Chapter 2

Models of Interatomic Forces in Covalent Solids

> Only quantum mechanics can account for the covalent bond.
>
> – Andre Guinier [23]

In recent years, many empirical potentials for Si have been developed and applied to a number of different systems, and more recently compared to each other [19, 24]. Some of these models have been extended to other covalent materials, like Ge [136, 25], C [26, 27], F [28], S [29], SiGe [30, 31, 32], SiC [31], SiF [33], SiO_2 [34] and $GeSe_2$ [35], but by far the most testing of potentials has occured for Si, making it the ideal candidate for theoretical study into the fundamental issues of covalent bonding and representation by an empirical potential. Existing models differ in degree of sophistication, functional form, fitting strategy and range of interaction, and each can accurately model various special atomic configurations. Surfaces and small clusters are the most difficult to handle [19, 36], but even bulk material (crystalline and amorphous phases, solid defects and the liquid phase) has resisted a transferable description by a single potential. Realistic simulations of important bulk phenomena such as defect mobility, radiation damage, sintering, melting and crystallization are still problematic. In this chapter, we review existing potentials and approximations of quantum mechanical models in order to reach important conclusions about the desirable features of a successful interatomic potential.

2.1 Review of Empirical Potentials

The usual approach for deriving empirical potentials is to guess a functional form, motivated by physical intuition, and then to adjust parameters to fit *ab initio* total energy data for various atomic structures. A covalent material presents a difficult challenge because complex quantum-mechanical effects such as chemical bond formation and rupture, hybridization, metalization, charge transfer and bond bending must be described by an effective interaction between atoms in which the electronic degrees of freedom have somehow been "integrated out" [17]. In the case of Si, the abundance of potentials in the literature illustrates the difficulty of the problem and lack of specific theoretical guidance. In spite of the wide range of functional forms and fitting strategies, all proposed models possess comparable (and insufficient) overall accuracy [19]. It has proven almost impossible to attribute the successes or failures of a potential to specific features of its functional form. Nevertheless, much can be learned from past experience, and it is clear that a well-chosen functional form is more useful than elaborate fitting strategies.

To appreciate this point we compare and contrast some representative potentials for silicon. The pioneering potential of Stillinger and Weber (SW) has only eight parameters and was fitted to a few experimental properties of solid (cubic diamond) and liquid silicon [13]. The model takes the form of a third order cluster potential [17] in which the total energy of an atomic configuration $\{\vec{R}_{ij}\}$ is expressed as a linear combination of two- and three-body terms,

$$E = \sum_{ij} V_2(R_{ij}) + \sum_{ijk} V_3(\vec{R}_{ij}, \vec{R}_{ik}), \tag{2.1}$$

where $\vec{R}_{ij} = \vec{R}_j - \vec{R}_i$, $R_{ij} = |\vec{R}_{ij}|$ and we use the convention that multiple summation is over all permutations of distinct indices. The range of the SW potential is just short of the second neighbor distance in the ground state diamond lattice, so the pair interaction $V_2(r)$ has a deep well at the first neighbor distance to represent the restoring force against stretching sp^3 hybrid covalent bonds. The three-body interaction is expressed as a separable product of radial functions $g(r)$ and an angular function $h(\theta)$

$$V_3(\vec{r}_1, \vec{r}_2) = g(r_1)g(r_2)h(l_{12}), \tag{2.2}$$

where $l_{12} = \cos\theta_{12} = \vec{r}_1 \cdot \vec{r}_2/(r_1 r_2)$. The angular function, $h(l) = (l + 1/3)^2$, has a minimum of zero at the tetrahedral angle to represent the angular preference of sp^3 bonds, and the radial function $g(r)$ decreases with distance to reduce this effect when bonds are stretched. The SW three-body term captures the directed nature of covalent sp^3 bonds in a simple way that selects the diamond lattice over close-packed structures. Although the various terms lose their physical significance for distortions of the diamond lattice large enough to destroy sp^3 hybridization, the SW potential seems to give a reasonable description of many experimentally relevant states, such as point defects, certain surface structures, and the liquid and amorphous phases [19]. The SW potential continues to be a favorite choice in the literature, due in large part to its appealing simplicity and apparent physical content.

Another popular and innovative empirical model is the Tersoff potential, with three versions generally called T1 [18], T2 [37], and T3 [38]. The original version T1 has only six adjustable parameters, fitted to a small database of bulk polytypes. Subsequent versions involve seven more parameters to improve elastic properties. The Tersoff functional form is fundamentally different from the SW form in that the strength of individual bonds is affected by the presence of surrounding atoms. Using Carlsson's terminology, the Tersoff potential is a third order cluster functional [17] with the cluster sums appearing in nonlinear combinations. As suggested by theoretical arguments [39, 16, 20], the energy is the sum of a repulsive pair interaction $\phi_R(r)$ and an attractive interaction $p(\zeta)\phi_A(r)$ that depends on the local bonding environment, which is characterized by a scalar quantity ζ,

$$E = \sum_{ij} [\phi_R(R_{ij}) + p(\zeta_{ij})\phi_A(R_{ij})] \tag{2.3}$$

$$\zeta_{ij} = \sum_{k} V_3(\vec{R}_{ij}, \vec{R}_{ik}), \tag{2.4}$$

where the function $p(\zeta)$ represents the Pauling bond order. The three-body interaction has the form of Eq. (2.2) with the important difference that the angular function, although still positive, may not have a minimum at the tetrahedral angle. The T1, T2 and T3 angular functions are qualitatively different, possessing minima at 180^o, 90^o and 126.745^o, respectively. The original versions cannot describe the liquid and amorphous states[1]. The Tersoff format has greater theoretical justification away from the diamond lattice than SW, but the three fitted versions do not outperform the SW potential overall, perhaps due to their handling of angular forces [19]. Nevertheless, the Tersoff potential (or rather, family of potentials) is another example of a successful model for bulk properties with a physically motivated functional form and simple fitting strategy.

The majority of Si empirical potentials fall into either the generic SW [41, 42, 43] or Tersoff [44, 45, 46, 47, 48, 49] formats just described[2], but there are notable exceptions that provide further insight into successful approaches for designing potentials. First, a number of potentials possess functional forms that have either limited validity or no physical motivation at all, suggesting that fitting without theoretical guidance is not the optimal approach. The Valence Force Field and related potentials [50, 51, 52,

[1]This can be improved by changing the cutoff distance [40], but an uncontrolled change of the cutoff affects every other property of the potential, thus creating a new (and untested) model. It is sometimes said that "the Tersoff potential" can describe many properties (elastic constants, solid defects, surfaces, liquid, amorphous, clusters, ...), when in fact it been necessary to refit and modify the model for each circumstance, always at the cost of other desirable properties. The original SW potential, on the other hand, has provided reasonable transferability without any modifications.

[2]Most potentials for other covalent solids also assume either the SW [136, 25, 28, 29, 33] or Tersoff [26, 31, 32] functional form.

53] (of which there are over 40 in the literature [52]) involve scalar products of the vectors connecting atomic positions, an approximation that is strictly valid only for small departures from equilibrium. Thus, extending these models to highly distorted bonding environments undermines their theoretical basis. The potential of Pearson *et. al.* [54], as the authors emphasize, is not physically motivated, but rather results from an exercise in fitting. Their use of Lennard-Jones two-body terms and Axilrod-Teller three-body terms, characteristic of Van der Waals forces, has no justification for covalent materials. The potential of Mistriotis, Flytzanis and Farantos (MFF) [55] is an interesting attempt to include four-body interactions. Although the importance of four-body terms is certainly worth exploring, the inclusion of a four-body term in a linear cluster expansion is not unique, and theoretical analysis tends to favor nonlinear functionals [17, 16, 20].

A natural strategy to improve on the SW and Tersoff models is to replace simple functional forms with more flexible ones and complement them with more elaborate fitting schemes. The Bolding and Andersen (BA) potential [56] generalizes the Tersoff format with up to five-body interactions and over 30 adjustable parameters fit to an unusually wide range of structures. Although it has not been thoroughly tested, the BA potential appears to describe simultaneously bulk phases, defects, surfaces and small clusters, a claim that no other potential can make [19]. However, its complexity makes it difficult to interpret physically, and since a large fitting database was used, it is unclear whether the potential can reliably describe structures to which it was not explicitly fit. In this vein, the spline-fitted potentials of the Force Matching Method [57, 58] represent the opposite extreme of the SW and Tersoff approaches: physical motivation is bypassed in favor of elaborate fitting. These potentials involve complex combinations of cubic splines, which have effectively hundreds of adjustable parameters, and the strategy of matching forces on all atoms in various defect structures is the most elaborate attempted thus far. Although the method may be worth pursuing as an alternative, it has not yet produced competitive potentials [59]. Moreover, even if a reliable potential could result from such fitting strategies, it would make it hard to interpret the results of atomistic simulations in terms of simple principles of chemical bonding. Such interpretation is essential if any physical insight is to be gained from computer simulations.

In spite of relentless efforts, no potential has demonstrated a transferable description of silicon in all its forms [19] leading us to another important conclusion: it may be too ambitious to attempt a simultaneous fit of all of the important atomic structures (bulk crystalline, amorphous and liquid phases, surfaces, and clusters) since qualitatively different aspects of bonding are at work in different types of structures. Theory and general experience suggest that the main ingredient needed to differentiate between surface and bulk bonding preferences is a more sophisticated description of the local atomic environment. A notable example in this respect is the innovative Thermodynamic Interatomic Force Field (TIFF) potential of Chelikowsky *et. al.* [22], which includes a quantity called the "dangling bond vector" that is a weighted average of the vectors pointing to the neighbors of an atom,

$$\vec{D}_i = -\sum_{j \neq i} \vec{R}_{ij} f(R_{ij}), \tag{2.5}$$

where $f(r)$ is a cutoff function. For symmetric configurations characteristic of the ideal (or slightly distorted) bulk material, the dangling bond vector vanishes (or is exceedingly small). Conversely, a nonzero value of the dangling bond vector indicates an asymmetric distribution of neighbors. While the dangling bond vector description appears to be very useful for undercoordinated structures like surfaces and small clusters, in this thesis our focus is on bulk material and thus we only consider simpler, scalar environment descriptions. We shall see in the next chapter that the dangling bond vector also improves accuracy in the bulk, but it is not the leading order environment variable. Our review of the current state-of-the-art of empirical potentials shows that a goal of fundamental importance is to obtain the best possible description of condensed phases and defects with a simple, theoretically justified functional form.

2.2 Quantum-Mechanical Approximations

An alternative to fitting guessed functional forms is to derive potentials by systematic approximation of quantum-mechanical models. So far, this approach has failed to produce superior potentials, but important connections between electronic structure and effective interatomic potentials have been revealed. Although attempts are being made to directly approximate Density Functional Theory [60], the most useful contributions involve approximating various Tight Binding (TB) models, which can themselves be derived as approximations of first principles theories [61]. These methods are based on low order moment approximations of the TB local density of states (LDOS), which is used to express the average band energy as the sum of occupied bonding states [63, 17, 62, 20, 67, 64, 65, 66, 68]. Pettifor has derived a many-body potential, similar in form to the Tersoff potential, by approximation of the TB bond order [20]. More recently, an angular dependence remarkably close to the T3 angular function has been derived for σ bonding from the lowest order two-site term in the Bond Order Potential (BOP) expansion [21, 67], but the analytically derived function has a flat minimum at 120^o and thus differs qualitatively with the T1 and T2 potentials (the latter being the most successful version overall). A simple physical principle explains these results: a σ bond is most weakened (desaturated) by the presence of an another atom when the resulting angle is small ($\theta < 100^o$) because in such cases the atom lies near the bond axis, thus interfering with the σ orbital where it is most concentrated. Working within the same framework of the TB LDOS, Carlsson and coworkers have derived potentials with the Generalized Embedded Atom Method [64, 65, 66]. Harrison has arrived at a similar model by expanding the average band energy in the ratio of the width of the bonding band to the bond-antibond splitting, the relevant small parameter in semiconductors [68]. These potentials resemble the SW potential in its description of angular forces with an additive three-body term, particularly for small distortions of the diamond lattice. The transition to metallic behavior in overcoordinated structures involves interbond interactions similar to the Tersoff and embedded atom potentials.

Many-body potentials can be derived from quantum-mechanical models if we restrict our attention to important small sets of configurations. Using a basis of sp^3 hybrid orbitals in a TB model, Carlsson *et. al.* [17, 64] have shown that a generalization of the SW format, in which Eq. (2.2) is replaced by a form similar to that used by Biswas and Hamann (BH) [41],

$$V_3(\vec{r}_1, \vec{r}_2) = \sum_{m=0}^{2} g_m(r_1) g_m(r_2)\, l_{12}^m, \tag{2.6}$$

is valid in the vicinity of the equilibrium diamond lattice. In general, the fourth moment controls the essential band gap of a semiconductor, implying four-body interactions, but the separable, three-body SW/BH terms are a consequence of the open topology of the diamond lattice: the only four-atom hopping circuit between first neighbors is the self-retracing path $i \to j \to i \to k \to i$ [17].

We can make analogous arguments for the graphitic lattice to draw conclusions about sp^2 hybrid bonds. Ignoring the weak, long-range interaction between hexagonal planes, we can assume a TB basis of sp^2 hybrid orbitals and follow Carlsson's derivation. Because the self-retracing path is also the only first neighbor hopping circuit in a graphitic plane, a cluster expansion with the generic BH three-body interaction is also valid for hexagonal configurations, with the functions in Eqs. (2.1) and (2.6) differing from their diamond sp^3 counterparts, as described below. These calculations also suggest that a locally valid cluster expansion should acquire strong environment dependence for large distortions from the reference configuration [17].

In summary, these studies provide theoretical evidence that the linear three-body SW/BH format is appropriate near equilibrium structures, while the nonlinear many-body Tersoff format describes general trends across different bulk structures. For the asymmetric configurations found in surfaces and small clusters, these theories also suggest that a more complicated environment dependence than Tersoff's is needed, like the dangling bond vector of the TIFF potential [20, 64]. Direct approximation of quantum models can provide insight into the origins of interatomic forces, but apparently cannot produce improved potentials. The reason may be that the long chain of approximations connecting first principles and empirical theories is uncontrolled, in the sense that there is no small parameter which can provide an asymptotic bound for the neglected terms for a wide range of configurations. The expansion parameters in these studies are the dimensionless ratios of high to low order moments of the TB LDOS, which may not be small, especially for defect structures with states in the band gap or for metallic states like the liquid, the β-tin crystal structure, and certain surfaces. Low order moments capture general trends in energy, but cannot be expected to maintain quantitative accuracy.

2.3 The Need for Theoretical Guidance

A wide range of ideas about interatomic forces have been advanced and tested. Some are suggested by approximations of quantum theories or by empirical trends in chemical data, but most merely reflect physical intuition. Improving upon current models remains a supremely frustrating proposition. Increased flexibility and sophistication in fitting does not seem to help; the most successful models tend to be the simplest. Still, one wonders if a computer might somehow be programmed to determine the functional form of interatomic forces with minimal human input.

This tantalizing possibility is currently being explored using genetic algorithms [69]. By randomly generating and exchanging functional elements (mathematical operations like addition and multiplication) and selecting the fittest individuals (assessed by predictions of important energies), a population of potentials evolves until an superior form emerges. Unfortunately, it seems that left to its own devices the computer cannot even find a reasonable pair potential. It turns out that "directed" genetic algorithms, which fill in portions of a human-engineered functional template, are required for the method to be successful.

Although the directed genetic algorithm approach may ultimately be fruitful, the focus must clearly be on the theoretical direction. The same may be said of the guess-and-fit approach reviewed earlier. In the next two chapters, we shall develop two methods for obtaining information about interatomic forces directly from experimental or *ab initio* data: analysis of elastic constants and inversion of cohesive energy curves. These powerful analytic techniques have roots in the literature of solid state physics preceding the recent flurry of activity in designing interatomic potentials. With some essential innovations, we shall see that these methods can unequivocally select amongst competing intuitive ideas and provide much needed constraints on the functional form of interactions in covalent solids.