PHYSICAL REVIEW B VOLUME 35, NUMBER 6 15 FEBRUARY 1987-II

Development of a many-body Tersoff-type potential for silicon

Brian W. Dodson
Sandia National Laboratories, Albuquerque, New Mexico 87185
(Received 27 August 1986)

An empirical potential for silicon recently introduced by Tersoff has been found to be unusable for global structural calculations since the bcc structure has lower energy than the diamond structure, by roughly 1 eV/atom, with this potential. Simulated annealing optimization studies of the fitting parameters to the quantum-mechanical data base for silicon reveal that serious structural pathologies (relative to the behavior of silicon) cannot be avoided. We introduce a modification of the Tersoff potential which retains the accurate treatment of surface and defect properties exhibited by that potential but which yields a reasonable fit to the bulk and high-density structural properties of silicon. The resulting potential can be used for global structural calculations.

INTRODUCTION

There has been considerable recent interest in atomistic simulation of the structural properties of tetrahedral semiconductors. This interest is fueled by the various surface reconstructions observed in silicon and by the need to understand nonequilibrium structures and processing techniques which are finding application in semiconductor devices. Because of the open tetrahedral structure of bulk silicon, simple pairwise potentials are inadequate. Further, since the interest is generally in studying rather large, nonperiodic structures, calculation of the structural energies using quantum mechanics is not practical. Therefore, recent emphasis has been on developing classical empirical three-body potentials. Such potentials can stabilize the diamond structure while remaining tractable on the problems of interest.

The oldest empirical three-body potential for diamond-phase semiconductors is the valence-force potential.[1] The essence of this approach, which has been used with considerable success for study of phonon and small-deformation elastic properties, is to expand the potential energy in terms of atomic displacements around some reference structure. In the case of silicon, this reference structure is the diamond phase. Such potentials are perturbative in nature, and cannot properly be applied to systems with nonideal topology, such as defect structures, surfaces, and melting.

To properly treat these more general structures, potentials which accurately span the entire available configuration space are needed. This need has been addressed in several forms in the last couple of years. The first attempt was that of Pearson, Takai, Halicioglu, and Tiller (PTHT).[2] They treated the pairwise interaction by a Lennard-Jones term and the three-body term by the Axilrod-Teller potential for van der Waals interactions of three bodies.[3] There is no physical reason to choose these terms to represent the behavior of silicon, and the potential is used simply as a fitting function. The PTHT potential does not predict reasonable elastic properties for bulk silicon or its high-pressure polymorphs, but fits small clusters fairly well.

Another attempt, which met with considerably more success, was made by Stillinger and Weber (SW),[4] who designed an empirical three-body potential to model the melting of silicon. They used an Lennard-Jones pairwise potential and included a term in the three-body potential which discriminates in favor of the ideal tetrahedral structure. This has some of the flavor of the valence-force-field approach, but is formulated so that the possibility of nonideal topologies is not excluded, but is only discouraged by the potential. Studies of the SW potential have indicated that it replicates the elastic properties of bulk silicon, but that it does not handle some surface structures, low-coordination-number geometries, or high-pressure polymorphs properly.[5] This potential has been used to study the melting of silicon,[4] faceting at the (100) crystal-melt interface,[6] and the stability of coherently strained SiGe layers on an Si(111) substrate.[7] All of these studies compared well with the corresponding experiments and/or overlapping continuum theory, lending credence to the applicability of the SW potential.

An improved class of empirical potentials for silicon and other covalently bonded semiconductors is based loosely on ideas from quantum chemistry and observations concerning the universality of mechanical behavior of solids.[8,9] The essential idea here, introduced first by Abell,[10] is that the bonding can be properly described by pairwise interactions (for which a Morse-type potential is a reasonable approximation), but the strength of the pairwise interaction is influenced by the local environment, e.g., by many-body interference terms. This picture of covalent binding was, apparently independently, developed and implemented by Biswas and Hamann for silicon.[11] They express the environment near the pair bond by expansion in Legendre polynomials which are then used to modify a generalized Morse-type pairwise potential. The (still empirical) three-body potential is then fit to the density-functional theory (DFT) structural data base for silicon.[12] The resulting potential works well on bulk elastic properties and the high-density polymorphs which make up the data base. However, the potential fails badly for very small clusters, and in addition is too time consuming for most large-scale calculations. Accordingly, this potential has not been widely tested despite its apparent promise.

Another approach closer to that outlined by Abell has recently been implemented for silicon by Tersoff.[13] This potential will be considered in greater detail, as it is the basis for the work described in this paper. The system energy is the sum over the asymmetric potential Φ_{ij}

$$E = 0.5 \sum_{i,j} \Phi_{ij} \ . \tag{1}$$

The pairwise potential Φ_{ij} has the Morse form

$$\Phi_{ij} = f_{ij}(A e^{-\lambda_1 r_{ij}} - B_{ij} e^{-\lambda_2 r_{ij}}) \ , \tag{2}$$

where

$$B_{ij} = B_0 \exp(-Z_{ij}/b) \ , \tag{3}$$

$$\lambda_1 = 2\lambda_2 \ ,$$

and

$$Z_{ij} = \sum_{k(k \neq i,j)} [f_{ik}/f_{ij}]^4 \exp[4\lambda_2(r_{ij} - r_{ik})] \times [c + \exp(-d \cos \Theta_{jik})]^{-1} \ , \tag{4}$$

where f_{ij} is a smooth cutoff term centered around 3.0 Å. The six parameters A, B_0, λ_2, b, c, and d were chosen by fitting to the lattice constant and bulk modulus of diamond-phase silicon and the cohesive energies of diamond, simple cubic, and face-centered cubic silicon and the silicon dimer, based on experimental values and DFT calculations by Yin and Cohen.[12]

The Tersoff potential does a good job of describing elastic properties of bulk silicon and a reasonable job of determining the energies of many defect structures and surface reconstructions, as well as low coordination number geometries.[13] Unfortunately, we have determined that the crystal structure with lowest energy is not the diamond phase. Among simple crystal structures, the lowest energy, by more than 1 eV/atom (see Fig. 1), is associated with the bcc phase. This is a serious limitation to the applicability of this potential. For example, if one melts and then cools a model base on the Tersoff potential, it will not recrystallize in the tetrahedral phase. It can be used in the original form for calculation of some defect structures, but clearly cannot be used in growth simulations or general stability calculations. It will always be difficult to determine if it can be applied to a new problem, because the tetrahedral phase is not stable. In the remainder of this paper, we describe a technique to perform global optimization of the Tersoff potential using simulated annealing of the parameter set. This technique is then used to demonstrate that the Tersoff potential is intrinsically pathological in its description of the global structural energetics of silicon. We then introduce a modification of the Tersoff potential form which is satisfactory for global structural calculations while retaining accurate treatment of surface configurations.

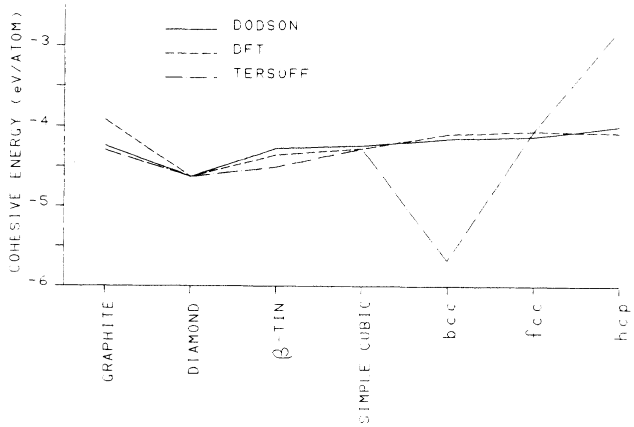

FIG. 1. Cohesive energies in eV/atom for various simple silicon lattices as generated by density-functional theory (DFT), the Tersoff potential, and the present work (Dodson).

GLOBAL OPTIMIZATION VIA SIMULATED ANNEALING

The problem of global optimization of multivariate functions has essentially no general theory or techniques for solution.[14] The object of optimization theory is to minimize an error function describing the deviation of a trial-fitting function from an ideal fit. This error function is often called the cost function. There are two general heuristic approaches in common use. These consist of iterative improvement of solutions based on searching parameter space for rearrangements which reduce the cost function and of procedures to determine the local stability of a given fit. These techniques usually get trapped in local minima, and their use for locating global minima is speculative at best. To partially overcome this tendency, it is customary to carry out the optimization process several times, starting from different initial parameters, and then accepting the best result. However, this technique is nearly useless in problems containing many low-lying local minima.

A new class of techniques designed to address the problem of global optimization was recently introduced by Kirkpatrick, Gelatt, and Vecchi (KGV).[15] These new techniques are collectively known as simulated annealing, and are based on an analogy between stochastic dynamics on thermodynamic phase space and the dynamics of optimization procedures on parameter space. In particular, consider recrystallization of a liquid. At high temperatures, the system is liquid, and the particles move freely. If the liquid is cooled slowly, order gradually begins to appear, first as a polycrystalline solid with defects and grain boundaries. As annealing (slow cooling) proceeds, the quality of the crystal improves until a single crystal is formed. The crystal is the lowest energy state of the system, and is the result of the slow-cooling process. If the cooling proceeds more rapidly, the final state will be polycrystalline, but will predominantly exhibit the bonding seen in the crystal. In the analogy to simulated annealing optimization, the cost function is the counterpart of the system energy. The temperature of the system is related to the permissible fluctuations of the cost function under perturbations of the parameter set. The liquid state represents the condition where the fluctuations of the cost function are large enough that the parameters may vary quite freely. As the "temperature" falls, the range of variation of the parameters reduces, and the correlations between the values of different parameters increases. The analog of the crystalline final state is the parameter set which results in the best fit based on the cost function chosen for optimization. The advantage of simulated annealing is that, in principle, the system can search over the complete configuration space to choose the global minimum, just as a thermodynamic system will (generally) visit all accessible regions of phase space.

One can simulate the annealing process by using one of several techniques to determine the path of the system through the parameter space as a function of temperature. These techniques include molecular dynamics, Monte Carlo methods, and Langevin dynamics. Although these techniques describe different paths through phase space, they all give the same thermodynamic averages, and thus can all be used for modeling the annealing process. We will concentrate on the Monte Carlo method. In this procedure, an atomic system is subjected to random changes in its configuration. The change is accepted if, as a result, the system energy decreases. If the energy increases, the change is accepted with probability $\exp(-\Delta E/k_B T)$, this being the Boltzmann probability distribution. This is known as the Metropolis Monte Carlo algorithm.[16] The essential point here is that a thermodynamic system cannot become trapped in a local-energy minimum when undergoing slow cooling, since changes in configuration which increase the energy are sometimes accepted. Thus, the system tends to climb out of local minima. The temperature determines the likely size of such upward motion, which becomes less important as the temperature decreases. Thus, sufficently slow cooling will result in a unique, or at worst a set of nearly degenerate, global minima.

To apply such an annealing procedure to nonthermodynamic problems requires the following features (the thermodynamic analog is listed in parenthesis).

(i) A parameter set description of the system (particle locations).

(ii) A procedure to randomly change the parameters (attempted random particle moves).

(iii) A cost function whose minimization is the goal of the optimization (system energy).

(iv) An annealing parameter T (temperature) and a schedule for annealing.

We are interested in optimizing the fit of an empirical potential function to various structures in the quantum mechanical data base, primarily as given by Yin and Cohen.[12] When we begin with the Tersoff potential [Eqs. (1)–(4)], the six parameters A, B_0, λ_2, b, c, and d form the parameter set analogous to particle locations. A procedure to randomly vary the parameter set is straightforward, as is the annealing parameter. The main difficulty is to develop an appropriate cost function. It is clear that one must develop some criterion for the quality of the fit to the quantum mechanical data base. It is also clear that such a criterion for fit will be, to some extent, arbitrary and inclusive of qualitative ideas concerning how the empirical potential should behave. We have chosen, for the sake of simplicity, a generalized difference-squared function as the energy function for the cost function:

$$E=\sum_k \alpha_i[\beta_i(\text{fit})-\beta_i]^2 , \tag{5}$$

where the sum is over the chosen points from the quantum mechanical data base. β_i is the desired value for the test structure i (from the data base), and $\beta_i(\text{fit})$ is the value for the test structure resulting from the current potential parameter set. (β can be system energy, interatomic distance, bulk modulus, etc.) The α_i are a measure of how important the fit for test structure i is felt to be. [This is where the arbitrariness of the energy analog enters; fortunately, the results are qualitatively quite insensitive to the details of the energy analog (for the Tersoff potential).]

The object of the simulated annealing optimization procedure is to find the parameter set for the Tersoff potential which optimizes the fit to the DFT data as described by the energy analog. The fitting goals are the cohesive energy, lattice parameter, and bulk modulus of tetragonal silicon, and the cohesive energy and lattice parameter of the silicon dimer and the graphitic, simple cubic, bcc, fcc, and hcp crystal phases of silicon, or subsets of the above. (Note that we have included more structural information than was used by Tersoff. This did not substantially improve the situation.) Fits are performed via Monte Carlo procedures as described earlier. Over a wide range of α_i values, we found that the diamond structure could not be made the global minimum among simple crystal structures without producing unacceptable errors in the energies and lattice constants of other important phases. The conclusion is that the Tersoff potential, as expressed in Eqs. (1)–(4), is not suitable for treatment of global structural energetics in silicon or, by extension, in other tetragonal semiconductors.

A MODIFIED TERSOFF-TYPE POTENTIAL FOR SILICON

We have been successful in developing a modified form for the potential based on the Tersoff potential. This modified potential retains the accurate treatment of surface structures but also yields the proper global limits as well as predicting reasonable properties for the high-density polymorphs. The formulation is essentially as expressed in Eq. (1)–(4), but the restriction that $\lambda_1=2\lambda_2$ is removed, and Eq. (3) becomes

$$B_{ij}=B_0 \exp[-(z_{ij})^{\eta}/b] . \tag{3'}$$

This modification seems to give us two extra parameters to adjust, but we will find that free adjustment actually gives $\lambda_1=2\lambda_2$ very nearly, so that the added freedom is not used.

Upon performing simulated annealing optimization of the parameter set for the modified potential described above, using as the data set the DFT calculations of Yin and Cohen[12] for lattice constant and cohesive energy for the diamond, simple cubic, bcc, fcc, and hcp phases, the bulk modulus of diamond phase silicon, and the atomic separation and binding energy of the Si_2 dimer, we find the parameters set as follows: $A = 1614.6$, $b = 3.4785$, $B_0 = 155.08$, $c = 0.8543$, $\lambda_1 = 2.7793$, $d = 3.9588$, $\lambda_2 = 1.3969$, $\eta = 0.6207$. It is noteworthy that η, the new exponent in Eq. (3a), which equaled 1 in the Tersoff potential, is now 0.6207. The size of the change of η indicates that our new potential is qualitatively different from that of Tersoff rather than representing a small quantitative change.

One can compare the performance of the new potential with that of Tersoff by examining Figs. 1 and 2, where we schematically plot the cohesive energy and equilibrium volume for a collection of high-density polymorphs. From Fig. 1, it is clear that the Tersoff bcc and hcp energies are respectively much too large and too small, and the corresponding interatomic distances (shown in Fig. 2) are also considerably in error. The tremendous improvement in the energies and equilibrium spacings of these phases resulting from use of the new potential is not due simply to their being included in the fitting routine for the new potential, but, as discussed earlier, is due to a qualitative change in the nature of the potential. Among the structures analyzed in Figs. 1 and 2, the graphitic and β-tin structures were included in neither fitting routine. The energy of the β-tin structure is much better modeled by the new potential, which gives an energy of 0.35 eV/atom above the diamond phase, compared with only 0.12 eV/atom for the Tersoff potential (the theoretical value is 0.27 eV). Neither potential works particularly well for the graphitic phase, predicting cohesive energies in the range of −4.24—4.30 eV/atom, compared to the theoretical value of −3.92 eV.[17] This difference is due to the π-bonding of the planar graphitic structure, which represents a difficult task for empirical potentials not specifically designed to treat π bonds.

The treatment of surface geometries by the new potential is rather similar to that provided by the Tersoff potential. To mention only two examples, we find that the (111) silicon surface relaxes inward by 30% of the first-layer separation, reducing the system energy by 0.12 eV/surface atom, compared to DFT results of (29—37)% and 0.15—0.17 eV/surface atom.[18] Also, the 2×1 (111) silicon reconstruction yields a reduction of energy of 0.12 eV/surface atom, compared to 0.35 eV/surface atom predicted by DFT.[19] Again, the effect of π bonding is seen, since this reconstruction depends fundamentally on such bonds. Most empirical potential models do not even predict this reconstruction.[13] These results for surface configurations are essentially indistinguishible from those of the Tersoff potential. Thus, the new potential seems to retain the many good features of the Tersoff potential while avoiding the obstructing pathologies.

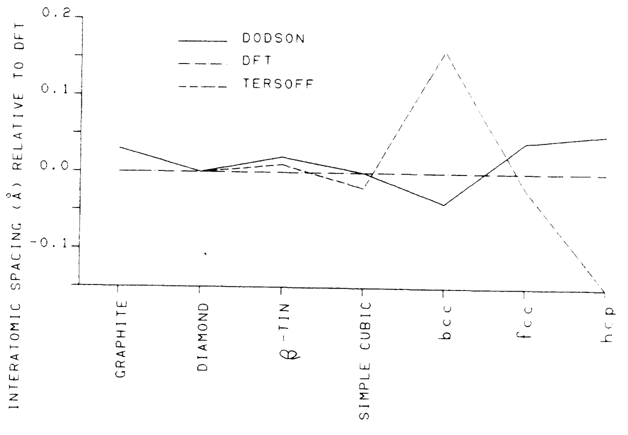

FIG. 2. Interatomic spacing in Å generated by the empirical potentials of Tersoff and Dodson relative to those predicted by density functional theory (DFT).

Note added. Tersoff has now developed a modified version of his original potential, which is currently undergoing evaluation by several groups.[20]

ACKNOWLEDGMENTS

This work was performed at Sandia National Laboratories and was supported by the U.S. Department of Energy under Contract No. DE-AC04-76DP00789.

---

[1] W. A. Harrison, *Electronic Structure* (Freeman, San Fransisco, 1980), Chap. 8.

[2] E. Pearson, T. Takai, T. Halicioglu, and W. A. Tiller, J. Cryst. Growth **70**, 33 (1984).

[3] B. M. Axilrod and E. Teller, J. Chem. Phys. **11**, 299 (1943).

[4] F. H. Stillinger and T. A. Weber, Phys. Rev. B **31**, 5262 (1985).

[5] B. W. Dodson, Phys. Rev. B **33**, 7361 (1986).

[6] U. Landman, W. D. Luedtke, R. N. Barnett, C. L. Cleveland, M. W. Ribarsky, E. Arnold, S. Ramseh, H. Baumgart, A. Martinez, and B. Khan, Phys. Rev. Lett. **56**, 155 (1986).

[7] B. W. Dodson and P. A. Taylor, App. Phys. Lett. **49**, 642 (1986).

[8] J. H. Rose, J. R. Smith, F. Guinea, and J. Ferrante, Phys. Rev. B **29**, 2963 (1984).

[9] B. W. Dodson, Phys. Rev. B **35**, 2619 (1987).

[10] G. C. Abell, Phys. Rev. B **31**, 6184 (1984).

[11] R. Biswas and D. R. Hamann, Phys. Rev. Lett. **55**, 2001 (1985).

[12] M. T. Yin and M. L. Cohen, Phys. Rev. B **26**, 3259 (1982).

[13] J. Tersoff, Phys. Rev. Lett. **56**, 632 (1986).

[14] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge, New York, 1986).

[15] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).

[16] K. Binder, in *Applications of the Monte Carlo Method* (Springer-Verlag, New York, 1984).

[17] M. T. Yin and M. L. Cohen, Phys. Rev. B **29**, 6996 (1984).

[18] J. E. Northrup and M. L. Cohen, Phys. Rev. Lett. **49**, 1349 (1982).

[19] K. C. Pandey, Phys. Rev. Lett. **49**, 223 (1982).

[20] J. Tersoff (private communication).