannier functions; see W. Kohn, *Phys. Rev.* **115**, 809 (1959); *Phys. Rev. B* **7**, 4388 (1973).
- <sup>15</sup>R. Haydock, V. Heine, and M. J. Kelly, *J. Phys. C* **5**, 2845 (1975).
- <sup>16</sup>R. McWeeny, *Rev. Mod. Phys.* **32**, 335 (1960).
- <sup>17</sup>See, for example, J. P. Lu and W. Yang, *Phys. Rev. B* **49**, 11 421 (1994); see also S. Itoh *et al.*, *ibid.* **53**, 2132 (1996).
- <sup>18</sup>B. A. McKinnon and T. C. Choy, *Phys. Rev. B* **52**, 14 531 (1995).
- <sup>19</sup>M. Menon and K. R. Subbaswamy, *Phys. Rev. B* **55**, 9231 (1997).
- <sup>20</sup>R. W. Nunes and D. Vanderbilt, *Phys. Rev. B* **50**, 17 611 (1994).
- <sup>21</sup>C. S. Jayanthi *et al.*, *Bull. Am. Phys. Soc.* **42**, 325 (1997).
- <sup>22</sup>Z. L. Xie *et al.*, *Phys. Rev. B* **55**, 1748 (1997).
- <sup>23</sup>S. Y. Wu *et al.*, *Phys. Rev. B* **49**, 7957 (1994).
- <sup>24</sup>S. Y. Wu *et al.*, *Comput. Phys. Commun.* **71**, 125 (1992).
- <sup>25</sup>C. S. Jayanthi *et al.* (unpublished).
- <sup>26</sup>K. Takayanagi *et al.*, *Surf. Sci.* **164**, 367 (1985).