Chapter 5

The Environment-Dependent Interatomic Potential

> I have restricted my work to ideal crystals though I am aware that the theory of the defects in real crystals is practically far more important. This I have left to a younger generation.
>
> – Max Born [70]

It is ironic that in answering Born's challenge to model defect structures we begin with analytic results for ideal crystals (from the preceding two chapters), some of which can be traced back to Born himself. In order to develop a model for the complex bonding in real crystals, however, a modern computer is required to search efficiently through the myriad of possible parameterizations, but not without significant human direction. Although we have seen that reasonable interatomic potentials can be derived analytically from experimental or *ab initio* data, inversion schemes become most powerful when used as theoretical guidance for fitting, for two basic reasons. The first is that inversion necessarily involves a restricted set of *ab initio* data. While the input data can be perfectly reproduced (unless it is overdetermined), it is desirable to allow an imperfect description of the inversion data in order to achieve a better overall fit of a wider *ab initio* database that includes low symmetry defect structures. The second drawback of inversion is that the class of tractable functional forms is rather limited due to issues of invertability, numerical stability, and physical validity. With the fitting approach, although there is less connection with first principles, we can explore the possibility of functional forms of greater complexity and sophistication. On the other hand, complex fitting schemes are difficult to implement; large parameter sets make it hard to judge transferability; and cumbersome functional forms obscure principles of chemical bonding and reduce the ease of force evaluation. A better approach is to incorporate the theoretically derived features of the previous chapters directly into a functional form, and then to fit the potential to a carefully chosen *ab initio* database with a minimal number of parameters. In this way, a reliable potential for bulk properties can be derived systematically, while keeping the functional form simple enough to allow for efficient computation of forces as well as intuitive interpretation of chemical bonding.

The results of this chapter are the product of almost ten years of hard work by many people, including E. Kaxiras, J. F. Justo, V. V. Bulatov, S. Yip, S. Ismail-Beigi, E. Chung and K. C. Pandey. In this chapter the current state of our empirical model for Si is presented with emphasis on the author's contributions. In Section 5.1, the theoretical results of the previous chapters are incorporated into a general functional form for interatomic forces in bulk covalent solids, called the "Environment-Dependent Interatomic Potential" (EDIP) [114], and in Section 5.2 the fitting and testing of an EDIP for bulk silicon is described.

5.1 Functional Form

5.1.1 Scalar Environment Description

Approximation of quantum models (Chapter 2.2) suggests that dependence on the local atomic environment is required to attain a transferable description of chemical bonding. In the case of silicon, these results are supported and quantitatively extended by the inversion of cohesive energy curves (Chapter 4). The simplest description of the local environment of an atom is the number of nearest neighbors. Following Tersoff [37], let us define an effective coordination number Z_i for atom i,

$$Z_i = \sum_{m \neq i} f(R_{im}) \tag{5.1}$$

where $f(R_{im})$ is a cutoff function that measures the contribution of neighbor m to the coordination of i in terms of the bond length R_{im}. The special sp^2 and sp^3 bonding geometries can be uniquely specified by their coordinations due to their high symmetry. Since environment dependence is not needed in those cases, it is natural to take the coordination number to be a constant, except when large distortions from equilibrium occur. Moreover, covalent bonds tend to involve only first neighbors, as indicated by *ab initio* charge density calculations of open structures like the diamond lattice [118]. Thus, the neighbor function is chosen to be exactly unity for typical covalent bond lengths, $r < c$, with a gentle drop to zero above a cutoff b that excludes second neighbors,

$$f(r) = \begin{cases} 1 & \text{if } r < c \\ \exp\left(\frac{\alpha}{1-x^{-3}}\right) & \text{if } c < r < b \\ 0 & \text{if } r > b \end{cases} . \tag{5.2}$$

where $x = (r-c)/(b-c)$. This particular choice of cutoff function is appealing because it has two continuous derivatives at the inner cutoff c, and is perfectly smooth at the outer cutoff b. The cutoffs b and c are restricted to lie between first and second neighbors of both the hexagonal plane and diamond lattice in equilibrium, so that their coordinations are 3 and 4, respectively. The cutoff function obtained from the fitting described in the next section is shown in Fig. 5.6. Although the elastic constant analysis of Chapter 3 suggests that weak, *vector* environment-dependence is needed for even for bulk elasticity (for C_{44} only), we suppress environment dependence completely near equilibrium ($f(r) = 1$ for $r < b$) in order to achieve remarkable computational efficiency, as described in Chapter 6.

Our scalar description of the atomic environment is similar to Tersoff's, but there are notable differences. First, the perspective is that of the atom rather than the bond: With our potential, the preferences for special bond angles, bond strengths and angular forces are the same for all bonds involving a particular atom. This is in contrast to the Tersoff format [18, 37, 38, 44, 56] in which a mixed bond–atom perspective is adopted: the contribution of atom i to the strength of bond (ij) is affected by the "interference" of other bonds (ik) involving atom i. This model provides an intuitive explanation for trends in chemical reaction paths of molecules [117] and allows for both covalent and metallic bonds to be centered at the same atom, as observed, for example, in *ab initio* charge densities for the BCT5 lattice [118], which lies between the covalent diamond lattice and the metallic β-tin lattice. However, the analysis of elastic properties discussed earlier favors the present approach for environment dependence near the diamond lattice. Another important difference between our model and Tersoff's is the separation of angular dependence from the bond order. As we shall see, this allows us to control independently the preferences for bond strengths, bond angles, and angular forces in a way that the Tersoff potential cannot. By keeping the bond order simple, we can also directly use the important theoretical results that motivated the Tersoff potential in the first place.

5.1.2 Coordination-Dependent Chemical Bonding

Our potential consists of coordination-dependent two- and three-body interactions corresponding to the defining features of covalent materials: *pair bonding* and *angular forces.* The energy of a configuration $\{\vec{R}_i\}$ is a sum over single-atom energies, $E = \sum_i E_i$, each expressed as a sum of pair and three-body interactions

$$E_i = \sum_j V_2(R_{ij}, Z_i) + \sum_{jk} V_3(\vec{R}_{ij}, \vec{R}_{ik}, Z_i), \tag{5.3}$$

depending on the coordination Z_i of the central atom. The pair functional $V_2(R_{ij}, Z_i)$ represents the strength of bond (ij), while the three-body functional $V_3(\vec{R}_{ij}, \vec{R}_{ik}, Z_i)$ represents preferences for special bond angles, due to hybridization, as well as the angular forces that resist bending away from those angles. From our atomic perspective, the pair interaction is broken into a sum of contributions from each atom, and similarly the three-body interaction is broken into a sum over the three angles in each triangle of atoms. Note that due to the environment dependence, the contributions to the bond strength from each pair of atoms are not symmetric, $V_2(R_{ij}, Z_i) \neq V_2(R_{ji}, Z_j)$, if the coordinations differ.

The basic idea behind our model is that chemical bonding for an arbitrary configuration can be expressed as a simple, three-body cluster potential that adapts itself to the local atomic environment. For condensed bulk structures, the environment may be sufficiently well-described by the scalar effective coordination number. The same basic form may also be appropriate for surfaces and molecules, but a generalized, vector or matrix environment-description should be required to successfully adapt pair bonding and angular forces to highly asymmetric configurations.

Pair Bonding

As implied by the results of Chapter 4.1.9, we adopt the bond order form for the pair interaction. Drawing on the popularity of the SW potential, we use those functional forms for the attractive and repulsive interactions,

$$V_2(r, Z) = A\left[\left(\frac{B}{r}\right)^\rho - p(Z)\right] \exp\left(\frac{\sigma}{r-a}\right), \tag{5.4}$$

which go to zero at the cutoff $r = a$ with all derivatives continuous. This choice can reproduce the shapes of inverted pair potentials for silicon. Because we have constructed Z, and hence $p(Z)$, to be constant near the diamond lattice, our pair interaction reduces exactly to the SW form for configurations near equilibrium, thus allowing us to obtain excellent elastic properties as explained in Chapter 3. Making this choice of repulsive term with the parameters [115] obtained by the fitting to defect structures (Chapter 5.2), we can follow the procedure of Chapter 4.1 to extract the implied bond order $p(Z)$ from *ab initio* cohesive energy curves for the following crystal structures (with coordinations given in parentheses): graphitic (3), diamond (4), BC-8 (4), BCT-5 (5), β-tin (6), SC (6) and BCC (8). These structures span the full range from three- and four-fold coordinated covalent bonding in sp^2 and sp^3 arrangements, to overcoordinated atoms in metallic phases. The inverted *ab initio* bond order versus coordination is shown in Fig. 5.1, along with two additional data points. Since we have only first neighbor interactions in the diamond lattice, we can obtain another bond order for three-fold coordination from the *ab initio* formation energy (3.3 eV) for an unrelaxed vacancy. An additional data point for unit coordination comes from the experimental binding energy (3.24 eV) and bond length (2.246 Å) of the Si_2 molecule [119].

The bond order data has a clear shoulder at $Z = Z_0 = 4$ where the predicted transition from covalent to metallic bonding occurs. For overcoordinated atoms with $Z > Z_0$, the bond order approaches its rough asymptotic behavior, $p \propto Z^{-1/2}$, characteristic of metallic band structure. For coordinations $Z \leq Z_0$, the bond order departs from the $Z^{-1/2}$ divergence, due to the formation of a band gap in the LDOS associated with covalent bonds. A natural choice to capture this shape is a Gaussian,

$$p(Z) = \mathrm{e}^{-\beta Z^2}. \tag{5.5}$$

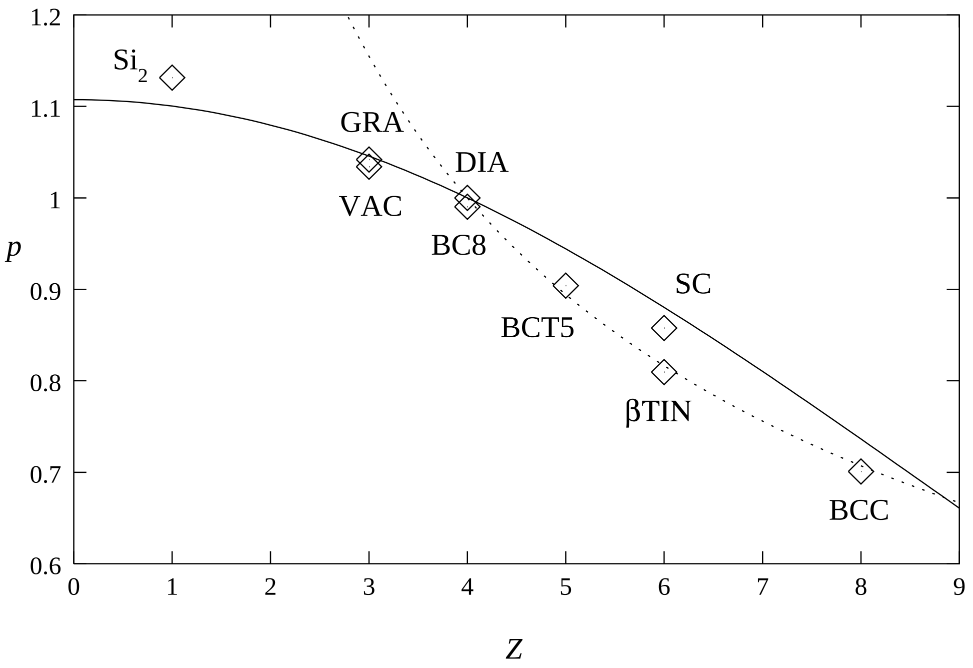


Figure 5.1: *Ab initio* values for the bond order in silicon as a function of coordination, obtained from the inversion of cohesive energy curves for the graphitic (GRA), cubic diamond (DIA), BC8, BCT5, SC, β-tin and BCC bulk structures and with additional points for the unrelaxed vacancy (VAC) and the dimer molecule (Si_2). For comparison the solid line shows the Gaussian $p(Z)$ obtained from fitting to defect structures. The dotted line shows the $1/\sqrt{Z}$ dependence, the theoretically predicted approximate behavior for $Z > 4$ (in the absence of angular forces).

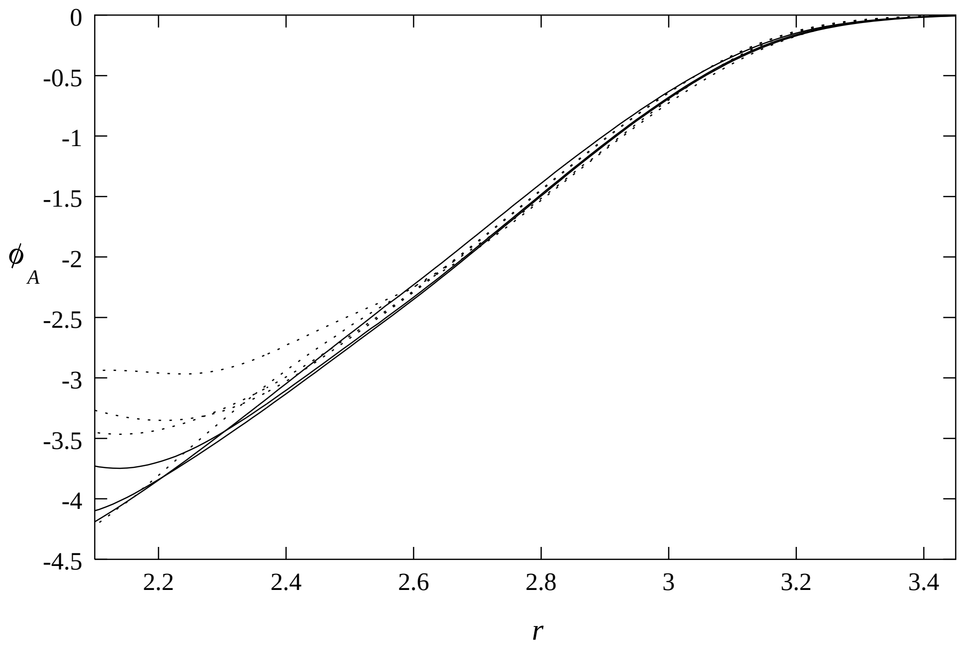


Figure 5.2: Attractive pair interactions from inversion of *ab initio* cohesive energy curves for the structures in Fig. 5.1 using the bond order and repulsive pair potential of our model. The solid lines are for the covalent structures with coordinations 3 and 4, while the dotted lines are for the overcoordinated metallic structures. The reasonable collapse of the attractive pair potentials indicates the validity of the bond order functional form of the pair interaction across a wide range of volumes and crystal structures.

In Fig. 5.1, we see that the bond order function we obtain from the fitting described in the next section is fairly close to the inversion data. It is intentionally somewhat too large for coordinations 5–8 to compensate for the small, but nonvanishing many-body energy for those structures, as explained below. The collapse of the attractive functions $\phi_A(r) = (V_2(r,Z) - V_A(r))/p(Z)$ with this choice of bond order shown in Fig. 5.2 is reasonably good, thus justifying the bond order formalism across a wide range of volumes. The deviation of the attractive functions for the metallic phases (weaker than the collapsed functions for the covalent phases) is consistent with our functional form: The pair attraction must be diminished to account for the nonvanishing (but small) many-body energy of metallic crystals. Our potential is the first to have a bond order in such close agreement with theory, which is a direct result of our novel treatment of angular forces.

Angular Forces

In a thorough comparative study of Si potentials, Balamane *et. al.* attribute the limitations of empirical models to the inadequate description of angular forces [19]. Our potential contains a number of innovations in handling angular forces, leading to a significant improvement over existing models in reproducing *ab initio* data. Analysis of elastic properties shows that, at least near equilibrium, the three-body functional should be expressed as a single, separable product of a radial function $g(r)$ for both bonds and an angular function $h(\theta, Z)$,

$$V_3(\vec{R}_{ij}, \vec{R}_{ik}, Z_i) = g(R_{ij})g(R_{ik})h(l_{ijk}, Z_i). \qquad (5.6)$$

Although the radial functions could vary with coordination, in the interest of simplicity we have focused on the angular function as the most important source of coordination dependence. Inversion of *ab initio* cohesive energy curves suggests that a consistent choice for the radial functions is the monotonic SW form,

$$g(r) = \exp\left(\frac{\gamma}{r-b}\right), \tag{5.7}$$

which also goes to zero smoothly at a cutoff distance b, a value that may be smaller than the two body cutoff a. Having separate cutoffs for two and three-body interactions is reasonable because they describe fundamentally different features of bonding. Although the pair interaction might extend considerably beyond the equilibrium first neighbor distance, the angular forces should not be allowed to extend beyond first neighbors, if they are to be interpreted as representing the resistance to bending of covalent bonds.

Much of the new physics contained in our potential comes from the angular function $h(l, Z)$. Theoretical considerations suggest the following general form:

$$h(l, Z) = H\left(\frac{l+\tau(Z)}{w(Z)}\right), \tag{5.8}$$

where $H(x)$, $w(Z)$ and $\tau(Z)$ are generic functions whose essential properties we now describe. The overall shape of the angular function is given by $H(x)$, a nonnegative [17, 66] function with a quadratic minimum of zero at the origin, $H(0) = H'(0) = 0$ and $H''(0) > 0$. The theoretical studies described in Chapter 2.2 suggest a polynomial form for $H(x)$ (expansions in $l = \cos\theta$), but the exact shape is a fundamental gap in our theoretical understanding, requiring additional research. A useful tool in this regard may be direct inversion for the angular function from shear strain energy data, described in Appendix B.

Motivated by theory, the function $\tau(Z)$ is chosen to control the coordination-dependent minimum of the angular function, $l_o(Z) = \cos(\theta_o(Z)) = -\tau(Z)$, with the following form[116][1],

$$\tau(Z) = u_1 + u_2(u_3 e^{-u_4 Z} - e^{-2u_4 Z}). \tag{5.9}$$

[1] Khor and Das Sarma also used a shifted equilibrium bond angle within the Tersoff format, but they did not specify exactly how the equilibrium angle should depend upon the local environment in a general configuration [45, 46].

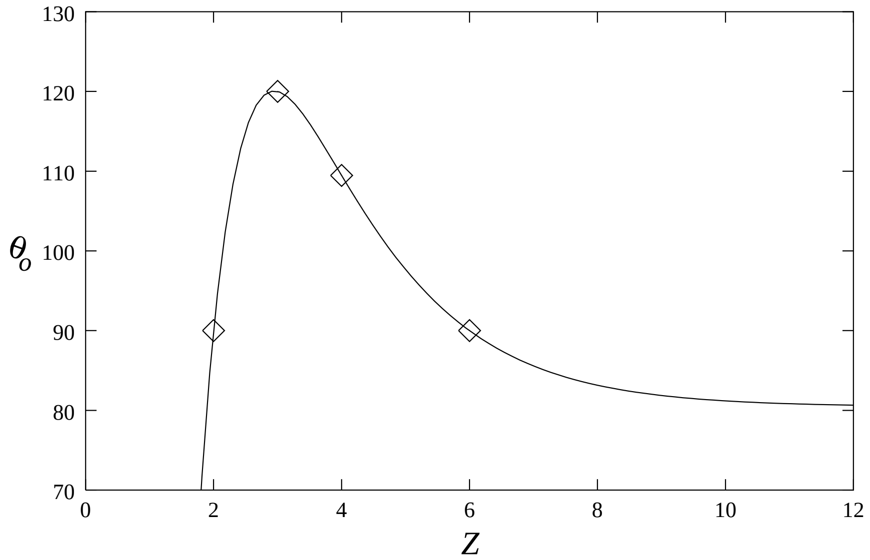


Figure 5.3: The coordination dependence of the preferred bond angle $\theta_o(Z)$ (in degrees), which interpolates the theoretically motivated points for $Z = 2, 3, 4, 6$, indicated by diamonds.

The parameters, $u_1 = -0.165799$, $u_2 = 32.557$, $u_3 = 0.286198$, and $u_4 = 0.66$, were chosen to make the preferred angle $\theta_o(Z) = \cos^{-1}[-\tau(Z)]$ interpolate smoothly between several theoretically motivated values, as shown in Fig. 5.3: We have already argued that $\tau(4) = 1/3$ and $\tau(3) = 1/2$ (so that sp^3 and sp^2 bonding correspond to the diamond and graphitic structures respectively), which determines two of the four parameters in $\tau(z)$. The remaining two parameters are selected so that $\tau(2) = \tau(6) = 0$ or $\theta_0(2) = \theta_0(6) = \pi/2$. For two-fold coordination, this choice reproduces the preference for bonding along two orthogonal p-states with the low energy, nonbonding s state fully occupied. For six-fold coordination, the choice $\theta_0(6) = \pi/2$ also reflects the p character of the bonds. However, structures with $Z = 6$ like SC and β-tin are metallic, with delocalized electrons that tend to invalidate the concept of bond-bending underlying the angular function, a crucial point we shall address shortly. The vanishing many-body energies for the graphitic plane and diamond structures allow fitting of the pair interactions $V_2(r,3)$ and $V_2(r,4)$ to be guided by Eq. (3.20), which determines $V_2''(r_d,4)$ from the bulk modulus, and Eq. (3.41), which requires $V_2''(r_h,3)/V_2''(r_d,4) \approx 1.4$. Moreover, the shifting of the minimum of the angular function in our model incorporates coordination-dependent hybridization in a way that other potentials cannot.

As an important aside, let us note that this choice of $\tau(Z)$ is not unique. It simply captures the essential feature of an environment-dependent minimum of the angular function. The BOP expansion suggests a similar (and simpler) function from Eq. (3.26): if we assume $p_\sigma = Z - 1$, then $\tau_\sigma(Z) = 1/(Z-1)$. The only major difference with the EDIP function is $\tau_\sigma(2) = 1$, indicating a favored angle of 180° for $Z = 2$, which also makes sense for sp-bonding. Otherwise, the EDIP function is in excellent agreement with the BOP expansion for $Z > 2$. Although we shall not pursue it here, another reasonable starting point to build a coordination-dependent, many-body potential for Si may be to let $p_\sigma = Z - 1$ in the BOP expansion.

Through the function $w(Z)$, the EDIP angular function has another novel coordination dependence to represent the covalent to metallic transition. The width of the minimum $w(Z)$ is broadened with increasing coordination, thus reducing the angular stiffness of the bonds as they become more metallic. Similarly, as coordination is decreased from 4 to 3, the width of the minimum is increased to reproduce the smaller angular stiffness of sp^2 bonds compared to that of sp^3 bonds. Thus, the function $w(Z)$ should have a minimum at $Z_0 = 4$ and diverge with increasing Z. Fitting of the model can be guided by Eq. (3.21), which determines $w(4)$ from the second shear modulus, and by Eq.(3.42), which requires $w(3)/w(4) \approx \sqrt{2}$. These features can be captured with the choice,

$$w(Z) = w_o Z^{-4\delta} e^{\delta Z}, \tag{5.10}$$

where $\delta \approx 2.3$. The softening of the angular function is important because it allows the decrease in cohesive energy per atom concomitant with overcoordination to be modeled by a weakening of pair interactions. In contrast, cluster potentials like SW and the inverted potential of Chapter 4.3 penalize overcoordinated structures with increased three-body energy that overcomes the decrease in pair bonding energy. This is an unphysical feature, since overcoordinated structures do not even have covalent bonds, and the many-body energy cannot be viewed as a consequence of stretching sp^3 bonds far from the tetrahedral geometry. In this sense, the reasonably good description of liquid Si (a metal with about 6 neighbors per atom) with the SW potential appears to be fortuitous. The large overestimation of the bulk modulus of β-tin (another six-fold coordinated metal) by SW and the inverted potential is another sign of the unphysically strong angular forces.

The coordination dependence of our angular function makes it possible for the first time to reproduce the well-known behavior of the bond order. The reason is that the contribution of the three-body functional to the total energy is suppressed for ideal crystals and overcoordinated structures. The shifting of the minimum makes the three-body energy vanish identically for sp^2 and sp^3 hybrids, and the variable width greatly reduces the three-body energy in metallic structures. With the three-body energy suppressed, we can use our knowledge of the bond order for the graphitic, diamond, β-tin and other lattices from inversion of cohesive energy curves to capture the energetics of these structures in the pair interaction, as described above. Several other potentials have tried to incorporate the bond order predicted from theory, but the uncontrolled many-body energy makes it impossible to connect directly with theory. Our treatment of angular forces is intuitively appealing because the forces primarily model the bending of covalent bonds, with the control of global energetics left to the pair interactions.

5.1.3 Discussion

In summary, recent theoretical innovations have been used to arrive at a functional form that describes the dependence of chemical bonding on the local coordination number. Bond order, hybridization, metalization and angular stiffness are all described in qualitative agreement with theory. Consistent with our motivation, we have kept the form as simple as possible, reproducing the essential physics with little more complexity than existing potentials. The fitted implementation of the model described in the next section involves only 13 adjustable parameters. We have theoretical estimates of almost half of the parameters, thus greatly narrowing the region of parameter space to be explored during fitting. The remaining parameters are chosen to fit important bulk defect structures.

Considering the theory behind our model, we can anticipate its range of applicability. We have shown that the structure and energetics of the diamond lattice can be almost perfectly reproduced. Because small distortions of sp^3 hybrids are accurately modeled, we would also expect a good description of the amorphous phase. Defect structures involving sp^2 hybridization should also be well described. In general, the model should perform best whenever the coordination number can adequately specify the local atomic environment. This certainly includes sp^2 and sp^3 hybridization and some metallic states, but might also include more general situations in which atoms are more or less symmetrically distributed, like the liquid and amorphous phases and reconstructed dislocation cores and grain boundaries. The theory behind the model begins to break down for noninteger coordinations, since our effective coordination number is only a way of smoothly interpolating between well-understood local structures. More seriously, no attempt is made to handle asymmetric distributions of neighbors, which are abundant in surfaces and small clusters. Theory suggests that our model may be fitted to provide a good description of condensed phases and defects in bulk tetrahedral semiconductors, such as Si, Ge and with minor extensions perhaps alloys such as SiGe, that can be understood in terms of simple principles of covalent bonding.

5.2 Fitting and Tests for Bulk Silicon

5.2.1 Fitting to Defect Structures

After choosing the functional form, the next important selection to be made in constructing an empirical potential is the set of experimental or *ab initio* data used to determine the adjustable parameters. Just as the functional form is designed to handle particular bonding arrangements, so too must the input data be carefully chosen to include representative properties within the target range of transferability of the potential. Experience shows that it is very difficult (if not impossible) to simultaneously fit to all the important classes of bonding states. Instead, the most successful approach is to focus on a single set of structures sharing similar bonding characteristics, and then to extend the range of transferability incrementally by adding new physics to the functional form and expanding the database. In the case of silicon, we have begun by working with bulk crystal and defects, with disordered phases as the next step (still in progress). These types of structures are within the theoretically predicted range of validity of the EDIP functional form, so we may hope to obtain a successful potential from a carefully controlled fitting strategy. Our first goal is a superior potential for simulations of condensed phases and bulk defects, such as defect diffusion, plastic deformation, radiation damage, amorphous vibrations, sintering, bulk melting and solid phase epitaxial growth. Ignoring other important structures like surfaces and small clusters is reasonable because a satisfactory degree of transferability for bulk defect structures not yet been attained by any model.

The *ab initio* fitting database used to construct the latest version of EDIP for silicon [115], compiled primarily by E. Kaxiras, includes: the diamond structure (cohesive energy and lattice constant), the experimental diamond elastic constants (C_{11}, C_{12} and C_{44})[78], formation energies of unrelaxed point defects (vacancy and tetrahedral and hexagonal interstitials)[122, 123, 124], the concerted exchange (CE) path for selfdiffusion (sampled at 10° intervals)[111], and selected points on the generalized stacking fault (GSF) energy surface [120, 121]. These target properties are used to optimize the 13 adjustable parameters: A, B, ρ, β, σ, a, b, c, λ, γ, α, Q_o and μ. (The last two parameters are defined below.) The fitting of the parameter vector $\{c_i\}$ to the input data $\{E_j\}$ (energies and elastic constants) is accomplished with a simulated annealing algorithm[109] that minimizes a least-squares cost function,

$$\Phi(\{c_i\}) = \sum_j \left(\frac{\tilde{E}_j - E_j}{\sigma_j} \right)^2 + P(\{c_i\}), \tag{5.11}$$

where $\tilde{E}_j$ is the value of property E_j predicted by the potential with the given parameter set, σ_j^{-1} are fitting weights (*e.g.* large for diamond cohesive energy and small for a particular CE data point), and $P(\{c_i\})$ is a nonnegative function that punishes unacceptable properties (like an energy lower than diamond or an incorrect equilibrium crystal volume). The art of fitting potentials lies in careful choice of all of these variables. Letting the annealing program run blindly always leads to frustration. Instead, one must regularly monitor the fitting routine and make adjustments to the weights, set of input properties, the penalty function, the random walk of the parameter vector, and the annealing schedule in response to the improvement or deterioration of the overall fit. Additional testing of the potential in parallel to the fitting is also required, for example, to search for spurious low-energy structures or expose trends behind poorly described structures (like large coordinations or small angles). A good fitted model is one which can survive a long barrage of attempts to invalidate it.

The parameters obtained from the latest simulated annealing fit, due to J. F. Justo in close collaboration with V. V. Bulatov, S. Yip, E. Kaxiras and the author, are given in Table 5.1. The fitted pair bonding function $V_2(r, Z)$ is plotted in Fig. 5.4 for several values of the coordination. Note the close similarity between $V_2(r, Z)$ and the inverted pair potentials for silicon shown in Fig 4.5, a built-in feature of the EDIP functional form.

The angular dependence $h(l, Z)$ is depicted in Fig. 5.5 through the three-body energy

Table 5.1: Values of the parameters that define the latest version of EDIP for bulk silicon.

A = 12.360638 eV	B = 1.6039258 Å	ρ = 1.3950202
a = 3.4557809 Å	b = 3.1640691 Å	c = 2.4504896 Å
σ = 1.3386900 Å	λ = 0.4610305 eV	γ = 0.2037403 Å
Q_o = 135.14236	μ = 0.7468472	β = 0.0063757
α = 4.0000000		

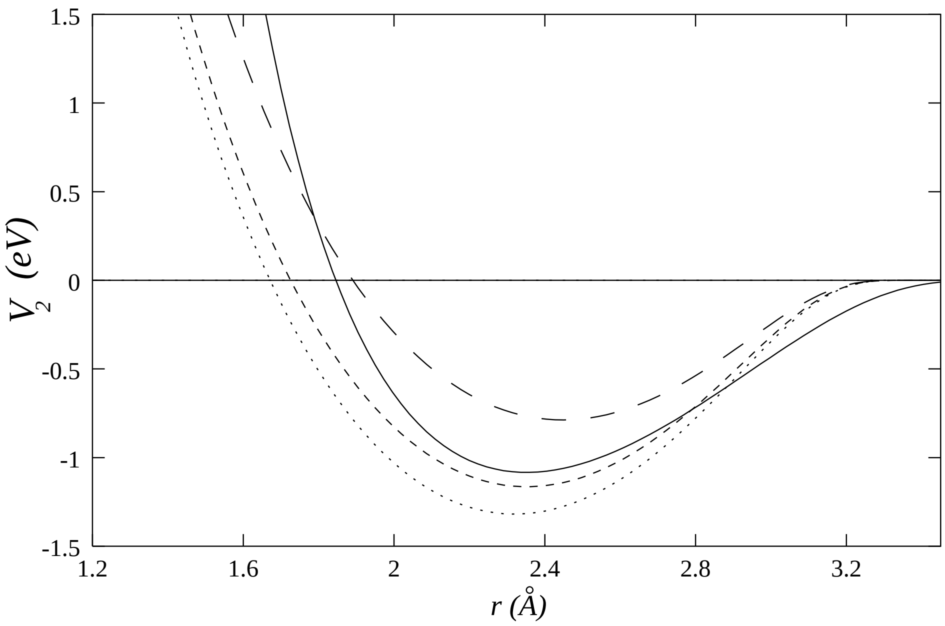


Figure 5.4: The two-body interaction $V_2(r, Z)$ as a function of separation r for different coordinations: $Z = 3$ (dotted line), $Z = 4$ (small-dash line), and $Z = 6$ (large-dash line), compared with the SW (solid line) pair interaction.

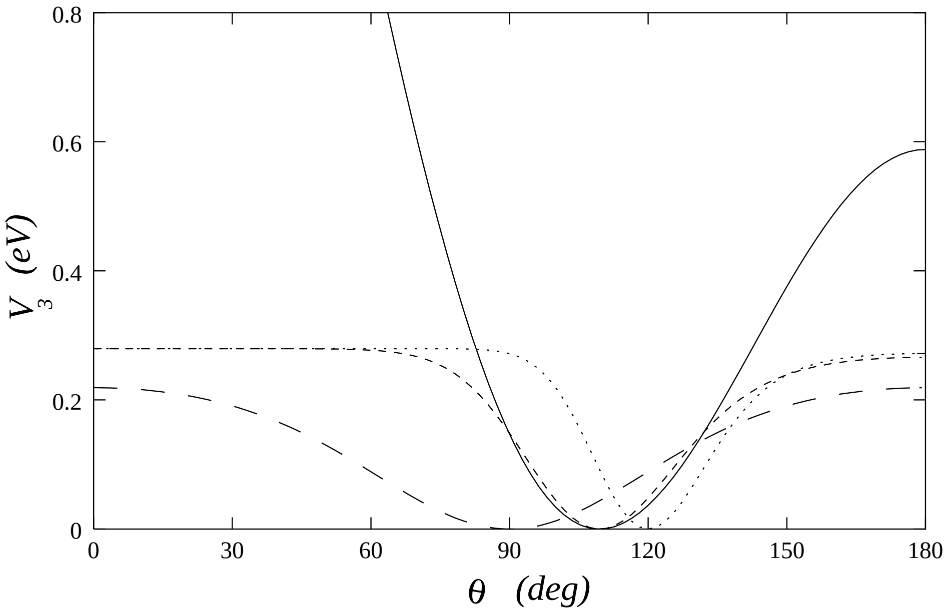


Figure 5.5: The three-body interaction $V_3(r, r, \cos\theta, Z)$ for a pair of bonds of fixed length r = 2.35 Å subtending an angle θ . The V_3 term is shown for several coordinations: $Z = 3$ (dotted line), $Z = 4$ (small-dash line), and $Z = 6$ (large-dash line), and compared with the SW (solid line) three-body interaction.

$V_3(r, r, l, Z)$ for a triplet of atoms with two bond lengths fixed at the equilibrium bond length $r = 2.35$ Å in differently coordinated environments. The shift of the minimum and the variable width are clearly seen, reflecting the physical trends we have already discussed. Note that for $Z = 12$ angular forces are almost completely suppressed, with $V_3 < 0.01$ eV for all angles. There is one inconsistency: the width of the angular function (greater angular forces) is smaller for $Z = 3$ than for $Z = 4$, the reverse of the theoretical prediction in Chapter 3. Unfortunately, the theoretical result came just after the last fit, but it may be incorporated into subsequent versions.

The particular choice of the shape of the angular function in the current version of EDIP for Si, due to J. F. Justo, resembles the angular dependence of the MFF potential [55]. An advantage of the MFF form over SW is its increased flexibility, containing two independent parameters instead of one. In order to make the similarity explicit, the MFF notation, $Q(Z) = w(Z)^{-1/2}$, is adopted,

$$h(l, Z) = \lambda \left[1 - \exp\left(-Q(Z)(l + \tau(Z))^2\right)\right]. \tag{5.12}$$

Although this inverted-Gaussian shape is identical to the MFF angular function, the EDIP dependence is much more sophisticated due to its environment dependence. The current choice for $Q(Z)$ is monotonically decreasing,

$$Q(Z) = Q_o \mathrm{e}^{-\mu Z}, \tag{5.13}$$

in contrast with the function postulated in Eq. (5.10), which requires that $Q(Z)$ have a maximum at $Z = 4$. Therefore, the functional form used in this version captures the covalent to metallic transition, but does not reproduce the relative bending strength of sp^2 and sp^3 hybrid covalent bonds. This may not be a serious problem, since the theoretical result comes from elastic moduli of the diamond and graphitic lattices, which is only relevant for angles near 109° and 120°, respectively. Nevertheless, the utility of the function in Eq. (5.10) is currently being explored, and, should it succeed, it might show that theory can steer the fitting process in a useful, but nonintuitive, direction.

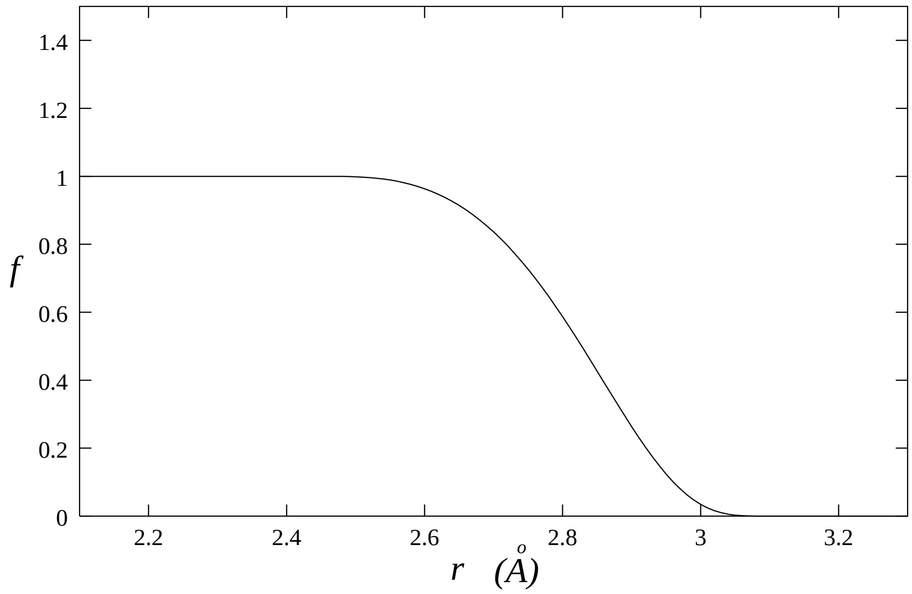


Figure 5.6: The function $f(r)$ that determines the contribution of each neighbor to the effective coordination Z.

The fitted neighbor function $f(r)$ is shown in Fig. 5.6. Note that the inner cutoff c, the largest radius for a full contribution to the coordination, is somewhat small to accurately count neighbors in closepacked structures. The coordinations measured for the *ab initio* equilibrium metallic crystals are: $Z = 5.863$ for β-tin (6), $Z = 5.994$ for SC (6), $Z = 7.801$ for BCC (8) and $Z = 11.319$ for FCC (12), where the actual coordinations are given in parentheses. Note, however, that the coordinations for all covalently bonded structures are correct, as well as for the mixed bonded BCT5. Thus, we have an accurate representation of the typical bulk atomic environments, placing the potential on firm theoretical ground.

5.2.2 Tests for Bulk Properties and Defects

Crystal Structure

The experimental binding energy, lattice constant and bulk modulus of the ground state diamond structure are perfectly reproduced by EDIP (as we enforce with large fitting

	EXPT	EDIP	SW	T3	TB
C_{11}	1.67	1.71	1.61	1.43	1.45
C_{12}	0.65	0.63	0.82	0.75	0.85
C_{44}	0.81	0.72	0.60	0.69	0.53
C_{44}^o	1.11	1.11	1.17	1.19	1.35
B	0.99	0.99	1.08	0.98	1.05
$C_{11} - C_{12}$	1.02	1.08	0.79	0.68	0.60
$C_{12} - C_{44}$	-0.16	-0.09	0.22	0.06	0.32

Table 5.2: Elastic constants of the diamond phase of silicon in Mbar, from experiment (EXPT) [78] (and first principles for C_{44}^o [79]) compared with the predictions of empirical potentials EDIP and SW (from the formulae of Cowley [71]) and T3 [19], as well as a tight binding model (TB) [80]. The dimensionless Kleinman internal strain parameter is also compared with experiment ([82, 83, 84].

weights). The cohesive energy curves of the low energy bulk phases are reasonably well described in the typical volume range for condensed phases of 15–20 Å^3, including some equilibrium volumes and energies (β-tin and BC8). There are also, however, some interesting inaccuracies in the cohesive energy curves, whose discussion we shall postpone until the next section.

Elastic Constants

The elastic constants of EDIP are given in Table 5.2 and compared with experimental and *ab initio* data, as well as the predictions of other empirical models. The overall agreement with experiment is excellent, a marked improvement over existing models, both empirical potentials and semi-empirical tight-binding Hamiltonians. This is not surprising, because, using the theoretical arguments of Chapter 3, we have built realistic elastic behavior directly into the EDIP functional form. The fitted EDIP potential reproduces the experimental values of C_{11} and C_{12} to within three percent, and therefore, as a consequence of the new elastic constant relation, Eq. (3.15), a good value of C^o_{44} is ensured. Indeed, the EDIP and *ab initio* C^o_{44} are identical. The second shear modulus $C_{11} - C_{12}$ is also very good, and the bulk modulus of EDIP is in perfect agreement with experiment. The value of C_{44} with internal relaxation is low, as it is with all empirical models, but the EDIP value is closer to experiment than the others, being only 11% too small. In contrast, with the most popular and successful version of the Tersoff potential (T2) the predicted value of C_{44} is 0.103 Mbar, almost an order of magnitude to small. The quality of elastic constants of the fitted EDIP is at the theoretical limit of its functional form, as calculated in Chapter 3. Also as predicted by that analysis, the Kleinman internal strain parameter is underestimated by EDIP, $\zeta = 0.494$, compared with the experimental, SW and T3 values, 0.74, 0.63 and 0.67, respectively. Finally, EDIP is the only transferable potential for silicon known to predict the negative Cauchy discrepancy, $C_{12} - C_{44}$, aside from the embedded-atom potential of Baskes was explicitly fit to it [48]. Even a number of semi-empirical tight-binding models cannot reproduce this important property[80], and those that can are not as accurate as EDIP [125].

In spite of its realistic elastic constants, however, EDIP does not provide a significant improvement over SW and T3 for phonon spectra[126]. The EDIP phonon frequencies are overestimated like those of the other potentials, especially along the zone boundary.

Point Defects

Point defect formation energies for vacancy (V), tetrahedral interstitial (I_T) and hexagonal interstitial (I_H), computed with EDIP are given in Table 5.3 and compared with LDA and other empirical models[2]. As an important example of an activated complex,

[2]The unrelaxed point defect energies in the fitting database, were computed using a plane wave basis with at least a 12 Ry cutoff and adequate sampling of the Brillouin zone for reciprocal space integrations.

		LDA	EDIP	SW	T3	TB
V	E_f	3.3-4.3	3.94	4.63	4.10	4.4
	ΔE_f	0.4-0.6	1.81	0.40	1.2	0.45
I_T	E_f	3.7-4.8	5.49	12.21	6.92	4.5
	ΔE_f	0.1-0.2	1.16	6.96	3.47	0.5
I_H	E_f	4.3-5.0	6.19	17.10	8.22	6.3
	ΔE_f	0.6-1.1	0.85	10.15	3.61	1.3
CE	E_f	5.5	6.55	7.90	6.50	5.5
	ΔE_f	0.9	2.51	3.26		1.8

Table 5.3: Ideal formation energies E_f^{ideal} of point defects (in eV) and relaxation energies $\Delta E = E_f^{ideal} - E_f^{relaxed}$ with EDIP using a 54 atom unit cell, compared with *ab initio* (LDA) [111, 122, 123, 124], SW and T3 [19, 42] and tight-binding (TB) [80] results.

energies for the saddle point of the concerted exchange path (CE) [111] are also given. The unrelaxed formation energies for all four defect configurations are included in the fitting database, and thus the EDIP values are fairly close to LDA. Note that the SW (and to a lesser degree T3) unrelaxed interstitial energies are much higher, indicating unphysical intolerance to overcoordinated structures with small angles, a point we have already noticed in the context of the inverted angular function of Chapter 4.3. Although the relaxed defect formation energies with SW and T3 are reasonable, they clearly fail in predicting the energy released upon relaxation from the ideal configuration. On the other hand, the EDIP relaxation energies are in fair agreement with *ab initio* calculations, in spite of only being fit to ideal structures. Note that EDIP predicts outward relaxation of the vacancy due to its first-neighbor, bond-order functional form, in analogy with the (111) 1×1 surface. Another important structure not included in the database is the split $< 111 >$ interstitial, whose (relaxed) formation energy with EDIP is 2.95 eV, compared with 4.68 eV for SW and 3.30 eV from *ab initio* calculations[127]. In agreement with *ab initio* results, EDIP predicts the split interstitial to be lowest energy interstitial configuration, while SW does not.

Finally, let us consider the concerted exchange path in more detail. Although the saddle point energies might suggest that EDIP, SW and T3 are equally good for this complicated and important path of local deformations, that is not the case. As shown in Fig. 4.13, EDIP outperforms SW, T3 and the inverted potential in describing the path overall (but keep in mind that only EDIP was fit to this data). EDIP still predicts unphysical minimum around $55°$, but it is not as pronounced as with the inverted potential. SW predicts a somewhat smaller metastable minimum at $65°$, but the saddle point energy is overestimated. T3 fails more seriously than any of the other models,

Supercells for the point defect calculations included 53-55 atoms, so that long range elastic relaxation energies, on the order of 0.01 eV, are ignored. For example, the formation energies for the unrelaxed defects used in the fitting are: 3.3 eV for the vacancy, 3.7 eV and 4.3 eV for the tetrahedral and hexagonal interstitials, and 5.47 eV for the concerted exchange saddle point configuration.

because there is a minimum rather than a saddle point at 90°, as well as another deep, unphysical minimum at 45°.

Extended Defects

The only extended defects in the fitting database are generalized stacking faults (GSF) of the {111} glide plane[120, 121]. In particular, three points (including energy saddles) on the glide set and three on the shuffle set are included. Generalized stacking fault energy surfaces are important as *ab initio* atomistic input to lattice-continuum models like the Peierls-Nabarro theory of dislocations, and test the ability of the potential to handle bond rupture and formation. Cross sections of the {111} glide set GSF energy surface in the high symmetry $< 11\overline{2} >$ and $< 1\overline{1}0 >$ directions computed with EDIP (by J. F. Justo [126]), SW and first principles are shown in Fig. 5.7. EDIP provides an good description of the lowest energy $< 11\overline{2} >$ cut, which passes from an ideal lattice to a stable stacking fault. The path to the stable stacking fault is more faithfully reproduced by EDIP than SW (although some points were fit). In the $< 110 >$ cut, which connects equivalent ideal crystal structures in the direction of the Burger's vector for a full dislocation, EDIP also outperforms SW. Although SW better fits the energy of the saddle point, this energy is very large and hence is physically irrelevant. In the important part of the curve ($E < 5$ eV/Å^2), EDIP is very close to the *ab initio* data points, in spite of not being fit to them.

Note that all first neighbor potentials, like SW and EDIP, predict a vanishing stable stacking fault energy. Although this may compromise some situations, such as long-range partial dislocation interactions, it is not a serious problem since the *ab initio* stable stacking fault energy is quite small, 0.006 eV/Å^2 (0.15 eV per atom on the glide plane), and the accuracy of empirical potentials is rarely better than a few tenths of an eV per atom.

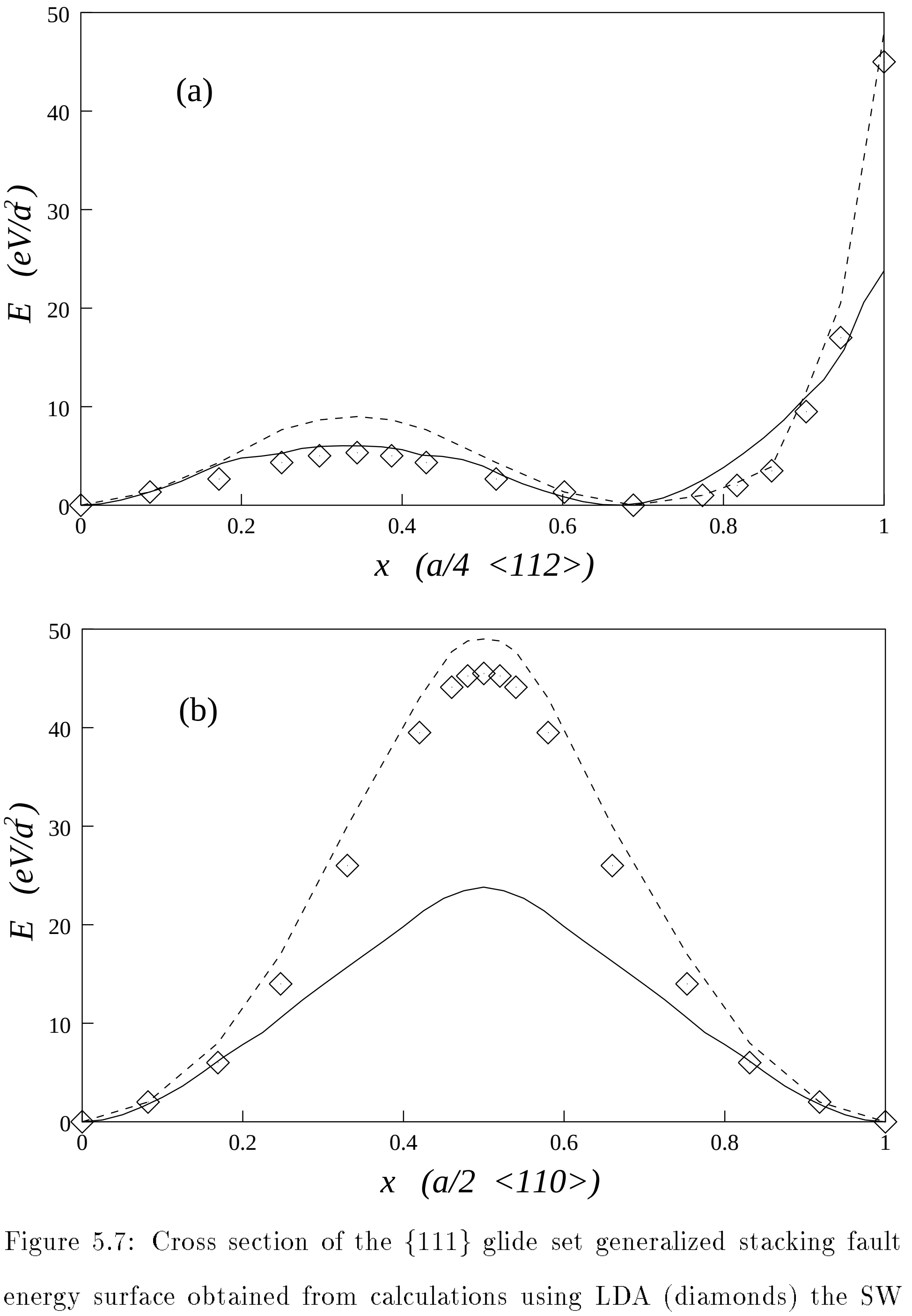


Figure 5.7: Cross section of the $\{111\}$ glide set generalized stacking fault energy surface obtained from calculations using LDA (diamonds) the SW (dashed line) and EDIP (solid line) along the (a) $<11\overline{2}>$ and (b) $<1\overline{1}0>$ directions.

Dislocation Cores

A much more stringent test of the transferability of EDIP comes from dislocation cores, about which no information is included in the fitting database. The lowest energy dislocations in tetravalent semiconductors like silicon are the 60° full and full screw in the $< 110 >$ / $\{111\}$ glide set slip system [128]. The 60° full lowers its energy by dissociating into 30° and 90° partial dislocations separated by a stable stacking fault, and the screw dissociates into two oppositely oriented 30° partials. It is the mobility of these partial dislocations that is believed to control plastic deformation in silicon. Due to the long-range $(1/r)$ stress fields around dislocations, any calculation of dislocation core energies or dynamics must involve a large number of atoms. Thus, partial dislocation cores in silicon provide an important materials application that is well-suited for theoretical study with empirical potentials, as long as sufficient transferability can be demonstrated to build confidence in their predictions. Unfortunately, no previous empirical potential is known to describe partial dislocations in silicon, even qualitatively.

In contrast, recent calculations with EDIP (by J. F. Justo in collaboration with V. V. Bulatov and S. Yip) have shown remarkable transferability for dislocations, making realistic simulations and energy studies possible. Here we shall only consider the highlights of core reconstructions; for a complete treatment with EDIP, including anti-phase defects, kink complexes and crossed dislocations, the reader is referred to Ref. [126]. First let us consider the 90°-partial dislocation core, which forms a simple asymmetric reconstruction by connecting dangling bonds across the core, thus lowering the energy by 0.87 eV/$\mathcal{B}$ according to first principles calculations [129], where $\mathcal{B}$ is the periodicity along the dislocation line. The SW potential cannot handle the core of the 90°-partial at all, because it does not support any reconstruction, *i.e.* the ideal, unreconstructed configuration is the stable minimum of energy [130]. The T3 potential does support reconstruction, but the predicted energy gain of 0.37 eV/$\mathcal{B}$ is too small by more than a factor of two [130]. In contrast, EDIP not only predicts the proper reconstruction, but also nearly the correct energy, 0.80 eV/$\mathcal{B}$.

Now let us turn to the 30°-partial dislocation core, which reconstructs by forming dimers to eliminate dangling bonds, much like the (100) surface. The *ab initio* reconstruction energy is 0.43 eV/$\mathcal{B}$ [131]. In this case, the SW potential predicts the correct reconstruction, but the energy 0.84 eV/$\mathcal{B}$ is too large by almost a factor of two. The T3 potential fails to describe the 30°-partial even qualitatively, because the reconstruction energy is negative[3]. Once again, in this case, EDIP not only predicts the reconstruction, but also reproduces the energy. In fact, the EDIP reconstruction energy of 0.45 eV/$\mathcal{B}$ is remarkably (and perhaps fortuitously) close to the *ab initio* value.

In summary, EDIP has attained an unprecedented degree of transferability for the silicon bulk crystal and defects. It is the first potential capable of a full description of partial dislocations with quantitative accuracy that appears to surpass even semi-empirical tight-binding models. The excellent elastic constants ensure an accurate treatment of long-range forces, and the core reconstructions and core defects are also well-described, in spite of not being explicitly fit.

Amidst these successes, however, there are hints of problems. For example, partially reconstructed metastable energy minima are predicted by EDIP along the reaction path connecting the ideal and reconstructed cores of both the 30°- and 90°-partial dislocations [126]. These minima are unphysical, but they will not affect transition rates or stability of the reconstructions during simulations. Similar oscillations in energy have also recently been observed for large shear distortions by E. Tadmor. These kinds of artificial energy changes occur when neighboring atoms pass through the cutoffs a and b, thus changing the local environment and making subtle contributions to the energy before they come close enough to form strong covalent bonds, which we have seen are quite well described by EDIP. In the setting of bulk defects, the density remains fairly con-

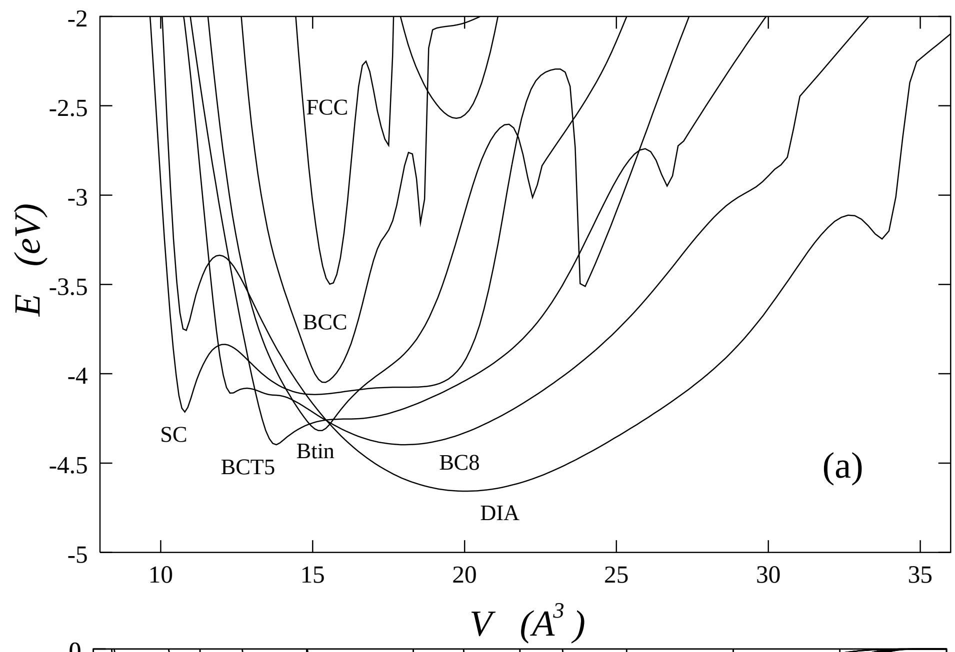
[3]This means that a metastable energy minimum appears at the reconstructed configuration, but it is higher in energy than the ideal structure, and hence is not favored.

stant, and typically only a small fraction of bond lengths fall near the cutoff distances during a simulation. In other cases of interest where volume changes are important, however, like melting or pressure-induced phase transitions, we may anticipate more serious problems.

5.2.3 Cohesive Energy Curves

We have paid special attention to cohesive energy curves for bulk crystal phases in building the theoretical foundation for EDIP, so it is natural to ask how the fitted potential performs. Cohesive energy curves computed with EDIP for the crystal structures we considered in Chapter 4 are shown in Fig. 5.8 (a)[4]. Unfortunately, the curves display large oscillations and bear little resemblance (across the entire volume range) to the *ab initio* curves of Fig. 4.1. Not everything is wrong: the diamond and BC8 curves are in close agreement with the *ab initio* curves across a broad range of volumes near their minima, indicating an accurate description of the lowest energy sp^3 bonding state. For volumes close to the minima of the corresponding *ab initio* curves, the EDIP curves are also not unreasonable. The obviously unphysical features lie away from those volumes, on both sides. At large volumes, every curve, even diamond, has a strange wiggle. In Fig. 5.8 (b), we see that those oscillations lie in the range where first neighbors contribute only partially to the EDIP coordination, $0 < f(r) < 1$. Another wiggle appears at smaller volumes roughly where second neighbors start to change the coordination. These volumes are not within the theoretical regime of validity for EDIP, which requires a correct determination of the coordination at the bulk density. So, these oscillations are not problematic for most bulk simulations, but there are other problems which are more serious. If the curves are recomputed with Z fixed at the value appropriate for each lattice, as shown in Fig. 5.9, then we see that the oscillations at high density are

[4]Note that these curves are constructed by dilating the *ab initio* structures. Allowing internal relaxation with EDIP might bring down some energies, but the general trends and oscillations should be insensitive to relaxation.



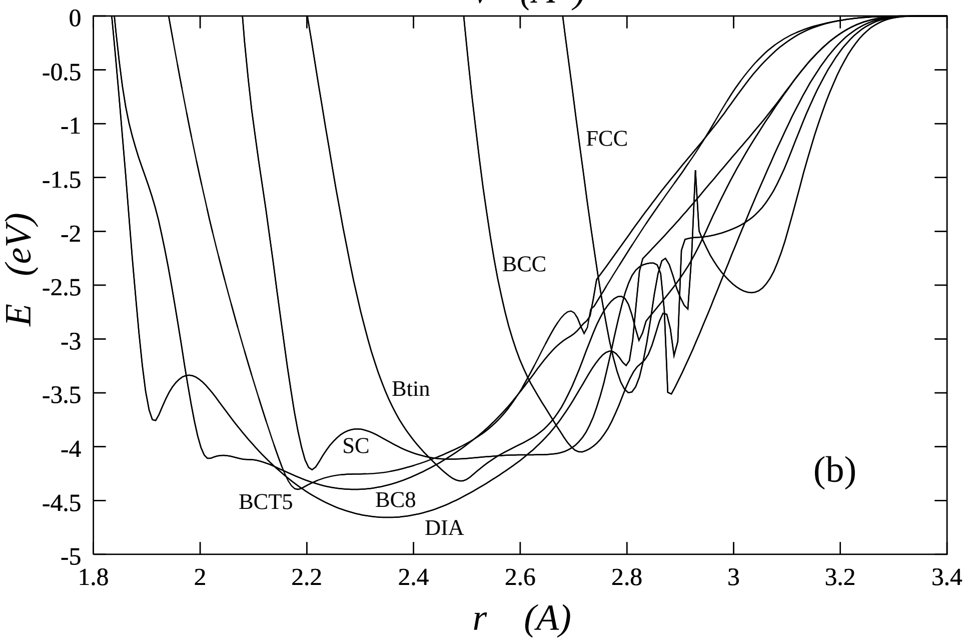


Figure 5.8: Cohesive energy curves versus volume (a) and first neighbor distance (b) computed with EDIP for various silicon crystal structures.

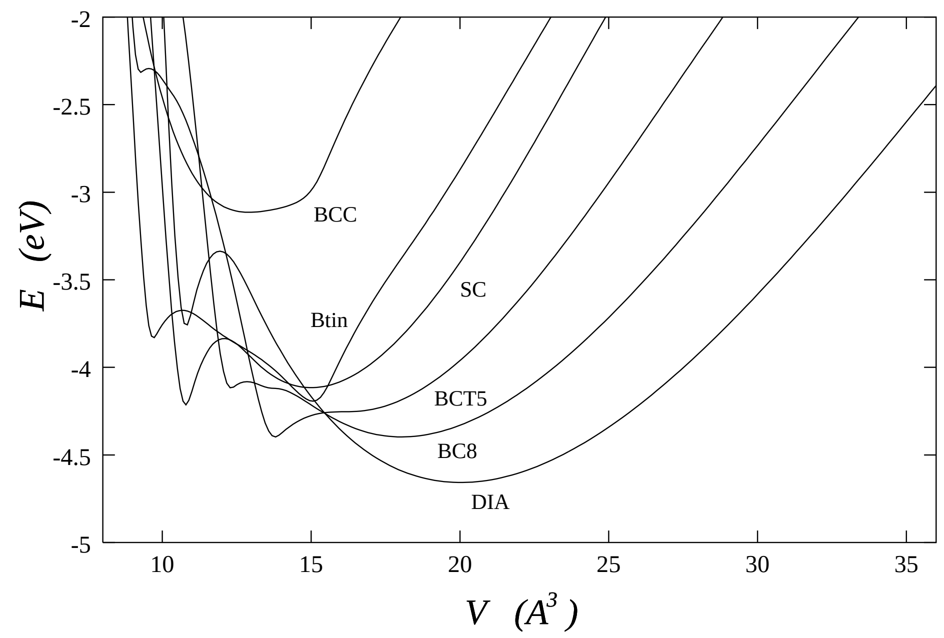


Figure 5.9: Cohesive energy curves computed with EDIP, with the coordination number artificially fixed at the correct equilibrium value for each crystal.

not removed, and thus must also have another cause.

The oscillations at small volumes are important to understand because they are the source of the most unphysical bulk properties. For example, the dips in the BCT5 and SC curves of Fig. 5.8 (a) influence pressure-induced phase transitions. With some correction possible for internal relaxation, the first high pressure phase transition of EDIP is from diamond to one of these structures, probably BCT5, and not to the experimentally observed β-tin. The primary reason for these dips is not changing coordination due to second neighbors, which occurs at a slightly smaller volume when second neighbors have $r < 3.00$ Å, as seen in Fig. 5.6 (and actually helps raise the energy by weakening pair bonds). Instead, the unphysical decreases in energy for small volumes occur where second neighbors fall into the range $3.05 < r < 3.20$ Å. This is the distance where the potential permits two-body attractions but not three-body repulsions. This phenomenon

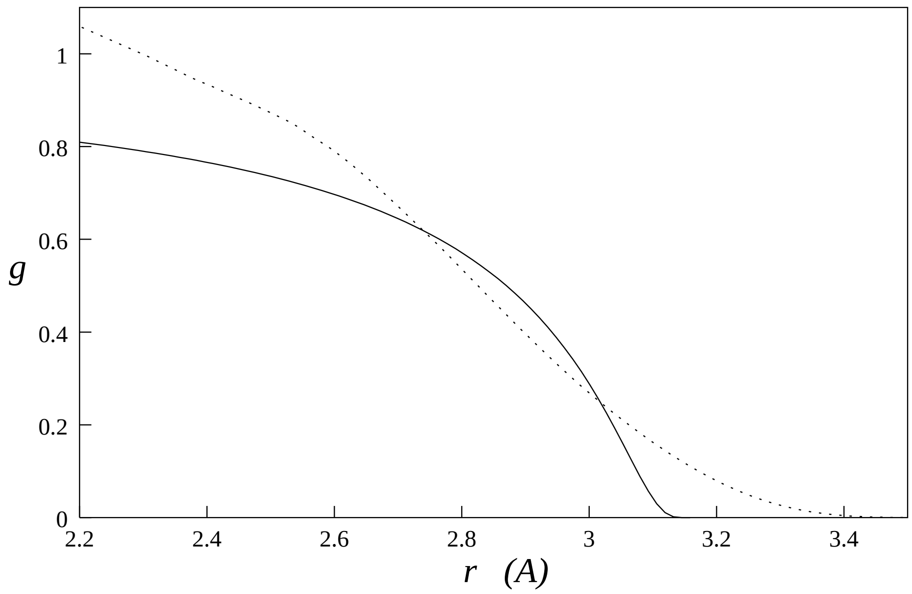


Figure 5.10: Three-body radial functions $g(r)$ for silicon from (a) the fitting of EDIP and (b) the inversion of the cohesive energy curves.

is a direct consequence of our allowing the three-body cutoff b to drift to a smaller value than the pair cutoff a in the fitting process. In general, there is an artificial and sharp decrease in the cohesive energy every time another shell of neighbors passes into the intermediate range between pair and three-body interactions. We shall see later that this is a serious problem for molecular dynamics simulations of disordered phases, but it should be possible to rectify it by slightly increasing the three-body cutoff. Refitting with $b = a$ for $g(r)$ (with coordination neighbor function the same, using the original $b < a$) is currently being explored.

The main factor leading to the sharpness of these and the coordination-induced oscillations is the sudden rise in the fitted $g(r)$ at the cutoff b. Our experience with three-body inversion clearly dictates that $g(r)$ must rise very gently near the cutoff. Fig. 5.10 compares the fitted $g(r)$ with the inverted function of Fig. 4.11, showing the relative abruptness of the fitted function. Requiring a larger value of γ (say, $\gamma > 0.5$) during fitting should remove many of these problems while preserving energies of covalently bonded structures.

5.2.4 Discussion

In summary, by fitting the EDIP functional form to a carefully chosen database of bulk crystal properties and defect formation energies, a superior potential has been produced, with a well-defined range of applicability. The potential performs remarkably well for bulk material near the equilibrium density where most bonds are covalent sp^3 hybrids. This includes isolated defects embedded in a nearly perfect diamond crystal with local relaxation and reconstruction. If the volume is increased so that bulk chemical bonds are broken, then the theory behind the functional form breaks down, and, understandably, the potential predicts unrealistic forces. At smaller volumes typical of metallic bonding, we have also observed unphysical effects, but theory suggests that the EDIP functional form may be refit to predict correct transitions and crystal stability of metallic phases. There are always going to be problems associated with neighbors passing through the cutoffs and contributing only partially to the coordination, but the hope is that they may not influence the most important structures, namely those corresponding to energy minima and saddle points, which tend to involve strong bonds in bulk defects. The only really serious problems we have found, after unusually comprehensive testing for bulk crystals and defects, are the low volume oscillations in cohesive energy curves, but these have been linked to the shorter range of many-body forces compared to pair interactions. It is possible that with minor refitting a transferable EDIP potential for bulk phases and defects can be produced, which would be a major advance over existing empirical models for silicon. In the meantime, it is interesting to ask how the current version of EDIP performs for disordered phases, like the liquid and amorphous states.