Appendix B

Direct Inversion for Angular Forces

> The main reason behind the inability [of empirical potentials for silicon] to be more transferable is an inadequate description of angular forces.
>
> – H. Balamane, T. Halicioglu and W. A. Tiller [19]

Understanding the nature of angular forces in covalent solids has been a primary objective in this thesis. For small distortions from ideal configurations characterized by sp^3 and sp^2 hybrid covalent bonds, elastic constant analysis has been quite useful. For highly distorted and overcoordinated structures, inversion of cohesive energy curves is another fruitful method rooted in first principles. In the inversions presented in the main text, we have worked exclusively with cohesive energy versus volume curves. Although we have obtained angular functions by comparison of many such curves for different crystal structures, that approach only involves a small, discrete set of angles. Ideally, the angular function should somehow be extracted from shear strains or other reaction coordinates in which angles vary continuously with little change in volume.

In this appendix, we present mathematical proofs that this is at least a formal, if not also practical, possibility: within certain limitations, cohesive energy curves for angle-changing reaction coordinates can be inverted exactly to obtain the elusive angular dependence.

B.1 Formulation of the Problem

Our first task is to mathematically formulate the inversion problem in such a way that we might be able to solve it. We begin with the same assumptions about the functional form the interatomic potential as in Chapter 4, namely the separable, three-body cluster potential. This assumption, which precludes environment-dependence, shall be the main limitation on our results. As before we assume that the pair interaction is known (either by inversion or by assumption), so that the many-body portion of the energy, $F(\varepsilon)$, as a function of strain ε can be identified and expressed in terms of radial and angular functions,

$$F[g,h](\varepsilon) = \sum_i \sum_j \sum_{k>j} g(R_{ij}) g(R_{ik}) h(l_{ijk}), \tag{B.1}$$

where $l_{ijk} = \cos\theta_{ijk}$. A cohesive energy curve involves only one continuous degree of freedom, and hence is insufficient to determine more than one single-variable function in the interatomic potential. In Chapter 4, we use the idea of recursion to determine $g[h,F](r)$ with an assumed angular dependence in the case of uniform dilation. Here we show that the converse is also possible, to obtain $h[g,F](l)$ with an assumed radial function from shear strain data. In principle, a two-dimensional cohesive energy surface $F[g,h](\varepsilon_1,\varepsilon_2)$ could be inverted to simultaneously obtain both unknown functions $g[F](r)$ and $h[F](l)$, but since technical difficulties increase dramatically with the complexity of the attempted inversion, it is probably not worth pursuing such an ambitious course of action.

Let us distill the mathematics of Chapter 4 down to the central idea, which may be applied to the current problem: *recursively extend the solution from a point where the desired function is known to ones where it is not.* In the case of $g(r)$, we assume $\lim_{r\to\infty} g(r) = 0$, without loss of generality. This gives a starting point for recursive inversion, because in an infinitely expanded crystal, every bond is, of course, infinitely long. Inversion then proceeds by contracting the crystal to condensed volumes. In the case of $h(l)$, things are not so simple, but we can still make progress with reasonable assumptions.

We are fortunate that covalent structures have small coordinations, and hence far fewer angles than metallic structures. In the extreme cases of the diamond and graphitic structures, there is only one nearest neighbor bond angle. Elastic constant analysis shows that it is quite reasonable to assume $h(l) = 0$ in these cases and restrict the range of $g(r)$ to include only first neighbors, supplying two convenient starting points for inversion. Therefore, let $\varepsilon = 0$ correspond to the ideal diamond or graphitic lattice, and let l_o be the cosine of the first neighbor bond angle, $-1/3$ or $-1/2$, respectively. For finite strain ε, small enough not to alter the interaction topology (first neighbors), there will be a discrete set of M bond angle cosines. For each $l_m(\varepsilon)$, $m = 1, \ldots, M$, define the following quantities (analogous to α_{pq} in Section 4.2),

$$G_m(\varepsilon) = \sum_i \sum_j \sum_{k>j, l_{ijk}=l_m(\varepsilon)} g(R_{ij})g(R_{ik}), \tag{B.2}$$

where summation is over triplets forming the specified bond angle. The many-body energy then is expressed as,

$$F(\varepsilon) = \sum_{m=1}^{M} G_m(\varepsilon)h(l_m(\varepsilon)). \tag{B.3}$$

Note that it is not necessary for ε to be a homogeneous strain of the lattice for any of our arguments. In general, ε can represent any reaction coordinate that deforms the ideal crystal while maintaining invertability. Unfortunately, there is always a finite value of ε at which the inversion problem becomes singular.

B.2 The Constrained Case of an Even Angular Function

A complication is that a single strain will continuously sample the unknown function $h(l)$ in *two* directions away from l_o, corresponding to larger and smaller angles. (In the radial case, we start with infinite separation, and hence proceed in only one direction, namely smaller bond length.) As a first example, let us avoid this difficulty by artificially constraining the angular function to be even in the variable $\delta l = l - l_o$, *i.e.* $h(\delta l) = h(-\delta l)$. In that case the two independent directions collapse into one, and we have the recursion formula,

$$h(l_1(\varepsilon)) = \frac{1}{G_1}\left(F(\varepsilon) - \sum_{m=2}^{M} G_m(\varepsilon)h(l_m(\varepsilon))\right), \tag{B.4}$$

where the cosine deviations are numbered in order of decreasing magnitude, $|\delta l_1| > |\delta l_2| > \ldots > |\delta l_M|$ for each value of the strain. An explicit formula (without the unknown function on the right hand side) could in principle we obtained by recursive substitution, but in many cases it would quite cumbersome. Instead, we use a recursion procedure: start with $\varepsilon = 0$, where $h(l)$ is known, and solve for $h(\delta l_1(\varepsilon))$ in order of increasing ε.

The recursion formula above is easy to write down, but nontrivial to understand and apply, because issues of invertability and numerical stability are more subtle here than in the radial function case. In order for the inverse to exist, $l_1(\varepsilon)$ must be a strictly increasing function, so that at every stage $l_m(\varepsilon)$ is known for $m > 1$ from smaller values of ε. In general, it is not possible to find a reaction coordinate which opens the largest angle indefinitely, so there will be a singularity when $dl_1/d\varepsilon = 0$. The inversion may also become singular at a smaller value of ε if $l_1(\varepsilon)$ is discontinuous, which can occur when the interaction topology changes. For example, as ε is increased beyond a certain critical value ε_c, second neighbors may enter the interaction range (defined by the assumed function $g(r)$) and suddenly introduce new angles outside the range $0 \leq \delta l < \delta l_1(\varepsilon)$, where $h(l)$ is known from $\varepsilon < \varepsilon_c$. Specifically, invertability is destroyed if $\lim_{\varepsilon\to\varepsilon_c^+} l_2 > \lim_{\varepsilon\to\varepsilon_c^-} l_1$, because we are then left with one equation and two unknowns (l_1 and l_2) at $\varepsilon = \varepsilon_c$.

The main source of numerical instability here (analogous to closely spaced first neighbor shells in the radial function case) is $(\delta l_1 - \delta l_2)/\delta l_1 \ll 1$. It is possible for this quantity to vanish for certain reaction coordinates at a nonzero value δl, but usually not with homogeneous stains. The universal problem with this instability is in starting the inversion procedure. At first all angular deviations are zero, and hence the right hand side of Eq. B.4 vanishes. Thus, the starting point for the recursion, $\varepsilon = 0$, is in fact a singularity! The trick to avoid the singularity and eliminate the associated instability is to use the analytic results of Chapter 3. Using the elastic constant formulae, it is straightforward to derive the curvature of the inverted angular function near the minimum that is appropriate for a given initial strain. Therefore, we can start the recursion at a small, positive value of ε and assume a known, parabolic form for $h(l)$ near the minimum, which should not present any problems. (This is equivalent to starting the radial function inversion at a large but finite first neighbor distance with an assumed tail rather than at infinite separation, where the radial function is actually known to vanish.)

B.3 The General Case of a Skewed Angular Function

The preceding case serves as an illustration containing most of the technical difficulties of the general case and giving the basic idea. The only idea needed to remove the (surely unreasonable) assumption of an even angular function is to consider *two* reaction coordinates, ε_1 and ε_2, in order to determine the two branches of the angular function for $\delta l > 0$ and $\delta l < 0$. Of course, each reaction path involves angles from both branches, so it will be necessary to invert both strain energy curves, $F_1(\varepsilon_1)$ and $F_2(\varepsilon_2)$, simultaneously. For the inverse to exist, however, the two reaction paths must be carefully chosen. We now discuss two general classes of reaction path pairs that lead to (formally) tractable inversion schemes.

B.3.1 Volume-Scaled Reaction Pairs

Once one useful reaction path ε_1 has been chosen, a second path ensuring invertability of the pair can be obtained by scaling the volume of the first path. Thus, the configuration corresponding to ε_2 on the second reaction path, has exactly the same structure (mainly, the same angles) as $\varepsilon_1 = \varepsilon_2$ on the first path, but with a different overall volume (which scales every bond length equally). The ratio of the two volumes may vary with ε_1, but may also be fixed for the entire reaction path. For example, we could consider simple shears strains $\varepsilon_1 = \varepsilon_2 = \gamma_4$ of the diamond lattice at two different volumes, corresponding to unstrained first neighbor distances $r_1 = 2.35$ Å and $r_2 = 2.55$ Å.

With such a "volume-scaled reaction pair", the inversion for the angular function proceeds as follows. For a particular value of $\varepsilon_1 = \varepsilon_2 = \varepsilon$, there is a set of M angles (the same in each of the two structures) whose cosines we number in decreasing order, $l_1 > l_2 > \ldots > l_M$. At a given stage in the inversion, the unknown variables are $h(l_1(\varepsilon))$ and $h(l_M(\varepsilon))$, while $h(l)$ is known for the range in between, $l_M(\varepsilon) < l < l_1(\varepsilon)$. In general, compared to the ideal angle l_o, both the positive and negative branches of $h(l)$ will be determined, since $\delta l_1 > 0$ and $\delta l_M < 0$. As before, define the radial quantities $G_{nm}(\varepsilon)$ via Eq. (B.2), for the two strain paths $n = 1, 2$, which again differ only in the overall volume. The recursion formula is a 2×2 linear system in this case,

$$G_{n1} h_1 + G_{nM} h_M = F_n - \sum_{m=2}^{M-1} G_{nm} h_m, \quad n = 1, 2, \tag{B.5}$$

where we use the shorthand notation, $h_m = h(l_m(\varepsilon))$, $G_{nm} = G_{nm}(\varepsilon)$ and $F_n = F_n(\varepsilon)$. If the volume difference between the two reaction paths is small, the determinant of the coefficient matrix, $|\mathrm{G}| = G_{11}G_{2M} - G_{21}G_{1M}$, at leading order is a linear combination of first derivatives of $g(r)$. These quantities can be small, thus causing instability, $|\mathrm{G}| \approx 0$. However, if the assumed $g(r)$ is not too flat, it may be possible to control the instability by choosing a large enough volume ratio.

B.3.2 Opening-Closing Reaction Pairs

Another approach involves choosing a pair of structurally different reaction paths such that for every state $(\varepsilon_1, \varepsilon_2)$ the largest angle among both structures is in one while the smallest angle is in the other. In this case one reaction path determines the largest angle (greatest opening of the ideal angle) and the other determines the smallest (greatest closing ideal angle). An example of such an "opening-closing reaction pair" might be a uniaxial strain taken in opposite directions, $\varepsilon_2 \propto -\varepsilon_1$ (compression and tension). For an opening-closing pair, in some sense, the matrix G of the previous section becomes diagonal. Although opening-closing pairs should provide greater numerical stability than volume-scaled pairs, the former may be difficult to construct, because a crystal deformation that produces very large angles (openings) typically also produces very small angles (closings).

To design an inversion procedure, couple the two reaction coordinates with a single parameter, $\varepsilon_1 = \varepsilon$ and $\varepsilon_2 = \alpha(\varepsilon)\varepsilon$. In the simplest case, take $\alpha(\varepsilon) = 1$, but the coupling function can be tuned to extend invertability or control numerical instability. Next independently number the cosines in the two structures, $\{l_{1m}(\varepsilon)\}$ and $\{l_{2m}(\varepsilon)\}$, in decreasing order, $l_{n1} > l_{n2} > \ldots > l_{nM_n}$. In general, the number of angles in each structure may be different, $M_1 \neq M_2$. The opening-closing condition requires that the largest cosine overall is in the first structure, $l_{max}(\varepsilon) = l_{11}(\varepsilon)$, and the smallest is in the other, $l_{min}(\varepsilon) = l_{2M_2}(\varepsilon)$. With these definitions, the recursion formulae for the unknown function $h(l)$ are,

$$h_{max} = \frac{1}{G_{11}} \left(F_1 - \sum_{m=2}^{M_1} G_{1m} h_{1m} \right), \tag{B.6}$$

$$h_{min} = \frac{1}{G_{2M_2}} \left(F_2 - \sum_{m=1}^{M_2-1} G_{2m} h_{2m} \right). \tag{B.7}$$

In general, the inverse will exist as long as $l_{max}(\varepsilon)$ and $l_{min}(\varepsilon)$ are continuous monotonic functions (increasing and decreasing, respectively). As before, the procedure must be started with parabolic forms for $h_{max}(\varepsilon)$ and $h_{min}(\varepsilon)$, derived analytically (*e.g.* from elastic constant formulae).

B.4 Conclusion

One can formally invert (pairs of) cohesive energy curves for carefully chosen reaction coordinates to obtain angular functions, $h[g, F](l)$. The technical complications related to invertability and numerical stability are more subtle than in radial function case, so it is not clear that direct angular inversions will succeed in practice. If meaningful inversions are possible, one might envision iterating between radial and angular inversions or selecting an optimal radial function that causes collapse of inverted angular functions, by analogy to Section 4.3. However, such inversions are likely to become so complex that no advantage would be gained over the usual fitting approach.

A better use of these formulae would be in testing the assumptions of hypothetical models of bonding like EDIP directly against *ab initio* data. Although these inversions may be challenging to perform, the results are desperately needed. For example, by comparing inverted angular functions from strains of the graphitic and diamond lattices, we could test the EDIP conjecture that the shape of the angular function $H(x)$ is the same for $Z = 3$ and $Z = 4$. We could also see to what extent a single angular function can handle diverse reaction paths at fixed coordination. These assumptions are surely imperfect, but it is not clear *a priori* how bad they will turn out to be. Perhaps, as in the case of the Harrison model for unrelaxed elasticity, the EDIP model will be fortuitously good. Alternatively, the results may be so bad that different hypotheses are required. In any case, producing meaningful *ab initio* angular functions without any *ad hoc* fitting would be a welcome theoretical advance.