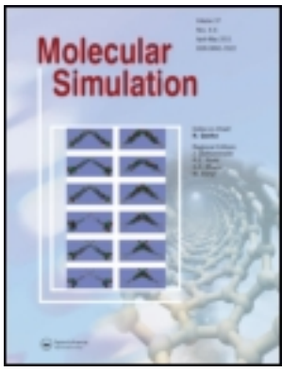
This article was downloaded by: [University of Saskatchewan Library]
On: 11 September 2012, At: 23:59
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gmos20

The General Utility Lattice Program (GULP)

Julian D. Gale & Andrew L. Rohl
[a] Department of Applied Chemistry, Nanochemistry Research Institute, Curtin University of Technology, P.O. Box U, 6845, Perth, WA

Version of record first published: 30 Sep 2011.

To cite this article: Julian D. Gale & Andrew L. Rohl (2003): The General Utility Lattice Program (GULP) , Molecular Simulation, 29:5, 291-341

To link to this article: http://dx.doi.org/10.1080/0892702031000104887

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Molecular Simulation, 2003 Vol. 29 (5), pp. 291–341

Taylor & Francis
Taylor & Francis Group

The General Utility Lattice Program (GULP)*

JULIAN D. GALE† and ANDREW L. ROHL

Department of Applied Chemistry, Nanochemistry Research Institute, Curtin University of Technology, P.O. Box U1987, Perth 6845, WA

(Received February 2003; In final form February 2003)

The General Utility Lattice Program (GULP) has been extended to include the ability to simulate polymers and surfaces, as well as adding many other new features, and the current status of the program is fully documented. Both the background theory is described, as well as providing a concise review of some of the previous applications in order to demonstrate the range of its use. Examples are presented of work performed using the new compatibilities of the software, including the calculation of Born effective charges, mechanical properties as a function of applied pressure, calculation of frequency-dependent dielectric data, surface reconstructions of calcite and the performance of a linear-scaling algorithm for bond-order potentials.

Keywords: GULP; Born effective charges; Linear-scaling algorithm; Bond-order potentials; Surfaces; Lattice dynamics; Forcefields

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

INTRODUCTION

The simulation of ionic materials has a long history going back over most of the last century. It began with lattice energy calculations based on experimental crystal structures through the use of Madelung's constant [1]. This was then expanded through the inclusion of short-range repulsive interactions, as found in the work of Born–Lande and Born–Mayer [2], in order that the crystal structure be a minimum with respect to isotropic expansion or compression. For many simple ionic materials a reasonable estimate of the lattice energy may even be obtained without knowledge of the structure, as demonstrated by the work of Kapustinskii [3]. Over the last few decades, atomistic simulation, in which we are only concerned with atoms, rather than electrons and sub-atomic particles, has developed significantly with the widespread use of computers. Correspondingly, the field has evolved from one that was initially concerned with reproducing experimental numbers, to one where predictions are being made, and insight is being offered.

The widespread use of atomistic simulation for solid state materials clearly resulted from the availability of computer programs for the task, just as much as the advent of the hardware to run the calculations. In the early days of solid state forcefield simulation for ionic materials, much of the work in the UK was centered around the Atomic Energy Authority at Harwell. Consequently, a number of computer codes arose from this work, such as HADES [4], MIDAS [5], PLUTO [6], METAPOCS and CASCADE [7]. Eventually, these migrated into the academic domain, leading to the THB suite of codes, including THBREL, THBFIT and THBPHON from Leslie. Further development of these programs led to the PARAPOCS code from Parker and Price [8], for free energy minimization of solids using numerical derivatives, and the DMAREL code from the group of Price for the simulation of molecular crystals through the use of distributed multipoles [9]. There were also several other prominent codes developed contemporaneously to the above family, in particular the WMIN code of Busing [10], the PCK series of programs from Williams [11], and the UNISOFT program of Eckold *et al.* [12]. While the codes mentioned above focus specifically on static lattice and quasi-harmonic

*Invited article
†Corresponding author.

ISSN 0892-7022 print/ISSN 1029-0435 online © 2003 Taylor & Francis Ltd
DOI: 10.1080/0892702031000104887

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

approaches to simulation, it should not be forgotten that there was a much larger, parallel, development of forcefield software for performing molecular dynamics simulation leading to programs such as GROMOS [13], AMBER [14], CHARMM [15] and DLPOLY [16,17], to name but a few.

This article focuses on the General Utility Lattice Program (GULP) which began development in the early 1990s, and is therefore subsequent to much of the aforementioned software, but implements many of the same ideas. However, there are increasingly many new, and unique, developments as well. The key philosophy was to try to bring together many of the facilities required for solid state simulation, with particular emphasis on static lattice/lattice dynamical methods, in a single package, and to try to make it as easy to use as possible. Of course, this is an aim and the degree of success depends on the perspective of the end user! It is important to also mention here the programs METADISE from Parker and co-workers [18], and SHELL from the group of Allan [19], which are also contemporary simulation codes sharing some of the same ideas.

In this work, the specific aim is to document the very latest version of GULP, 3.0, which includes many new features over previous versions, as well as having been considerably rewritten. Given that the previous release was version 1.3, the above use of 3.0 may come as a surprise. However, there was a development version of GULP that included periodic semi-empirical quantum mechanical methods from the MINDO and MNDO family, and this was nominally referred to as version 2.0 [20]. Hence, the extra increment is to avoid any possible confusion. Firstly, we detail the background theory to the underlying methods, some of which is not readily available in the literature. Secondly, we present a brief review of the utilization of the code to date, in order to highlight the scope of its previous application. Finally, we present some results illustrating the new capabilities of the latest version.

METHODS

The starting point for the majority simulation techniques is the calculation of the energy, and so it will be for this article. Most methods are based around the initial determination of the internal energy, with subsequent treatment of the nuclear degrees of freedom in order to determine the appropriate free energy to the ensemble of interest. In principle, the internal energy of a solid is a manybody quantity that explicitly depends upon the positions and momenta of all electrons and nuclei. However, this is an intractable problem to solve at any level of theory, and thus approximations must be made to simplify the situation. To tackle this we assume that the effect of the electrons will largely be subsumed into an effective atom, and that the energy can be decomposed into an expansion in terms of interactions between different subsets of the total number of atoms, N:

$$U = \sum_{i=1}^{N} U_i + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} U_{ij} + \frac{1}{6}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} U_{ijk} + \ldots,$$

where the first term represents the self energies of the atoms, the second the pairwise interaction, etc. This decomposition is exact if performed to a high enough order. However, we know that the contribution from higher order terms becomes progressively smaller for most systems, and so we choose to neglect the terms beyond a certain point and introduce a degree of parameterization of the remaining terms to compensate. Justification for this is forthcoming from quantum mechanics. It is well known that the Hartree–Fock method is a reasonable first approximation for the description of many systems, albeit with a systematic quantitative error for most observables. Here, the highest term included is a four-center integral, which indicates that including up to four-body terms should be a reasonable approach, which, for example, is indeed found to be the case for most organic systems. Furthermore, it is intuitively obvious that the further the two atoms are apart, the weaker will be their interaction. Thus, the introduction of distance cut-offs is a natural way to simplify the computational task.

The form of the explicit interaction between atoms is usually chosen based on physical insights as to the nature of the forces between the particles. For instance, if considering a covalent diatomic molecule the natural representation of the potential energy surface would be a Morse potential since this is harmonic at the minimum and leads to dissociation at large bond lengths, in accord with spectroscopic observation. In the following sections we will review some of the common types of potentials that are widely used, as well as some novel approaches which point towards the future of forcefield methods.

Coulomb Interaction

When considering ionic materials, the Coulomb interaction is by far the dominant term and can represent, typically, up to 90% of the total energy. Despite having the simplest form, just being given by Coulomb's law;

$$U_{ij}^{\text{Coulomb}} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

it is in fact the most complicated to evaluate for periodic systems (subsequently atomic units will be employed and the factor of $4\pi\epsilon_0$ will be omitted). This is because the Coulomb energy is given by a conditionally convergent series, i.e. the Coulomb energy is ill defined for an infinite three-dimensional (3-D) material unless certain additional conditions are specified. The reason for this can be readily understood—the interaction between ions decays as the inverse power of r, but the number of interacting ions increases with the surface area of a sphere, which is given by $4\pi r^2$. Hence, the energy density of interaction increases with distance, rather than decaying. One solution to the problem, proposed by Evjen [21], is to sum over charge-neutral groups of atoms. However, by far the most widely employed approach is the method of Ewald [22] for 3-D materials. Here the conditions of charge neutrality and zero dipole moment are imposed to yield a convergent series with a well-defined limit. To accelerate the evaluation, the Coulomb term is subjected to a Laplace transformation and then separated into two components, one of which is rapidly convergent in real space, and a second which decays quickly in reciprocal space. Conceptually, this approach can be viewed as adding and subtracting a Gaussian charge distribution centered about each ion [23]. The resulting expressions for real and reciprocal space, as well as the self-energy of the ion, are given below:

$$U^{\text{real}} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{r_{ij}} erfc\left(\eta^{\frac{1}{2}} r_{ij}\right)$$

$$U^{\text{recip}} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{G}\frac{4\pi}{V} q_i q_j \exp\left(iG\cdot r_{ij}\right)\frac{\exp\left(-\frac{G^2}{4\eta}\right)}{G^2}$$

$$U^{\text{self}} = -\sum_{i=1}^{N} q_i^2 \left(\frac{\eta}{\pi}\right)^{\frac{1}{2}}$$

$$U^{\text{electrostatic}} = U^{\text{real}} + U^{\text{recip}} + U^{\text{self}}$$

Here q is the charge on an ion, G is a reciprocal lattice vector (where the special case $G = 0$ is excluded), V is the volume of the unit cell and η is a parameter that controls the division of work between real and reciprocal space. It should also be noted that although the reciprocal space term is written as a two-body interaction over pairs of atoms, it can be rewritten as a single sum over ions for more efficient evaluation. The above still leaves open the choice of cut-off radii for real and reciprocal space. One approach to defining these in a consistent fashion is to minimize the total number of terms to be evaluated in both series for a given specified accuracy, A [24]. This leads to the following expressions:

$$\eta_{\text{opt}} = \left(\frac{Nw\pi^3}{V}\right)^{\frac{1}{3}}$$

$$r_{\max} = \left(\frac{-\ln(A)}{\eta}\right)^{\frac{1}{2}}$$

$$G_{\max} = 2\eta^{\frac{1}{2}}(-\ln(A))^{\frac{1}{2}}.$$

Note that the above expressions contain one difference from the original derivation, in that a weight parameter, w, has been included that represents the relative computational expense of calculating a term in real and reciprocal space. Tuning of this parameter can lead to significant benefits for large systems. There have been several modifications proposed for the basic Ewald summation that accelerate its evaluation for large systems, most notably the particle-mesh [25], and fast multipole methods [26,27]. Furthermore, there are competitive approaches that operate purely in real space for large unit cells that scale linearly with increasing size, such as the hierarchical fast multipole methods, though care must be taken to obtain the same limiting result by imposing the zero dipole requirement. This latter approach can also be applied to accelerating the calculation of the Coulomb energy of finite clusters.

In principle, it is possible to calculate the Coulomb energy of a system with a net dipole, μ, as well. The nature of the correction to the Ewald energy can be determined as a function of the shape of the crystal, and the formula below can also be employed [28]:

$$U^{\text{dipole}} = \frac{2\pi}{3V}\mu^2.$$

However, the complication of using the above correction is that it depends on the macroscopic dipole of the crystal, and even on any compensating electric field due to the environment surrounding the particle. Hence, the dipole moment is usually ill defined, since it depends on the surfaces, as well as the bulk material. Even if we neglect surface effects, the definition of the dipole moment is ambiguous since the operator is not invariant under translation of atomic images by a lattice vector. Consequently, we will take the Ewald result as being definitive.

Similarly, it is possible to relax the charge neutrality constraint, and to perform calculations on charged supercells [29], provided care is taken when constructing thermodynamic cycles. This is often used when probing defect energetics, an alternative to the Mott–Littleton method. Here the net charge, Q ($Q = \sum q_i$) is neutralized by

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

a uniform background charge, leading to an energy correction of:

$$U^{\text{background}} = -\frac{\pi}{2V\eta}Q^2.$$

So far we have only considered how to handle infinite 3-D solids, but the same issues exist for lower dimensionalities. Again for a two-dimensional (2-D) slab, the Coulomb sum is only conditionally convergent and so an analogous approach to the Ewald sum is usually taken, originally devised by Parry [30,31]. Here the slab, or surface, is taken so as to be oriented with the surface vectors in the *xy* plane and with the surface normal lying parallel to *z*. The energy contributions in real and reciprocal space are given by:

$$U^{\text{real}} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_iq_j}{r_{ij}}erfc\left(\eta^{\frac{1}{2}}r_{ij}\right)$$

$$U_1^{\text{recip}} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{G}\frac{\pi}{A}\frac{q_iq_j\exp(iG\cdot r_{ij})}{|G|} \times \left[\exp(|G|z_{ij})erfc\left(\frac{|G|}{2\eta^{\frac{1}{2}}}+\eta^{\frac{1}{2}}z_{ij}\right) + \exp(-|G|z_{ij})erfc\left(\frac{|G|}{2\eta^{\frac{1}{2}}}-\eta^{\frac{1}{2}}z_{ij}\right)\right]$$

$$U_2^{\text{recip}} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{2\pi q_iq_j}{A}\left[z_{ij}\cdot erf\left(\eta^{\frac{1}{2}}z_{ij}\right)+\frac{\exp\left(-\eta z_{ij}^2\right)}{(\pi\eta)^{\frac{1}{2}}}\right]$$

$$U^{\text{self}} = -\sum_{i=1}^{N}q_i^2\left(\frac{\eta}{\pi}\right)^{\frac{1}{2}}.$$

Note, there are now two terms in reciprocal space involving the 2-D reciprocal lattice vector, *G*, and here *A* is the surface area of the repeat unit, while z_{ij} is the component of the distance between two ions parallel to the surface normal. Again it is possible to relax the dipolar and charge neutrality constraints within the repeat directions. However, the approach to correcting the energy is far more uncertain, and it is necessary to make approximations [32]. As per the 3-D case, the optimum value of the convergence parameter can also be determined [33]:

$$\eta_{\text{opt}} = \frac{\pi w}{A}.$$

Recently, another approach has been proposed for the calculation of the 2-D Coulomb sum, which is reported to be faster than the Parry method for all cases, and especially beneficial as the number of atoms increases [34].

Lowering the dimensionality further, we arrive at one-dimensional (1-D) periodic systems which represents a single polymer strand, for example. At this dimensionality the Coulomb sum becomes absolutely convergent, though at a very slow rate when performed directly in real space. Summing over charge neutral units accelerates the process, though convergence is still somewhat tardy. While there have been several proposed approaches to accelerating the Coulomb sum, we find that the method proposed by Saunders *et al.* [35], in which a neutralizing background charge is applied, is effective. Here there are three contributions to the energy given by:

$$U_1^{\text{real}} = \frac{1}{2}\sum_{m=-M}^{+M}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_iq_j}{r_{ij}+ma}$$

$$U_2^{\text{real}} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_iq_j}{a} \times \left[\ln\left(\sqrt{(u+x)^2+y^2+z^2}+u+x\right) + \ln\left(\sqrt{(u-x)^2+y^2+z^2}+u-x\right)-2\ln a\right]$$

$$U_3^{\text{real}} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}q_iq_j[\xi(M,r_{ij})+\xi(M,-r_{ij})],$$

where the first term is summed over all images of *i* and *j* in the unit cells from $-M$ to *M*, *a* is the 1-D repeat parameter in the *x* direction and the remaining variables and functions are defined as below:

$$u = a\left(M+\frac{1}{2}\right)$$

$$\xi(M,r_{ij}) = -\sum_{i=1}^{M}E_ia^{2i-1}W_{2i-1}(u+x,y^2+z^2)$$

$$W_n(u+x,\alpha) = \left(\frac{\partial}{\partial u}\right)^n((u+x)^2+\alpha)^{-\frac{1}{2}}.$$

Although the required extent of the summations, given by *M*, to achieve a given precision is not known *a priori*, the method can be implemented in an iterative fashion so that the degree of convergence is tested as the number of lattice repeats is increased.

One alternative approach to the above methods for performing the Coulomb sum in any dimensionality is that due to Wolf *et al.* [36]. Their approach involves a purely real space summation that is asymptotic to the Ewald limit given a suitable choice of convergence parameters. It is based around the concept of ensuring that the sum of the charges of all ions within a spherical cut-off region is equal to zero and

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

that the potential goes smoothly to zero at that cut-off. This is achieved by placing an image of every ion that a given atom interacts with at the cut-off boundary, but in the diametrically opposite direction. While this approach can be applied directly to the Coulomb potential, convergence is slow. Better results are obtained by using a damped form of the potential, which is chosen to be equivalent to the real space component of the Ewald sum, with subtraction of the associated self-energy of the corresponding Gaussian. In this form, the expression for the energy is:

$$U^{\text{wolf}} = \frac{1}{2}\sum_i \sum_j q_i q_j \left(\frac{erfc(\alpha r_{ij})}{r_{ij}} - \lim_{r_{ij}\to r_{\text{cut}}} \left\{ \frac{erfc(\alpha r_{ij})}{r_{ij}} \right\} \right) - \sum_i q_i^2 \left\{ \frac{erfc(\alpha r_{\text{cut}})}{2r_{\text{cut}}} + \frac{\alpha}{\pi^{\frac{1}{2}}} \right\},$$

where α is the convergence parameter, closely related to the factor η in the Ewald sum and r_{cut} is the cut-off radius. There is a trade-off to be made in the choice of parameters within this sum. The smaller the value of α, the closer the converged value will be to the Ewald limit. However, the cut-off radius required to achieve convergence is also increased. This summation method has been implemented within GULP for 1-, 2- and 3-D calculations, though by default the term due to the limit of the distance approaching the cut-off is omitted from the derivatives in order to keep them analytically correct at the expense of the loss of smoothing.

Before leaving the topic of how Coulomb interactions are evaluated, it is important to note the special case of molecular mechanics forcefields. Here the Coulomb interaction, and usually the dispersion one too, is subtracted for interactions which are between neighbors (i.e. bonded or 1–2) and next nearest neighbors (i.e. which have a common bonded atom, 1–3) according to the connectivity. This is done so that the parameters in the two- and three-body potentials can be directly equated with experimentally observable quantities, such as force constants from spectroscopy. Furthermore, long-range interactions for atoms that are 1–4 connected are often scaled, usually by 1/2.

So far we have discussed the methods for evaluating Coulomb sums. However, we have yet to comment on how the atomic charges are determined. In most simulations, the charges on ions are fixed (i.e. independent of geometry) for simplicity and their magnitude is determined parametrically along with other forcefield parameters, or extracted from quantum mechanical information. In the later situation there is the question as to what is the appropriate charge to take since it depends on how the density matrix is partitioned. Although Mulliken analysis [37] is commonly used as a standard, this is not necessarily the optimum charge definition for use in a forcefield. Arguably, a better choice would be to employ the Born effective charges [38] that describe the response of the ions to an electric field, about which more will be said later. If the charges are purely regarded as parameters then the use of formal charges is convenient as it removes a degree of freedom and maximizes the transferability of the forcefield, especially to charged defects.

An alternative to using fixed charges is to allow the charges to be determined as a function of the geometry. It is well documented that this environment dependence of the charge is very important in some cases. For example, the binding energy of water in ice is much greater than that in the water dimer. This arises due to the increased ionicity of the O—H bond in the solid state and cannot be described correctly by a simple two-body model. In order to implement a variable charge scheme, a simple Hamiltonian is needed that is practical for forcefield simulations. Consequently, most approaches to geometry-dependent charges have been based around the concept of electronegativity equalization [39]. Here the energy of an atom is expanded with respect to the charge, q, where the first derivative of the energy with respect to charge is the electronegativity, χ, and the second is the hardness, μ:

$$U_i = U_i^0 + \chi_i^0 q_i + \frac{1}{2}\mu_i^0 q_i^2 + \frac{1}{2} q_i V_i.$$

The final term in the above expression is the interaction with the Coulomb potential due to other atoms within the system. By solving the coupled set of equations for all atoms simultaneously, this leads to a set of charges that balance the chemical potential of the system. There are two variants of this method in general use. In the first, typified by the work of Mortier and co-workers [40], the Coulomb interaction, J, is described by a simple $1/r$ form. The alternative, as used by Rappe and Goddard in their QEq method [41], is to use a damped Coulomb potential that allows for the fact that at short distances the interaction arises from the overlap of electron density, rather than from just simple point ions. Hence, in QEq the potential is calculated based on the interaction of s orbitals with appropriate exponents. A further variant tries to encapsulate the short range damping of the Coulomb interaction in a mathematically more efficient form [42];

$$J_{ij} = \frac{1}{(r_{ij}^3 + \gamma_{ij}^{-3})^{\frac{1}{3}}},$$

where the pairwise terms γ_{ij} are typically determined according to combination rules in order to minimize

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

the number of free parameters:

$$\gamma_{ij} = \sqrt{\gamma_i \gamma_j}.$$

In the case of hydrogen, the variation of charge is so extreme between the hydride and proton limits that it was necessary to make the electronegativity a function of the charge itself in the original QEq scheme. As a result, the solution for the charges now requires an iterative self-consistent process. Here we propose a modified formulation from that of Rappe and Goddard that drastically simplifies the calculation of analytic derivatives:

$$U_H = U_H^0 + \chi_H^0 q_H + \frac{1}{2} J_{HH}^0 \left(1 + \frac{2q_H}{3\zeta_H^0}\right) q_H^2 + \frac{1}{2} q_H V_H.$$

The reason for the simplification is based on the Hellmann–Feynman theorem and can be understood as follows. If we consider the Cartesian first derivatives of the variable charge energy we arrive at:

$$\left(\frac{dU}{d\alpha}\right) = \left(\frac{\partial U}{\partial \alpha}\right)_q + \left(\frac{\partial U}{\partial q}\right)_\alpha \left(\frac{\partial q}{\partial \alpha}\right),$$

where the first term represents the conventional fixed charge derivative, and the second term is the contribution from the variation in charge as a function of structure. However, if the charges at each geometry are chosen so as to minimize the total energy of the system, then the first derivative of the internal energy with respect to charge is zero and so the correction disappears. Consequently, it only becomes necessary to evaluate the first derivatives of the charges with respect to position when calculating the second derivative matrix.

Dispersion Interactions

After the Coulomb energy, the most long-ranged of the energy contributions is usually the dispersion term. From quantum theory we know that the form of the interaction is a series of terms in increasing inverse powers of the interatomic distance, where even powers are usually the most significant:

$$U_{ij}^{\text{dispersion}} = -\frac{C_6}{r_{ij}^6} - \frac{C_8}{r_{ij}^8} - \frac{C_{10}}{r_{ij}^{10}} - \ldots.$$

The first term represents the instanteous dipole–instanteous dipole interaction energy, and the subsequent terms correspond to interactions between higher order fluctuating moments. Often in simulations only the first term, C_6/r^6, is considered as the dominant contribution. Again, the dispersion term can cause difficulties due to the slow convergence with respect to a radial cut-off. Although the series is absolutely convergent, unlike the Coulomb sum, the fact that all contributions are attractive implies that there is no cancellation between shells of atoms. The problem can be remedied by using an Ewald-style summation [43] to accelerate convergence for materials where the dispersion term is large:

$$U_{C_6}^{\text{recip}} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij} \left(\frac{\pi^{\frac{3}{2}}}{12V}\right) \sum_G \exp(iG\cdot r_{ij}) G^3 \times \left[\pi^{\frac{1}{2}} erfc\left(\frac{G}{2\eta^{\frac{1}{2}}}\right) + \left(\frac{4\eta^{\frac{3}{2}}}{G^3} - \frac{2\eta^{\frac{1}{2}}}{G}\right) \exp\left(-\frac{G^2}{4\eta}\right)\right]$$

$$U_{C_6}^{\text{real}} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{C_{ij}}{r^6} \left(1 + \eta r_{ij}^2 + \frac{\eta^2 r_{ij}^4}{2}\right) \exp\left(-\eta r_{ij}^2\right)$$

$$U_{C_6}^{\text{self}} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} -\frac{C_{ij}}{3V}(\pi\eta)^{\frac{3}{2}} + \sum_{i=1}^{N} \frac{C_{ii}\eta^3}{6}.$$

There can also be problems with the dispersion energy at short-range since it tends to negative infinity faster than some expressions for the repulsion between atoms, thus leading to collapse of the system. This can be resolved by recognizing that the above expansion for the dispersion energy is only valid for non-overlapping systems, and that at short-range the contribution decays to zero as it becomes part of the intrinsic correlation energy of the atom. To allow for this, Tang and Toennies [44] proposed that the dispersion energy is exponentially damped as the distance tends to zero according to the function:

$$f_{2n}(r_{ij}) = 1 - \left\{\sum_{k=0}^{2n} \frac{(br_{ij})^k}{k!}\right\} \exp(-br_{ij}).$$

Two-body Short-range Interactions

Contributions to the energy must be included that represent the interaction between atoms when they are bonded, or ions when they are in the immediate coordination shells. For the ionic case, a repulsive potential is usually adequate, with the most common choices being either a positive term, which varies inversely with distance, or an exponential form. These lead to the Lennard–Jones and Buckingham potentials, respectively, when combined with the attractive C_6 term:

$$U_{ij}^{\text{Buckingham}} = A \exp\left(-\frac{r_{ij}}{\rho}\right) - \frac{C_6}{r_{ij}^6}$$

$$U_{ij}^{\text{Lennard-Jones}} = \frac{C_m}{r_{ij}^m} - \frac{C_6}{r_{ij}^6}.$$

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

The Buckingham potential is easier to justify from a theoretical perspective since the repulsion between overlapping electron densities, due to the Pauli principle, which take an exponential form at reasonable distances. However, the Lennard–Jones potential, where the exponent is typically 9–12, is more robust since the repulsion increases faster with decreasing distance than the attractive dispersion term.

For covalently bonded atoms, it is often preferable to Coulomb subtract the interaction and to describe it with either a harmonic or Morse potential. In doing so, the result is a potential where the parameters have physical significance. For instance, in the case of the Morse potential the parameters become the dissociation energy of the diatomic species, the equilibrium bond length and a third term, which coupled with the dissociation energy, is related to the vibrational frequency for the stretching mode.

This does not represent an exhaustive list of the forms used to describe short-range interactions, but most other forms are closely related to the above functional forms, with the exception of the use of a spline function, which consists of a tabulation of function values versus distance. A full list of the two-body potential functional forms presently available is given in Table I.

Polarizability

The Coulomb interaction introduced previously is just the first term of an expansion involving moments of the charge density of an atom which includes the monopole, dipole, quadrupole, etc. Unlike the monopole term, it is generally unreasonable to assume that the dipole moment of an atom is fixed, since both the magnitude and direction readily alter within the crystalline environment according to the polarizability of the species. There are two approaches to modeling the polarizability that have been widely used, which we will now introduce.

The first, and most intuitive model is to use a point ion dipolar polarizability, α, which, in the presence of

TABLE I Two-body interatomic potentials currently available within GULP

Potential name	*Functional form*
Buckingham	$A\exp\left(-\frac{r}{\rho}\right)-\frac{C}{r^6}$
Lennard–Jones	$\frac{A}{r^m}-\frac{B}{r^n}$
Lennard–Jones (ε,σ)	$\varepsilon\left[\left(\frac{n}{m-n}\right)\left(\frac{\sigma}{r}\right)^m-\left(\frac{m}{m-n}\right)\left(\frac{\sigma}{r}\right)^n\right]$
Lennard–Jones (ε,σ) zero	$\varepsilon\left[\left(\frac{n}{m-n}\right)\left(\frac{m}{n}\right)^{\frac{m}{m-n}}\left(\frac{\sigma}{r}\right)^m-\left(\frac{m}{m-n}\right)\left(\frac{m}{n}\right)^{\frac{n}{m-n}}\left(\frac{\sigma}{r}\right)^n\right]$
Lennard − Jones buffered	$\frac{A}{(r+r_0)^m}-\frac{B}{(r+r_0)^n}$
Morse	$D_e[(1-\exp(-a(r-r_0)))^2-1]$
Harmonic	$\frac{1}{2}k_2(r-r_0)^2+\frac{1}{6}k_3(r-r_0)^3+\frac{1}{24}k_4(r-r_0)^4$
General/Del Re	$\frac{A\exp\left(-\frac{r}{\rho}\right)}{r^m}-\frac{C}{r^n}$
Coulomb-subtract	$-\frac{q_iq_j}{r}$
Four-range Buckingham	$A\exp\left(-\frac{r}{\rho}\right)$, $a_0+a_1r+a_2r^2+a_3r^3+a_4r^4+a_5r^5$, $b_0+b_1r+b_2r^2+b_3r^3$, $-\frac{C}{r^6}$
Stillinger–Weber 2-body	$A\exp\left(\frac{\rho}{r-r_{\text{cutoff}}}\right)\left(\frac{B}{r^4}-1\right)$
Inverse Gaussian	$-A\exp\left(-b(r-r_0)^2\right)$
Tang–Toennes	$-\left(\frac{C_6}{r^6}\right)f_6(r)-\left(\frac{C_8}{r^8}\right)f_8(r)-\left(\frac{C_{10}}{r^{10}}\right)f_{10}(r)$
Qtaper	$\left(\frac{q_iq_j}{r}\right)f(r)+C(1-f(r))$ where $f(r)$ is a taper function
Polynomial	$c_0+c_1r+c_2r^2+c_3r^3+c_4r^4+c_5r^5$
Qerfc	$\left(\frac{q_iq_j}{r}\right)erfc\left(\frac{r}{\rho}\right)$
Covalent exponential	$-D\exp\left(-\frac{a(r-r_0)^2}{2r}\right)$
Rydberg	$-A\left(1+B\left(\frac{r}{r_0}-1\right)\right)\exp\left(-B\left(\frac{r}{r_0}-1\right)\right)$
Fermi–Dirac	$\frac{A}{(1+\exp(B(r-r_0)))}$

Here r represents the interatomic distance of which the potential is a function, and q denotes the atomic charge of the species. All other values are parameters of the potentials.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

an electric field, V_f, will give rise to a dipole moment, μ, and energy of interaction as given below:

$$\mu = \alpha V_f$$

$$U^{\text{polarization}} = -\frac{1}{2}\alpha V_f^2.$$

This approach has the advantage that it is readily extended to higher order polarizabilities, such as quadrupolar, etc. [45]. It has been applied both in the area of molecular crystals, though often fixed moments are sufficient here [46], and, more recently, to ionic materials by Wilson and Madden [47]. The only disadvantage of this approach is that the polarizability is independent of the environment, which implies that it is undamped at extreme electric fields and can lead to a polarization catastrophe. It is well documented that the polarizability of the oxide ion is very sensitive to its location, since in the gas phase the second electron is unbound and only associates in the solid state due to the Madelung potential [48]. A further complication is that the scheme must involve a self-consistency cycle if the induced multipoles on one atomic center are allowed to interact with those on another, though in some approaches this is neglected for simplicity.

The second approach to the inclusion of dipolar polarizability is via the shell model first introduced by Dick and Overhauser [49]. Here a simple mechanical model is used, whereby an ion is divided into a core, which represents the nucleus and inner electrons of the ion and therefore has all of the mass associated with it, and a shell, which mimics the valence electrons. Although it is convenient to think in terms of this physical picture, it should not be taken too literally as in some situations the shell can carry a positive charge, particularly for metal cations. The core and shell are Coulombically screened from each other, but coupled by a harmonic spring of force constant k_{cs}. If the shell charge is q_{s}, then the polarizability of the ion *in vacuo* is given by;

$$\alpha = \frac{q_s^2}{k_{\text{cs}}}.$$

By convention, the short-range forces are specified to act on the shell, while the Coulomb potential acts on both. Hence, the short-range forces act to damp the polarizability by effectively increasing the spring constant, and thus the polarizability is now environment dependent. The shell model has been widely adopted within the ionic materials community, particularly within the UK. Although the same issue exists as for point ion polarizabilities, namely that self-consistency has to be achieved for the interaction of the dipoles due to the positions of the shells, the problem is transformed into a coordinate optimization one. This can be solved concurrently with the optimization of the atomic core positions. The main disadvantage of this approach is that it is not naturally extensible to higher order moments, though some attempts have been made, such as the spherical and elliptical breathing shell models. Furthermore, when performing molecular dynamics special treatment of the shells must be made by either using an adiabatic approach, in which the shells are optimized at every time step, or by using a technique analogous to the Car–Parrinello method [50], in which a fictitious mass is assigned to the shell [51].

As a final note on the topic of polarizability, it is impossible to distinguish from a phenomenological point of view between on site ion polarization and charge transfer between ions. This may explain why the combination of formal charges with the shell model has been so successful for modeling materials that are quite covalent, such as silica polymorphs. Provided the crystal symmetry is low enough, the shell model could be viewed as representing charge transfer/covalency.

Radial Interactions

There is a refinement to the conventional point particle shell model, which is the so-called breathing shell model, that introduces non-central ion forces [52]. Here the ion is assigned a finite radius, R_0, and then all the short-range repulsion potentials act upon the radius of the ion, rather than the nuclear position. A radial constraining potential is then added which represents the self-energy of the ion. Two functional forms are most commonly used:

$$U_i^{\text{BSM-Harmonic}} = \frac{1}{2}K_{\text{BSM}}(R_i - R_0)^2$$

$$U_i^{\text{BSM-Exponential}} = K'_{\text{BSM}}(\exp(\rho(R_i - R_0)) + \exp(-\rho(R_i - R_0))).$$

This model has two important consequences. Firstly, it allows the change of radius between two different coordination environments to be modeled—for example, octahedral versus tetrahedral. This represents an alternative to using different repulsive parameters in the Buckingham potential by scaling the A term according to $\exp(-\rho_{\text{tet}}/\rho_{\text{oct}})$ to correct for this effect. Secondly, the coupling of the repulsive interactions via a common shell radius creates a many-body effect that is able to describe the Cauchy violation ($C_{12} \neq C_{44}$) for rock salt structured materials.

Three-body Interactions

There are two physical interpretations for the introduction of three-body terms, depending on whether you take a covalent or ionic perspective. Within the former view, as adopted by molecular

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

TABLE II Three-body interatomic potentials currently available in GULP

Potential name	*Functional form*
Three (harmonic)	$\frac{1}{2}k_2(\theta-\theta_0)^2+\frac{1}{6}k_3(\theta-\theta_0)^3+\frac{1}{24}k_4(\theta-\theta_0)^4$
Three (exponential-harmonic)	$\frac{1}{2}k_2(\theta-\theta_0)^2\exp\left(-\frac{r_{12}}{\rho_{12}}\right)\exp\left(-\frac{r_{13}}{\rho_{13}}\right)$
Three (exponential)	$k\exp\left(-\frac{r_{12}}{\rho_{12}}\right)\exp\left(-\frac{r_{13}}{\rho_{13}}\right)\exp\left(-\frac{r_{23}}{\rho_{23}}\right)$
Axilrod–Teller	$k\frac{(1+3\cos(\theta_{123})\cos(\theta_{231})\cos(\theta_{312}))}{r_{12}^3r_{13}^3r_{23}^3}$
Stillinger–Weber 3-body	$k\exp\left(\frac{\rho_{12}}{r_{12}-r_{12}^{\text{cutoff}}}+\frac{\rho_{13}}{r_{13}-r_{13}^{\text{cutoff}}}\right)(\cos(\theta)-\cos(\theta_0))^2$
Bcross	$k(r_{12}-r_{12}^0)(r_{13}-r_{13}^0)$
Urey–Bradley	$\frac{1}{2}k_2(r_{23}-r_{23}^0)^2$
Vessal	$\frac{k_2((\theta_0-\pi)^2-(\theta-\pi)^2)^2}{8(\theta_0-\pi)^2}\exp\left(-\frac{r_{12}}{\rho_{12}}\right)\exp\left(-\frac{r_{13}}{\rho_{13}}\right)$
Cosine-harmonic	$\frac{1}{2}k_2(\cos(\theta)-\cos(\theta_0))^2$
Murrell–Mottram	$k\exp(-\rho Q_1)f_{MM}(Q_1,Q_2,Q_3)$
	$f_{MM}=c_0+c_1Q_1+c_2Q_1^2+c_3(Q_2^2+Q_3^2)+c_4Q_1^3+c_5Q_1(Q_2^2+Q_3^2)$
	$+(c_6+c_{10}Q_1)(Q_3^3-3Q_3Q_2^2)+c_7Q_1^4+c_8Q_1^2(Q_2^2+Q_3^2)+c_9(Q_2^2+Q_3^2)^2$
	$Q_1=\frac{R_1+R_2+R_3}{\sqrt{3}},Q_2=\frac{R_2-R_3}{\sqrt{2}},Q_3=\frac{2R_1-R_2-R_3}{\sqrt{6}}$
	$R_1=\frac{r_{12}-r_{12}^0}{r_{12}^0},R_2=\frac{r_{13}-r_{13}^0}{r_{13}^0},R_3=\frac{r_{23}-r_{23}^0}{r_{23}^0}$
Bacross	$(k_{12}(r_{12}-r_{12}^0)+k_{13}(r_{13}-r_{13}^0))(\theta-\theta_0)$
Linear-three-body	$k(1\pm\cos(n\theta))$
Bcoscross	$k(1+b\cos^m(n\theta))(r_{12}-r_{12}^0)(r_{13}-r_{13}^0)$

For potentials with a unique pivot atom, this atom is taken to be atom 1 and θ is the angle between the vectors r_{12} and r_{13}. All terms other than θ, θ_{123}, θ_{231}, θ_{312}, r_{12}, r_{13}, r_{23} are parameters of the potential.

mechanics, the three-body potential represents the repulsion between bond pairs, or even occasionally lone pairs. Hence, the form chosen is usually a harmonic one that penalizes deviation from the expected angle for the coordination environment, such 120° for a trigonal planar carbon atom:

$$U_{ijk}=\frac{1}{2}k_2(\theta-\theta_0)^2.$$

At the other end of the spectrum, ionic materials possess three-body forces due to the three-center dispersion contribution, particularly between the more polarizable anions. This is typically modeled by the Axilrod–Teller potential [53]:

$$U_{ijk}=k\frac{\left(1+3\cos\left(\theta_{ijk}\right)\cos\left(\theta_{jki}\right)\cos\left(\theta_{kij}\right)\right)}{r_{ij}^3r_{jk}^3r_{ik}^3}.$$

As with two-body potentials, there are many variations on the above themes, such as coupling the three-body potential to the interatomic distances, but the physical reasoning is often the same. A full tabulation of the three-body potentials is given in Table II.

Four-body Interactions

Specific four-body interactions are usually only included in molecular mechanics forcefields where

TABLE III Four-body interatomic potentials currently available in GULP

Potential name	*Functional form*
Torsional	$k_4(1\pm\cos(n\phi-\phi_0))$
Ryckaert–Bellemanns	$\sum_{n=0}^5c_n\cos^n\phi$
Out of plane	$k_2d^2+k_4d^4$
ESFF torsion	$k_1\sin^2\theta_{123}\sin^2\theta_{234}+k_2\sin^n\theta_{123}\sin^n\theta_{234}\cos(n\phi)$
Torsional harmonic	$\frac{1}{2}k_2(\phi-\phi_0)^2$

Here the potential acts on the sequence of atoms 1–2–3–4 (except for the out of plane potential), where the torsion angle ϕ lies between the plane containing atoms 1–2–3 and atoms 2–3–4, θ_{ijk} is the angle between the vectors r_{ji} and r_{jk} and d represents the distance of atom 1 out of the plane of atoms 2–3–4.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

TABLE IV Density functionals available within the embedded atom method

Functional	*Functional form*
Power law	$f(\rho) = \rho^{\frac{1}{n}}$
Banerjea and Smith [54]	$f(\rho) = c_0\left(1 - \frac{1}{n}\ln\left(\frac{\rho}{\rho_0}\right)\right)\left(\frac{\rho}{\rho_0}\right)^{\frac{1}{n}} + c_1\left(\frac{\rho}{\rho_0}\right)$

they act to describe torsional angles. Hence, the functional form usually involves the cosine of the torsional angle with factors that reflect the equilibrium torsional angle, ϕ_0, and the periodicity with respect to rotation about the central bond. The most widely used form is therefore:

$$U_{ijkl} = k_4(1 + m\cos(n\phi - \phi_0)).$$

The other form of torsional potential more occasionally found is one that employs a harmonic potential to describe the out of plane bending mode of a central atom that has a planar coordination geometry. This is of utility when describing aromatic systems, and has also been used in the modeling of the carbonate anion. An alternative to this potential is to use so-called improper torsions, where the planar geometry is maintained by specifying a torsional potential between atoms that are not bonded. A full list of the available functional forms of four-body potentials is given in Table III.

Many-body Interactions

Some important interactions for particular systems cannot be described within the above forcefield framework. Below we describe a few selected higher order interaction potentials that are of significance and that are becoming more widely used, despite the greater computational cost.

The Embedded Atom Method

The embedded atom model (EAM) is an approach that has been successful in the description of metallic systems. Its foundations lie within density functional theory, and are based on the tenet that the energy is a function of the electron density. To simplify things, the EAM considers that the electron density is a superposition of the atomic densities, and that instead of integrating the density across all space it is sufficient just to express the energy as a function of the density at the nucleus of an atom, summed over all particles:

$$U^{\mathrm{EAM}} = -\sum_{i=1}^{N} f(\rho_i).$$

The above equations encapsulate the idea that the interaction between any given pair of atoms is dependent on the number of other atoms within the coordination sphere. Within this generic scheme there are a number of variations, based around different functionals of the density (Table IV) and different representations of how the density varies with distance (Table V).

In the original work of Sutton and Chen [55], which developed and extended the ideas of Finnis and Sinclair [56], a square root was used as the density functional, while the density itself was represented as an inverse power of the interatomic distance. One of the beauties of the EAM is that, in principle, once the metal is parameterized it can be studied in other environments, such as alloys, without further modification. On the downside, the prediction of the relative stability of phases can be sensitive to the cut-off radius chosen, though if care is taken this problem can be surmounted [57].

TABLE V Functional forms for the distance dependence of the atomic density available within the embedded atom method

Density	*Functional form*
Power law	$\rho_{ij} = cr_{ij}^{-n}$
Exponential	$\rho_{ij} = cr_{ij}^{n}\exp(-d(r_{ij} - r_0))$
Gaussian	$\rho_{ij} = cr_{ij}^{n}\exp(-d(r_{ij} - r_0)^2)$
Cubic	$\rho_{ij} = c(r_0 - r_{ij})^3$

Bond Order Potentials

Related in many ways to the embedded atom method, but with a more sophisticated formalism, are so-called bond order potentials. It was recognized by Abell [58] that the local binding energy could be expressed as follows:

$$U^{\mathrm{BO}} = \sum_{i=2}^{N}\sum_{j=1}^{i-1}[U^{\mathrm{repulsive}}(r_{ij}) - B_{ij}U^{\mathrm{attractive}}(r_{ij})],$$

where B_{ij} is the bond order between the atoms i and j. The bond order is dependent on the local environment of both atoms and thereby converts an apparent two-body interaction into a many-body one. Several different formulations have been proposed, most notably by Tersoff [59,60], and also more recently by Pettifor and co-workers [61], where the latter use a more extensive analysis of the contributions to the bond order and appeal to first principles methods to extract the parameters. One particular model has had an enormous impact in recent years due to its applicability to carbon polymorphs and hydrocarbon systems, that due to Brenner and co-workers [62]. Unsurprisingly, it has been extensively applied to fullerenes, nanotubes and diamond, as systems of topical interest. One of the other reasons for the popularity of the model is the fact that Brenner makes his code freely available. An independent implementation of the Brenner model has been made within GULP, since the capabilities of the program require that

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

analytic derivatives to at least second order, and preferably third, are present, which is not the case for the existing code. To date, there exist three published variants of the Brenner potentials, but we have implemented just the latest of these models since it supercedes the previous two [63].

The terms in the expression for the energy in the Brenner model are expressed as follows:

$$U^{\text{repulsive}}(r) = Af^c(r)\left(1+\frac{Q}{r}\right)\exp\left(-\alpha r\right)$$

$$U^{\text{attractive}}(r) = f^c(r)\sum_{n=1}^{3} B_n \exp\left(-\beta_n r\right),$$

where A, Q, α, B_{1-3} and β_{1-3} are parameterized constants that depend on the atomic species, C or H, involved and $f^c(r)$ is a cosine tapering function to ensure that the potential goes smoothly to zero of the form:

$$f^c(r) = \begin{cases} 1 & r < r^{\min} \\ \frac{1}{2}\left[1+\cos\left(\frac{(r-r^{\min})}{(r^{\max}-r^{\min})}\pi\right)\right] & r^{\min} < r < r^{\max} \\ 0 & r > r^{\max} \end{cases}.$$

The bond order term itself is composed of several terms;

$$B_{ij} = \frac{1}{2}\left(b_{ij}^{\sigma-\pi} + b_{ji}^{\sigma-\pi}\right) + \frac{1}{2}\left(\Pi_{ij}^{\text{RC}} + b_{ij}^{\text{DH}}\right).$$

Note that the above expression for B_{ij} differs from the one given in the defining manuscript due to the factor of a half for the second term, but is required to obtain results that agree with those quoted. The first two terms in the above equation represent the influence of local bond lengths and angles about the atoms i and j, respectively, while the third term is a correction for radical character, and the fourth one for the influence of dihedral angles. Both of the last two terms are related to the degree of conjugation present. Full defining equations for these terms, along with the parameters, can be found in the original reference and the subsequent errata.

In the above many-body contributions to the bond order, bicubic and tricubic splines are used to interpolate parameter values. For the distributed Brenner potential code, the spline coefficients are precomputed and supplied as data files. In the present implementation the splines are performed internally on the fly. This has two advantages in that it both avoids possible transcription errors, as well as loss of precision through I/O, and allows for the possibility of parameter fitting to be readily implemented.

Because of the short-ranged nature of the Brenner potential we have implemented two different algorithms for the evaluation of the interactions. The first involves a conventional search over all atoms to find neighbors with a non-zero interaction. The second uses a spatial decomposition of the system into cubes of side length equal to the maximum range of the potential. Consequently, only atoms within neighboring cubes can possibly interact. This leads to a linear scaling algorithm that is far more efficient for large systems. A comparison is presented in the "Results" section.

While the Brenner model does have many strengths, such as its ability to describe bond dissociation, there are also a few limitations. Perhaps the most significant is the difficulty in describing long-range forces. For instance, there is no bonding between the sheets for graphite. There have been a number of remedies proposed, including adding on two-body potentials to describe these effects, either only between different molecules, or with a tapering that removes the interaction at short-range so as not to invalidate the parameterization. However, there are limitations to these approaches, though a more sophisticated expression for removing the contribution of long-range forces where the existing interactions due to the Brenner potential are significant shows promise [64]. As yet, this is still to be implemented.

The Angular Overlap Model (AOM)

One of the most problematic types of systems to model with a forcefield are those containing transition metal ions. Often the local symmetry of the environment about such ions is of a form that cannot be described within a simple dipolar shell model. The best example of this occurs for ions in a distorted octahedral environment where the anti-bonding e_g^* orbitals are partially occupied-the so-called Jahn–Teller distortion. It is impossible to reproduce these distortions within the framework of a purely atomistic model without the use of some unsatisfying constructs (for instance, specifying different interatomic potentials for different atoms of the same ligand element). However, the approximate inclusion of ligand field effects can be achieved through the addition of an empirical on-site Hamiltonian for transition metals based on the AOM [65].

The basis of the AOM approach is to assume that all ligands overlap with the transition metal d orbitals as a simple s orbital, thus excluding π-bonding effects. For species such as the oxide ion, this is a good approximation given its position in the spectrochemical series. In the original method, the energy of each d orbital is then evaluated according to the sum of the angular contributions over all the ligand positions, while the radial term is a constant, assuming that all ligands are equidistant. For the practical implementation of the technique in a forcefield context, it is necessary to modify the method since we require that

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

the interactions decay with distance (so that the number of ligands can vary continuously) and that the energy obtained be rotationally invariant (so that the answer obtained does not depend on the Cartesian frame of reference). Consequently, the orbital energies are defined by diagonalizing the following Hamiltonian matrix:

$$H_{dd'} = \sum^{\text{ligands}} f_{\text{radial}}(r)\Gamma_d\Gamma_{d'},$$

where the terms Γ represent the spherical harmonics of the orbitals as given below;

$$\Gamma_{z^2} = \frac{1}{4}(1 + 3\cos 2\theta)$$

$$\Gamma_{x^2-y^2} = \frac{\sqrt{3}}{4}\cos 2\phi(1 - \cos 2\theta)$$

$$\Gamma_{xy} = \frac{\sqrt{3}}{4}\sin 2\phi(1 - \cos 2\theta)$$

$$\Gamma_{xz} = \frac{\sqrt{3}}{2}\cos\phi\sin 2\theta$$

$$\Gamma_{yz} = \frac{\sqrt{3}}{2}\sin\phi\sin 2\theta$$

and the radial function is chosen to be of exponential form:

$$f_{\text{radial}}(r) = A\exp\left(-\frac{r}{\rho}\right).$$

Having obtained the orbital eigenvalues, the levels are populated with the appropriate number of up and down spin electrons according to the charge and electronic configuration of the transition metal ion. In order to maintain a continuous energy with respect to these occupations, particularly when degeneracies occur, a Fermi–Dirac function is used with the Fermi level being chosen in order to obtain the correct total number of electrons. Further details of the method are given in the papers by Woodley *et al.* [66,67], as well as details of the successful application of the method to $LaMnO_3$ where the $Mn^{3+}(d^4)$ ion is Jahn–Teller distorted.

One-body Interactions

Going to the other extreme of complexity from many-body interactions, we have the simplest possible atomistic model, namely the Einstein model [68]. This approach deviates from those that have gone before in that there is no interaction between particles and all atoms are simply tethered to their lattice site by harmonic springs:

$$U^{\text{Einstein}} = \frac{1}{2}\sum_{i=1}^{N} k_i\left((x_i - x_i^0)^2 + (y_i - y_i^0)^2 + (z_i - z_i^0)^2\right).$$

This model acts as a reference state since all interactions are purely harmonic and consequently the phonon density of states, and also the structure, are independent of temperature. This implies that all the quantities that can be derived from the vibrational partition function can be analytically determined for any temperature without approximation, unlike other potential models. Even a structure that consists of atoms coupled together by a series of harmonic potentials has implied anharmonicity that arises from the derivatives of the transformation matrix from the bond-oriented pairwise frame of reference to the Cartesian one.

Given that the results of the Einstein model can be derived without recourse to a structural description, there is no need to employ an atomistic simulation program in order to calculate the required quantities. However, it can be useful to combine the Einstein model with a conventional, more accurate, representation of the interactions for use in thermodynamic integration [69]. Since the free energy of the Einstein crystal is known under any conditions, it is possible to extract free energies from molecular dynamics via a series of perturbative runs in which the Einstein model is introduced to an anharmonic potential in a series of steps, as a function of a switching parameter, λ, in order to obtain the difference relative to the known value.

It is worth highlighting the differences between the Einstein model and all others within the program. Because of the lack of interatomic interactions, there can be no optimization of the structure and no strain related properties calculated. For the same reason, there is no phonon dispersion across the Brillouin zone. Finally, because the particles are tethered to lattice sites there is no translational invariance, and consequently all vibrational frequencies are positive and non-zero (assuming all force constants are specified to be likewise).

Potential Truncation

As previously mentioned, all potentials must have a finite range in order to be calculable. However, there is a range of conditions for a potential to act between two species, as well as various methods for handling the truncation of potentials. Hence, it is appropriate to describe a few of these issues here.

Radial Truncation

The natural way to truncate a potential is through the use of a spherical cut-off radius. It is also possible to specify a minimum radius from which the potential should act too. Hence, it is possible to create multiple distance ranges within a potential with different forcefield parameters, as well as to overlay different potentials. Of course, where there is a cut-off boundary, there will be a discontinuity in the energy for most

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

interaction terms, unless the potential smoothly tends to zero at that distance by design. This can lead to problems during energy minimization, since the point of lowest energy may not be a point of zero force anymore, and in other simulation techniques, such as molecular dynamics, where energy conservation will be affected by discontinuities. Other than increasing the cut-off radius, there are several approaches to minimizing these difficulties as described briefly in the subsequent subsections.

Cut and Shift

To avoid a discontinuity in the energy at the potential cut-off, a constant shift can be added to the potential so that the energy at the cut-off boundary becomes zero:

$$U_{ij}^{\text{shifted}}(r_{ij}) = U_{ij}(r_{ij}) - U_{ij}(r_{\text{cut}}),$$

where r_{cut} is the cut-off radius. However, the gradient is still discontinuous at the cut-off distance, so the procedure can be extended by adding a linear term in the distance such that the gradient is also zero at this point. While in principle this method can be applied to make any order of derivative go exactly to zero at the boundary by construction, the increasing powers of distance lead to the correction terms modifying the variation of the potential with distance more strongly as the order rises. This characteristic, that the potential is modified away from the point of cut-off, makes this method of smoothing less desirable than some.

Tapering

In this approach, the potential is multiplied by a smooth taper function that goes to zero, both for its value and its derivatives typically up to second order. This is usually applied over a short range from an inner radius, r_{taper}, to the cut-off distance:

$$\begin{array}{cc} U_{ij}(r_{ij}) & r_{ij} < r_{\text{taper}} \\ U_{ij}(r_{ij}) f^{\text{taper}}(r_{ij}) & r_{\text{taper}} < r_{ij} < r_{\text{cut}} \\ 0 & r_{ij} > r_{\text{cut}} \end{array}.$$

Hence, within r_{taper} the potential is unchanged. There are numerous possible functional forms that satisfy the required criteria for a taper function. Perhaps the two most commonly used are a fifth order polynomial or a cosine function with half a wavelength equal to $r_{\text{cut}} - r_{\text{taper}}$. Both of these are available within GULP as part of the overall two-body potential cut-off.

Molecular Mechanics

In the simulation of molecular systems, it is often preferable to use a molecular mechanics forcefield. This implies that certain cut-offs are determined by connectivity, rather than by distance alone. For example, particular potentials may only act between those atoms that are bonded, such as a harmonic force constant, whereas others may specifically only act between those that are not covalently linked. Fundamental to this is the notion of a bond between atoms. Within GULP these can either be determined automatically by comparing distances between pairs of atoms with the sum of the covalent radii, multiplied by a suitable tolerance factor, or alternatively the connectivity can be user specified. From this information, it is possible for the program to determine the number of molecules and, in the case of periodic systems, their dimensionality.

A number of options are possible that control how both the potentials and the Coulomb terms act, as described below:

- *Bonded*: A potential may act only between atoms that are bonded (1–2).
- *Exclude 1–2*: The case of bonded atoms is specifically omitted from the allowed interactions.
- *Exclude 1–3*: Interactions between bonded atoms and those with a common bonded atom are excluded.
- *Only 1–4*: A potential only acts between atoms that are three bonds apart. This can be useful in the description of torsional interactions.
- *Intramolecular*: Only interactions within a given molecule are permitted for the potential.
- *Intermolecular*: Only interactions between atoms of different molecules are allowed.
- *Molecular Coulomb subtraction*: All electrostatic interactions between atoms within the same molecule are excluded. This implies that the charges on the atoms within a molecule purely serve to describe its interaction with other molecules.
- *Molecular mechanics Coulomb treatment*: This implies that Coulomb terms are excluded for all 1–2 and 1–3 interactions. In addition, it is sometimes desirable to remove, or scale down, 1–4 electrostatic interactions too. The benefit of this is that the parameters of the intramolecular potentials then have a direct correspondence with the equilibrium bond lengths, bond angles and localized vibrational frequencies. It should be noted that two-body potentials can also be scaled for the 1–4 interaction, if so desired.

Partial Occupancy

Many materials have complex structures which include disorder of one form or another. This can typically consist of partial occupancy where there are more symmetry degenerate sites than there are ions to occupy them, or where there are mixtures of ions that share the same structural position. In principle, the only way to accurately model such systems is to construct a supercell of the material containing

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

a composition consistent with the required stoichiometry and then to search through configuration space for the most stable local minima, assuming the system is in thermodynamic equilibrium (which may not always be the case). This includes allowing for the contribution from the configurational entropy to the relative stability. While this approach has been taken for several situations, most notably some of the disordered polymorphs of alumina [70,71], it is demanding because of the shear number of possibilities which increases with a factorial dependence.

There is an approximate approach to the handling of disorder, which is to introduce a mean-field approach. Here all atoms are assigned an occupancy factor, o_i, where $0 \leq o_i \leq 1$, and all interactions are then scaled by the product of the relevant occupancies:

$$U_{ij}^{m-f} = o_i o_j U_{ij}$$

$$U_{ijk}^{m-f} = o_i o_j o_k U_{ijk}$$

and so on for high order terms. This approach can be utilized in any atomistic simulation program by scaling all the relevant potential parameters according to the above rules. However, for complex forcefields this can be tedious, and it also precludes fitting to multiple structures where the occupancies of ions of the same species are different. Hence, the inclusion of partial occupancies has been automated in GULP so that only the site occupancies have to be specified and everything else is handled by the program. This includes the adding of constraints so that atoms that share the same site move together as a single particle, as well as checking that the sum of the occupancies at any given site does not exceed one. It should be noted that the handling of partial occupancies requires particular care in determining phonons where the matrix elements for all coupled species must be condensed in order to obtain the correct number of modes.

The use of a mean field model is clearly an approximation. For structures, it can often work quite well, since crystallography returns an average anyway. However, for other quantities, such as thermodynamic data it has limitations. For example, the excess enthalpy of mixing of two phases is typically overestimated since the stabilization that arises from local structural distortions to accommodate particular species is omitted [72].

Structural Optimization

Having defined the internal energy of a system, the first task to be performed is to find the minimum energy structure for the present material. To be more precise, this will typically be a local minimum on the global potential energy surface that the starting coordinates lie closest to. Trying to locate the global energy minimum is a far more challenging task and one that has no guarantee of success, except for the simplest possible cases. There are several approaches to searching for global minima, including simulated annealing [73], via either the Monte Carlo or molecular dynamics methods, and genetic algorithms (GAs) [74,75]. Here we will focus the quest for a local minimum.

At any given point in configuration space, the internal energy may be expanded as a Taylor series:

$$U(x + \delta x) = U(x) + \frac{\partial U}{\partial x}\delta x + \frac{1}{2!}\frac{\partial^2 U}{\partial x^2}(\delta x)^2 + \ldots,$$

where the first derivatives can be collectively written as the gradient vector, g, and the second derivative matrix is referred to as the Hessian matrix, H. This expansion is usually truncated at either first or second order, since close to the minimum energy configuration we know that the system will behave harmonically.

If the expansion is truncated at first order then the minimization just involves calculating the energy and first derivatives, where the latter are used to determine the direction of movement and a line search is used to determine the magnitude of the step length. In the method of steepest descents this process is then repeated until convergence. However, this is known to be an inefficient strategy since all the previous information gained about the energy hypersurface is ignored. Hence, it is far more efficient to use the so-called conjugate gradients algorithm [76] where subsequent steps are made orthogonal to the previous search vectors. For a quadratic energy surface, this will converge to the minimum in a number of steps equal to the number of variables, N.

If we expand the energy to second order, and thereby use a Newton–Raphson procedure, then the displacement vector, Δx, from the current position to the minimum is given by the expression:

$$\Delta x = -H^{-1}g,$$

which is exact for a harmonic energy surface (i.e. if we know the inverse Hessian matrix and gradient vector at any given point then we can go to the minimum in one step). For a realistic energy surface, and starting away from the region close to the minimum, then the above expression becomes increasingly approximate. Furthermore, there is the danger that if the energy surface is close to some other stationary point, such as a transition state, then simply applying this formula iteratively may lead to a maximum, rather than the minimum. Consequently, the expression is modified to be

$$\Delta x = -\alpha H^{-1}g,$$

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

where α is a scalar quantity which is determined by performing a line search along the search direction to find the 1-D minimum and the procedure becomes iterative again, as per conjugate gradients.

By far, the most expensive step of the Newton–Raphson method, particularly once the size of the system increases, is the inversion of the Hessian. Furthermore, the Hessian may only vary slowly from one step to the next. It is therefore wasteful and undesirable to invert this matrix at every step of the optimization. This can be avoided through the use of updating formulae that use the change in the gradient and variables between cycles to modify the inverse Hessian such that it approaches the exact matrix. Two of the most widely employed updating schemes are those due to Davidon–Fletcher–Powell (DFP) [77] and Broyden–Fletcher–Goldfarb–Shanno (BFGS) [78], which are given below:

$$H_{i+1}^{\mathrm{DFP}} = H_i^{\mathrm{DFP}} + \frac{\Delta x \otimes \Delta x}{\Delta x \Delta g} - \frac{(H_i^{\mathrm{DFP}} \Delta g) \otimes (H_i^{\mathrm{DFP}} \Delta g)}{\Delta g H_i^{\mathrm{DFP}} \Delta g}$$

$$H_{i+1}^{\mathrm{BFGS}} = H_i^{\mathrm{BFGS}} + \frac{\Delta x \otimes \Delta x}{\Delta x \Delta g} - \frac{(H_i^{\mathrm{BFGS}} \Delta g) \otimes (H_i^{\mathrm{BFGS}} \Delta g)}{\Delta g H_i^{\mathrm{BFGS}} \Delta g} + \left[\Delta g H_i^{\mathrm{BFGS}} \Delta g\right] v \otimes v$$

$$v = \frac{\Delta x}{\Delta x \Delta g} - \frac{H_i^{\mathrm{BFGS}} \Delta g}{\Delta g H_i^{\mathrm{BFGS}} \Delta g}.$$

The BFGS algorithm is generally recognized as the more efficient one and so is the default optimizer. However, if performing a transition state search then DFP is preferred due to the different tendances of the updates with regard to targeting a positive definite character for the Hessian. The practical approach taken in a BFGS optimization is to initialize the Hessian by performing an exact inversion of the second derivatives and then subsequently updating for a number of cycles. Occasionally, it is necessary to calculate the exact inverse Hessian again. This is triggered by one of a number of possible situations:

- The maximum number of cycles for updating is exceeded.
- The angle between the gradient vector and the search vector exceeds a given threshold.
- The energy has dropped by more than a certain threshold in one cycle, so the curvature is likely to have altered.
- The energy cannot be lowered by line minimization along the current search vector.

There is one variation upon the BFGS strategy above, in which, rather than calculating the exact second derivative matrix and inverting it, the Hessian is initialized approximately and then updated such that it tends to the true Hessian. Two possible starting approximations to the Hessian are either to use a unit matrix or to perform a finite difference estimate of the on-diagonal elements only in order to achieve better preconditioning. By taking this approach the cost is similar to conjugate gradients, though with a higher memory requirement, but the convergence is typically twice as fast since more information from previous energy/gradient evaluations is retained.

The choice of appropriate algorithm for minimization depends on the size of the system and number of variables. For small systems, the Newton–Raphson methods that use second derivatives will be far more efficient. As the size of system increases, the computational cost of building and inverting the Hessian matrix will ultimately grow as N^3 while the memory requirements also increase as $N(N+1)/2$ for lower-half triangular storage. Hence, ultimately it will become necessary to use first the unit Hessian approach and then ultimately the conjugate gradients method as N continues to increase. This balance may also be influenced by the use of symmetry, where possible, as will be discussed later. A further issue is that for minimization where the initial structure lies outside the harmonic region, Newton–Raphson methods may not be particularly advantageous, whereas they will be as convergence is approached. Hence, GULP includes the facility to change the minimizer from, for example, conjugate gradients, to BFGS when a criterion is met, such as the gradient norm falling below a specified threshold.

Aside from the minimization algorithm itself, something should be said about the criteria for convergence of an optimization. Typically, some or all of the following can be checked for convergence:

- the gradient norm (root mean square gradient per variable),
- the maximum individual gradient component,
- the norm of the estimated displacement vector,
- the change in the function (energy, enthalpy or free energy) between successive cycles,
- the change in the variables between successive cycles.

Normally, all of these quantities are well converged (and often excessively so—i.e. well beyond chemical accuracy) with Newton–Raphson optimization for a forcefield. Occasionally, convergence is not achieved, which is nearly always indicative of a significant discontinuity in the energy surface, such as that caused by interatomic potential cut-offs. Ideally, all potentials should be tapered so that their value, first and second derivatives, go smoothly to zero at the cut-off radius. In practice, most repulsive potentials are negligibly small by a typical cut-off distance of 10 Å and so it is usually

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

the attractive potentials that mimic dispersion that are responsible. The other common cause are discontinuities in bonding—i.e. an atom is displaced beyond a bond cut-off during minimization and so the whole nature of the interaction potential changes. While this can be overcome by maintaining a fixed connectivity list, the real issue is that the forcefield is inadequate for the system being modeled.

All of the above discussion has related to the search for a local minimum where the first derivatives are zero and the second derivative matrix is positive definite. However, there are many instances in which it is necessary to locate other stationary points on the potential energy surface, and in particular transition states, such as for ion diffusion. There are a number of algorithms available for locating transition states, of which the most widely used class in solid state optimizations tend to be constrained minimizations. Here the reaction path is discretized in some way and the system minimized so that the variable changes are orthogonal to the direction of the pathway. Arguably, the most successful and widely used approach is the Nudged Elastic Band method [79]. However, the disadvantage of these constrained techniques is that some assumption usually has to be made about the nature of the mechanism before the calculations are performed. In contrast, there are algorithms which use Newton–Raphson techniques in order to locate stationary points with an arbitrary number of negative eigenvalues for the Hessian (where a transition state is the special case of a single negative eigenvalue). In the case of GULP, the method implemented is so-called rational functional optimization (RFO) [80].

In the RFO method, the inverse Hessian matrix is diagonalized to obtain the eigenvalues and eigenvectors. From the eigenvalues it is possible to examine whether the matrix has the required characteristics for the stationary point being sought. If the number of negative eigenvalues is incorrect, then the spectrum is level-shifted to correct this and the search direction constructed appropriately. By default, the Hessian modes with the smallest eigenvalues are followed towards the corresponding stationary point. However, eigenvector following can also be performed in which different modes from the spectrum are selected. Hence, the RFO optimizer will, in principle, locate various possible transition states starting from a given position. There are a few provisos though, most notably that there must be a non-zero projection of the gradient vector in the search direction leading to the stationary point. Furthermore, the step size must be quite small in order to ensure convergence to a transition state. Since the line search can no longer be used, the optimization relies on a fixed proportion of the search vector to achieve convergence. In the implementation in GULP, this step magnitude is correlated to the gradient norm, so that as the calculation approaches convergence, and therefore enters the harmonic region, the proportion of the search vector used approaches 100%. Hessian updating can also be applied while optimizing towards a transition state. However, because the rate of convergence depends more critically on the inverse Hessian being accurate, more frequent exact calculations are usually required. RFO can also be used to accelerate optimizations to minima once the harmonic region is approached, since it is able to better handle soft modes. In extreme cases, the number of cycles is reduced by an order of magnitude by use of RFO towards the end of an optimization, as well confirming that the Hessian has the required structure. A final note of caution is that this method does not guarantee that a local minimum has been achieved, even if the Hessian is confirmed as having all positive eigenvalues. This is because the Hessian only spans the space of the allowed optimization variables and there may be directions of negative curvature with respect to constrained degrees of freedom. Hence, a Γ-point phonon calculation should always be used as a final validation.

Often it is desirable to apply external forces to a system when performing a minimization. The most common example is the application of an isotropic external pressure, such as used in the determination of the equation of state for a material. This situation can be straightforwardly handled by making the objective function for minimization the enthalpy instead of the internal energy:

$$H = U + pV.$$

More complex is the case when an external force is to be applied to individual atoms. This arises when the program is being coupled to a separate quantum mechanical calculation as part of a QM/MM method (i.e. a quantum mechanics/molecular mechanics coupled calculation). Here the forces on the forcefield atoms arise from the quantum mechanical region, which is not explicitly considered in the present calculation. Adding the negative of the external forces to the internal derivatives is simple to implement. However, for minimization the internal energy is no longer a well-behaved objective function (i.e. the minimum of the internal energy does not correspond to a configuration of zero gradients). One solution to this would be to minimize the gradient norm instead, but this requires either one order higher of energy derivatives or that conjugate gradients are used to minimize, which is less efficient.

The approach implemented within GULP to handle the application of external forces is to define a new additional term in the internal energy, which is the work due to the external force:

$$U^{\text{external}} = -\int_{\text{initial}}^{\text{final}} F^{\text{external}} \alpha,$$

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

where α represents the vector of atomic coordinates and F^{external} is the vector of external forces. With this correction to the internal energy, the minimum coincides with force balance and so the full range of standard minimizers are applicable. However, the final energy is an arbitrary quantity unless taken with respect to a standard reference set of atomic positions.

The framework of structural optimization in GULP is designed to be flexible and to allow the user as much freedom as possible to control the degrees of freedom. Intrinsically, there are four main classes of variables possible within the forcefield models available, namely, atomic coordinates, the unit cell periodicity (for polymers, surfaces, and bulk materials), shell coordinates, and breathing shell radii. Each variable can be controlled by optionally specifying a binary flag that indicates whether it is to be varied or held fixed. Alternatively, there are keywords that select default settings of the flags, such that those quantities not constrained by symmetry are allowed to vary. In particular, the defaults automatically handle the fact that one atom must be fixed in order to remove the three translational degrees of freedom. For bulk systems, the first atom is usually the one chosen to be fixed, since the choice is arbitrary. In the case of slab calculations, the atom nearest the center of the slab is fixed since this is the one least likely to move by a substantial amount. When there is a center of symmetry in the space group, then there is no need to fix any atoms since the origin is already immoveable. The default optimization options also automatically construct the constraints necessary to preserve the space group symmetry as the atoms are displaced.

When minimizing periodic systems there are a couple of choices that have to be made. The first concerns the representation of the unit cell. This can be achieved through either the unit cell parameters, a, b, c, α, β, γ, or via the Cartesian cell vectors as a (3×3) matrix. Correspondingly, when optimizing the unit cell this can be done with respect to the cell parameters directly, the Cartesian cell vectors or by using a strain tensor. Here we chose to use the Voight strain tensor to scale the Cartesian lattice vectors:

$$\begin{vmatrix} 1+\epsilon_1 & \frac{1}{2}\epsilon_6 & \frac{1}{2}\epsilon_5 \\ \frac{1}{2}\epsilon_6 & 1+\epsilon_2 & \frac{1}{2}\epsilon_4 \\ \frac{1}{2}\epsilon_5 & \frac{1}{2}\epsilon_4 & 1+\epsilon_3 \end{vmatrix}.$$

Through using the six Voight strains as the variables, we have the same advantage as working with the unit cell parameters in that cell rotation is eliminated by construction. However, as will be demonstrated later, the derivatives with respect to strain are closely related to the Cartesian internal derivatives and therefore readily calculated at almost no extra cost. We should note that when strain derivatives are taken, they are done so with respect to infinitesimal strain about the current configuration, rather than with respect to a single constant reference cell. For 2-D systems, the strain tensor is correspondingly reduced to a (2×2) matrix and for polymers there is a single scalar strain value. At this point we should also comment on the absolute spatial orientation of systems, since this a matter of convention, rather than an absolute choice. For 3-D systems, GULP orients the a cell vector along the x-axis, the b cell vector in the xy-plane, and then the orientation of the c cell vector is fixed, except for an issue of chirality, which is resolved by selecting the orientation that has a positive dot product with the z-axis. For 2-D systems, the same two conventions are followed for the surface vectors, which leaves the surface normal parallel to the z-axis. Finally, for 1-D systems, the repeat direction is taken to lie along the x-axis.

Genetic Algorithms

The approach to minimization of a structure has already been discussed when considering the case of a local stationary point. Now we turn to consider the quest for a global minimum. In order to do this, a technique is required that allows barriers to be overcome in some way. While hyperdynamic [81] and basin-filling approaches are becoming increasingly useful, the more traditional approach has been to employ methods that incorporate random changes in the structure with a tendency to minimize a so-called cost function. Here the cost function is a quantity that represents a measure of the quality of the structure. In the most obvious case, this might be just the total energy of the system. However, because stochastic processes require many moves, it is often convenient to use a more computationally efficient cost function. For example, one form that has previously been successfully used for oxides is a combination of bond-valence sum rules with a short-ranged Coulomb penalty that prevents ions of like charge approaching each other [82].

There are two main approaches to the global optimization of structures via a random walk—the Monte Carlo method [83] and GAs [16,17]. Although both are in principle equally capable of performing the task, it could be argued that GAs are more naturally suited to coarse grained global optimization, while the Monte Carlo technique is also able to yield an integration within the phase space of a statistical ensemble. Here we will concentrate on the use of GAs.

In the GA, each optimization variable is discretized within a specified range. This is particularly well suited to fractional coordinates where there are natural bounds of 0 and 1. Then all that needs to be specified is

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

the number of allowed intervals within the range. If a coarse grid of points is specified then configuration space will be rapidly searched, but the solution will be very approximate and the global minimum may not be resolvable with the level of resolution available. Conversely, if a fine grid is utilized then the search is capable of locating a more accurate estimate of the global minimum. However, because of the large number of possible configurations, the optimal state may never be visited. Clearly, a hierarchical approach may be ideal, where a coarse discretization is initially used and then refined as the process progresses. The state of each configuration is represented by a binary number which is a concatenation of the binary representations of the grid point numbers for each individual variable.

The fundamental idea behind GAs is to perform a process that mimics natural selection through survival of the fittest. Here the concept of fitness equates to the value of the chosen cost function. Each cycle of the GA begins with an even number of so-called parents. In the first step these are typically randomly chosen initial configurations. These parents then are randomly paired up in order to breed to produce two children in order to maintain the size of the population. This is the tournament phase, in which a probability is set for the parent with the lower cost function to go through to the next generation and to form the children. A random number is chosen and compared to this probability in order to make the decision. Choosing this probability threshold is a balance between wanting to ensure that the better configurations survive against trying to maintain diversity in the population during the initial stages. More sophisticated tournament procedures have also been proposed where an exponential weighting is used in favor of the fitter candidates. Usually, the number of children is set equal to the number of parents in order to maintain a stable population size.

Apart from the tournament phase, there are two other processes that can occur, which are mutation and crossover. In the mutation phase, a number of binary digits are changed from 0 to 1 or *vice versa*. Again the chance of this occurring is determined by comparison of a random number to the mutation probability. During crossover, two configurations are selected, the binary strings split at a random point, and the two parts switched between configurations. Both mutation and crossover represent genetic defects that are designed to introduce randomness with the aim of exploring configuration space.

The above three GA steps are applied repeatedly with the aim of gradually reducing the cost functions of the population down and focusing in on the global minimum. When the procedure is stopped after the specified number of steps, then hopefully the configuration with the lowest cost function will be the global minimum, though this can never be known for certain for complex systems. If the cost function used was an approximation to the total energy that is ultimately desired to be the criterion for selection, then a selection of the best configurations can be subsequently minimized according to the total energy and hopefully the global minimum will lie amongst this set, even though it is not the structure with the absolutely lowest cost function.

There are several possible refinements to the GA procedure, which, somewhat appropriately, is itself still evolving. However, the above basic formulation is capable of correctly predicting the atomic coordinates of binary, and even some ternary oxides, given the unit cell. Indeed, the method had an early success through assisting in the solution of the previously unknown structure of Li_3RuO_4 [75]. More details of the background to, and application of, the GA for solid state structural optimization can be found elsewhere [82].

Calculation of Bulk Properties

Having determined the optimized structure for a material it is then possible to calculate a wide range of physical properties based on the curvature of the energy surface about the minimum. These include both mechanical properties, such as the bulk modulus and elastic constants, as well as dielectric properties. The expressions used for the calculation of the individual properties that may be determined routinely from the second derivative matrix are detailed below.

Elastic Constants

The elastic constants represent the second derivatives of the energy density with respect to strain:

$$C_{ij} = \frac{1}{V}\left(\frac{\partial^2 U}{\partial \epsilon_i \partial \epsilon_j}\right),$$

thereby describing the mechanical hardness of the material with respect to deformation. Since there are six possible strains within the notation scheme we employ, the elastic constant tensor is a 6×6 symmetric matrix. The 21 potentially independent matrix elements are usually reduced considerably by symmetry [84]. For example, for a cubic material the only unique elements are C_{11}, C_{12} and C_{44}. The calculation of elastic constants is potentially very useful, since the full tensor has only been measured experimentally for a very small percentage of all known solids. This is primarily because the practical determination typically requires single crystals with a size of a few micrometers at least.

In calculating the second derivatives of the energy with respect to strain it is important to allow for the response of all internal degrees of

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

freedom of the crystal to the perturbation. If we introduce the following notation for the second derivative matrices:

$$D_{\epsilon\epsilon} = \left(\frac{\partial^2 U}{\partial \boldsymbol{\epsilon} \partial \boldsymbol{\epsilon}}\right)_{\text{internal}}$$

$$D_{\epsilon i} = \left(\frac{\partial^2 U}{\partial \boldsymbol{\epsilon} \partial \alpha_i}\right)_{\epsilon}$$

$$D_{ij} = \left(\frac{\partial^2 U}{\partial \alpha_i \partial \beta_j}\right)_{\epsilon},$$

then the full expression for the elastic constant tensor can be written as:

$$C = \frac{1}{V}(D_{\epsilon\epsilon} - D_{\epsilon i} D_{ij}^{-1} D_{j\epsilon}).$$

The elastic compliances, S, can be readily calculated from the above expression by inverting the matrix (i.e. $S = C^{-1}$). The above expressions are valid for adiabatic conditions and at zero pressure. The adiabatic elastic constants can be extended to the case of a finite external pressure, p, according to the corrections:

$$C_{\alpha\beta\gamma\zeta} = \frac{1}{V}\left(\frac{\partial^2 U}{\partial \boldsymbol{\epsilon}_{\alpha\beta} \partial \boldsymbol{\epsilon}_{\gamma\zeta}}\right) + \frac{p}{2}(2\delta_{\alpha\beta}\delta_{\gamma\zeta} - \delta_{\alpha\zeta}\delta_{\beta\gamma} - \delta_{\alpha\gamma}\delta_{\beta\zeta}).$$

Here the strains are expressed directly in terms of the Cartesian components, as opposed to in the Voight notation, so as to make the corrections more transparent.

Under isothermal conditions the expressions for the elastic constant tensor are analogous, except for the fact that the differentials are with respect to the Helmholtz free energy, rather than the internal energy. Because the evaluation of the isothermal elastic constants requires the fourth derivative of the internal energy with respect to internal coordinates and strains, this is not presently implemented within GULP analytically. However, finite differences may be used in combination with free energy minimization in order to determine this tensor.

Bulk and Shear Moduli

Like the elastic constant tensor, the bulk (K) and shear (G) moduli contain information regarding the hardness of a material with respect to various types of deformation. Experimentally, a bulk modulus is much more facile to determine than the elastic constant tensor. If the structure of a material is studied as a function of applied isotropic pressure, then a plot of pressure versus volume can be fitted to an equation of state where the bulk modulus is one of the curve parameters. Typically, a third or fourth order Birch–Murgnahan equation of state is utilized. Alternatively, the bulk and shear moduli are also clearly related to the elements of the elastic constant. However, there is no unique definition of this transformation. Here we give three different definitions due to Reuss, Voight and Hill [84]. Below are the equations for the Reuss and Voight definitions, while the Hill values are defined as the average of the other two:

$$K_{\text{Voight}} = \frac{1}{9}(C_{11} + C_{22} + C_{33} + 2(C_{12} + C_{13} + C_{23}))$$

$$K_{\text{Reuss}} = (S_{11} + S_{22} + S_{33} + 2(S_{12} + S_{13} + S_{23}))^{-1}$$

$$G_{\text{Voight}} = \frac{1}{15}(C_{11} + C_{22} + C_{33} + 3(C_{44} + C_{55} + C_{66}) - C_{12} - C_{13} - C_{23}).$$

$$G_{\text{Reuss}} = \frac{15}{4(S_{11}+S_{22}+S_{33}-S_{12}-S_{13}-S_{23})+3(S_{44}+S_{55}+S_{66})}$$

Young's Moduli

When a uniaxial tension is applied to a material then the lengthening of the material is measured according to the strain. The ratio of stress to strain defines the value of the Young's modulus for that axis:

$$Y_\alpha = \frac{\sigma_{\alpha\alpha}}{\boldsymbol{\epsilon}_{\alpha\alpha}}.$$

Since a material will always increase in length under tension, the value of this quantity should always be positive. The Young's moduli in each of the Cartesian directions can be calculated from the elastic compliances:

$$Y_x = S_{11}^{-1}$$

$$Y_y = S_{22}^{-1}$$

$$Y_z = S_{33}^{-1}.$$

Poisson's Ratio

Complementary to Young's modulus is the Poisson ratio, which measures the change in a material at right angles to the uniaxial stress. Formally, it is defined as the ratio of lateral to longitudinal strain under a uniform, uniaxial stress. The expression used to calculate this property, assuming an isotropic medium, is given below [84]:

$$\sigma_\alpha(\beta) = -S_{\alpha\alpha\beta\beta} Y_\beta.$$

Because most materials naturally shrink orthogonal to an applied tension this leads to positive values for Poisson's ratio, with a theoretical maximum of 0.5.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

Typical values for many materials lie in the range 0.2–0.3, though negative values are also known. For an isotropic material this quantity can also be related to the bulk modulus:

$$K = \frac{1}{3}\frac{Y}{(1-2\sigma)}.$$

Acoustic Velocities

The acoustic velocities are key quantities in the interpretation of seismic data. The polycrystalline averages of these acoustic velocities in a solid can be derived from the bulk and shear moduli of the material, as well as the density, ρ. There are two values, that for a transverse wave, V_s, and that for a longitudinal wave, V_p, which are given by:

$$V_s = \sqrt{\frac{G}{\rho}}$$

$$V_p = \sqrt{\frac{4G+3K}{3\rho}}.$$

As to be expected, there is some degree of variability in the calculated values according to the definition of the bulk and shear moduli employed.

Static and High Frequency Dielectric Constants

The dielectric properties of a material are of crucial importance in many contexts, including those beyond the strictly bulk properties. For example, the response of a solid to a charged defect depends on the inverse of the dielectric constant. The actual value of the dielectric constant varies according to the frequency of the electromagnetic field applied. Commonly, two extreme values are quoted, namely the static and high frequency dielectric constants. In the static limit all degrees of freedom of the crystal, both nuclear and electronic, are able to respond to the electric field and therefore to provide screening. At the high frequency limit the oscillation is greater than the maximum vibrational frequency of the material and therefore only the electrons are able to respond to the perturbation fast enough. Here we describe how to calculate these extremal values, while the case of intermediate frequencies will be presented later.

The static dielectric constant (3×3) tensor can be determined from the Cartesian second derivative matrix of all particles, $D_{\alpha\beta}$, and the vector, q, containing the charges of all particles:

$$\varepsilon^0_{\alpha\beta} = \delta_{\alpha\beta} + \frac{4\pi}{V}(qD^{-1}_{\alpha\beta}q).$$

The expression for the high frequency dielectric constant is identical to that for the static equivalent, except that the second derivative matrix, $D_{\alpha\beta}$, now only includes the Cartesian components for any shells present within the model. If a core only model is being used then the high frequency dielectric tensor is just a unit matrix. Hence, information regarding the high frequency dielectric constants is particularly useful in determining the parameters of a shell model due to the relatively direct correlation.

Because the dielectric constant tensor depends on the inverse second derivative matrix, it has many of the characteristics of the Hessian matrix and is therefore quite a sensitive indicator of whether a potential model is sensible. Extreme values, particularly negative ones, instantly point to the fact that the potential model is inadequate or that the system wishes to undergo a symmetry change.

Refractive Indices

The refractive indices of a material, n, are simply related to the dielectric constant by:

$$n = \sqrt{\varepsilon}.$$

For orthorhombic, tetragonal or cubic unit cells, in the standard orientation, this represents a trivial mapping since the dielectric constant tensor is a diagonal matrix and so the square roots can be taken directly. For other crystal systems it is necessary to diagonalize the tensor first and then find the refractive indices in this new axis eigensystem.

Piezoelectric Constants

The piezoelectric constants are key quantities in many technological applications, since they govern the correlation between strain in a material and applied electric field for non-centrosymmetric materials (centrosymmetric materials necessarily have zero for all piezoelectric constants). There are several different types of piezoelectric constant too, depending on whether it is the polarization being induced by a given strain that it is being considered or the stress/strain induced by an applied electric field. In GULP, both the piezoelectric stress constants, d, and the piezoelectric strain constants, e, are calculated:

$$d_{\alpha i} = \frac{\partial P_\alpha}{\partial \sigma_i}$$

$$e_{\alpha i} = \frac{\partial P_\alpha}{\partial \epsilon_i}.$$

The two sets are related by a transformation involving either the elastic constant, or the elastic compliance tensor, depending on the direction. The piezoelectric strain constants are calculated from the Cartesian second derivative matrices

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

according to:

$$e_{\alpha i} = -\sum_{k=1}^{N-1}\left[\frac{4\pi}{V}q_k\left(\frac{\partial^2 U}{\partial\alpha\partial\beta}\right)^{-1}\left(\frac{\partial^2 U}{\partial\beta\partial\epsilon_i}\right)\right].$$

The above piezoelectric strain constants can be readily transformed into piezoelectric stress constants by multiplication by the elastic compliance tensor.

Electrostatic Potential, Electric Field and Electric Field Gradients

The electrostatic site potential is a measure of the Coulomb interaction per unit charge experienced by an ion at a given position in space. This can be formally defined by;

$$V_i = \sum_{j=1}^{N}\frac{q_j}{r_{ij}}$$

where the summation explicitly excludes the case where $i = j$. Clearly, this is closely defined to the definition of the electrostatic energy, the two simply being related by;

$$U^{\text{electrostatic}} = \frac{1}{2}\sum_{i=1}^{N} q_i V_i.$$

Similar considerations also apply with regard to periodic systems (i.e. an appropriate charge summation technique must be utilized with a net charge and dipole of zero).

The electrostatic potential is a useful quantity for qualitative predictions as to certain properties of a material. For instance, if a structure contains several distinct oxygen sites one might expect that there would be a correlation between site potential and basicity. However, the limitations of the point ion, or dipolar shell, model must be kept in mind. Certainly, the absolute site potential will be totally dependent on the choice of charges in the potential model, and therefore will be arbitrary, so at best only trends should be considered.

Based on the calculation of the site potential, the electric field, and the electric field gradient may also be determined as the first and second derivatives, respectively, of this quantity, while again excluding the case where $i = j$;

$$\frac{\partial V_i}{\partial\alpha} = \sum_{j=1}^{N} -\frac{q_j}{r_{ij}^3}\alpha_{ij}$$

$$\frac{\partial^2 V_i}{\partial\alpha\partial\beta} = \sum_{j=1}^{N}\frac{\left(3\alpha_{ij}\beta_{ij} - r_{ij}^2\delta_{\alpha\beta}\right)}{r_{ij}^5}.$$

The electric field gradient tensor is perhaps the most useful quantity since it may be transformed into the so-called asymmetry parameter, η, which can be important in the interpretation of solid state NMR data [85]. If we represent the tensorial components of the electric field gradient as $V_{\alpha\beta}$, then first the (3×3) matrix can be diagonalized to yield the three principal axis components, V_{xx}, V_{yy} and V_{zz}. If the components are labeled V_1, V_2 and V_3, such that $|V_3| > |V_2| > |V_1|$, then the asymmetry parameter is defined as;

$$\eta = \frac{|V_2| - |V_1|}{|V_3|},$$

which ensures that the values lie in the range 0–1. A high symmetry site, such as one of O_h point group where all components are equal, would therefore have an asymmetry of zero.

Born Effective Charges

The quantification of the charges associated with atoms/ions is one of the most problematic tasks in theoretical chemistry. Given that the atomic charge is not a direct quantum mechanical observable, it is therefore an artificial quantity, but one that is nonetheless useful in the development of understanding from quantitative results. Not surprisingly there are many different definitions of charge, the most famous of which is that due to Mulliken [37], where overlap density is apportioned equally to both nuclear centers. More recently, there has been considerable interest in Bader analysis [86] as a means of partitioning electron density. Both of these aforementioned schemes are based on the analysis of the density matrix, or the density itself. However, there is another family of charge definitions, which define the atomic charges in terms of the response of the dipole moment of the system with respect to perturbations. Since the dipole is a genuine observable, such charge definitions are particularly attractive for forcefields as they describe the charges that are consistent with the response of the material to atomic displacements.

The most widely used definition in this second category is the Born effective charge:

$$q_i^{\text{Born}} = \frac{\partial\mu^{\alpha}}{\partial\beta_i}.$$

Here the charge of an atom is now a symmetric (3×3) tensorial quantity since it describes the derivative of the three Cartesian components of the dipole moment with respect to the three Cartesian atomic displacements. While the definition of the dipole moment is complicated by the choice of the particular images of each ion, this complication does not affect the derivatives of the dipole moment within the present model. Differentiation of the dipole moment leads to the following expression for the Born effective

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

charge;

$$q_i^{\text{Born}} = q_i^{\text{core}}\delta_{\alpha\beta} - (D_{\text{core-shell}}\, D^{-1}_{\text{shell-shell}}\, q^{\text{shell}})_i,$$

where the first term on the right-hand side is the core charge of the ion, and the second term is the corresponding component of the product of the core–shell second derivative matrix with the inverse of the shell–shell second derivative matrix, scaled by the vector of shell charges, q^{shell}. Physically, the second term corresponds to the response of all the shells present to the atomic displacement of atom i, or in other words, the electronic contribution. Hence, for a rigid ion model, the Born effective charge tensor is equal to a diagonal matrix with all the diagonal elements equal to the core charge. Increasingly, the Born effective charges are being determined from *ab initio* calculations, thus creating a new avenue for determining shell model parameters beyond the fitting of dielectric constants. Beyond this, the Born effective charges are also especially important in the determination of Γ-point phonon frequencies, as will be discussed in the next section.

Phonons

Atoms must constantly be in motion as a consequence of the Heisenberg uncertainty principle and this is achieved through vibrations. At low temperatures the vibrations correspond to simple harmonic motion about the position of minimum energy, while as the temperature increases they become increasingly anharmonic. For a molecule, there will be $3N - 6$ vibrational modes (or $3N - 5$ for a linear system). In the case of an infinitely perfect 3D solid, there will be a corresponding infinite number of phonons. These phonons are described by calculating their values at points in reciprocal space, usually within the first Brillouin zone. Hence, we will obtain $3N$ phonons per k-point. The lowest three modes represent the so-called acoustic branch which tend to values of zero at the center of the Brillouin zone ($k = 0, 0, 0$), known as the Γ-point. At this point, the acoustic modes correspond to the pure translation of the crystal lattice, and thus they are modes of zero frequency. A plot of the vibrational frequencies versus k gives rise to phonon dispersion curves.

To calculate the vibrations or phonons of a system, the starting point is the force constant matrix, given by the second derivatives with respect to the atoms in Cartesian space. In the case of a solid, the terms must be multiplied by the corresponding phase factor, $\exp(ik\,r)$. Thus, the force constant matrix, F, between two atoms i and j is given by;

$$F_{i\alpha j\beta}(k) = \sum_R \left(\frac{\partial^2 U}{\partial\alpha\partial\beta}\right)\exp(ik(r_{ij} + R)).$$

The summation over R represents the sum over lattice vectors within the cut-off radius. This matrix of force constants is then converted to the dynamical matrix, D, by multiplying by the inverse square root masses of the ions:

$$D_{i\alpha j\beta}(k) = \frac{1}{(m_i m_j)^{\frac{1}{2}}} F_{i\alpha j\beta}(k).$$

The origin of the three acoustic phonons at the Γ-point for a solid, or in the regular vibrational spectrum for a molecule, arises from the sum rules that govern the energy derivatives. Firstly, the sum of all first derivatives, or alternatively forces, must be equal to zero in the absence of an external force:

$$\sum_{i=1}^{N}\left(\frac{\partial U}{\partial\alpha_i}\right) = 0.$$

Secondly, by further differentiating the above expression, it can be shown that the on-diagonal elements of the force constant matrix are equal to the negative sum of the off-diagonal elements:

$$\left(\frac{\partial^2 U}{\partial\alpha_i\partial\beta_i}\right) = -\sum_{j=1}^{N\prime}\left(\frac{\partial^2 U}{\partial\alpha_i\partial\beta_j}\right),$$

where the summation now excludes the case when $i = j$. It should be noted that if a phonon calculation is performed for a surface using a two-region strategy, as will be discussed later, then there will no longer be three modes of zero frequency at the zone-center. This is because the influence of region 2 acts as an external force, which breaks the translational invariance.

In forming the dynamical matrix from the force constant matrix there are a number of issues relating to particular forcefields and system types. The most common issue relates to the use of the shell model, in which the shells have zero mass. Since the number of vibrational modes is strictly given by three times the number of atoms, which in this case also corresponds the number of cores, then the shell co-ordinates cannot appear directly in the dynamical matrix. Instead, the shell contributions are factored into the force constants of the cores according to the expression:

$$F_{\text{cc}}^{\text{total}} = F_{\text{cc}} - F_{\text{cs}}F_{\text{ss}}^{-1}F_{\text{sc}},$$

where F_{cc}, F_{ss} and F_{sc} are the core–core, shell–shell and shell–core force constant matrices, respectively, and F_{cs} is just the transpose of F_{sc}. In the case of the breathing shell model, the shell index is extended to run over the shell radius, as well as its Cartesian co-ordinates.

At the Γ-point there is an extra complication in the calculation of the phonons. In materials where the atoms possess a charge, the degeneracy of the transverse optical (TO) and longitudinal optical

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

(LO) modes is broken due to the electric field that is generated during vibration. This effect is not allowed for in the usual analytic evaluation of the dynamical matrix. Furthermore, the precise splitting that occurs also depends on the direction of approach to the Γ-point in reciprocal space, k^{Γ}(i.e. the LO and TO modes are likely to be discontinuous at this point in phonon dispersion plots). If the Born effective charges are known then this non-analytic term [87] can be corrected for by adding a correction to the dynamical matrix [38]:

$$D^{na}_{i\alpha j\beta} = \frac{4\pi}{V\left(m_i m_j\right)^{\frac{1}{2}}} \frac{\left(k^{\Gamma} q_i^{\text{born}}\right)_{\alpha} \left(k^{\Gamma} q_j^{\text{born}}\right)_{\beta}}{\left(k^{\Gamma} \epsilon^{\infty} k^{\Gamma}\right)_{\alpha\beta}},$$

where terms have the meanings as previously defined. As to be expected, the influence of the charges on the splitting is mediated in inverse proportion to the high frequency dielectric constant tensor.

Two scenarios arise with regard to the calculation of the Γ-point phonons. If the value for a specific direction of approach is required, such as in the case of a phonon dispersion curve, then the value of k^{Γ}may be explicitly specified. The direction of approach is automatically set equal to the direction of the phonon dispersion curve when the Γ-point is part of one in GULP. Alternatively, if the intention is to compute the infra-red spectrum as a polycrystalline average, then the phonons at this point should be a spherical average over all possible orientations. To allow for this last possibility, there is the possibility within GULP to perform a numerical integration by sampling the phonon modes as a function of θ and ϕ in spherical polar coordinates and then averaging the resulting frequencies.

Vibrational Partition Function

Statistical mechanics allows the connection to be made between microscopic quantum energy levels and macroscopic thermodynamic observables [88]. Pivotal to this process is the partition function, Z^{tot}, for the system. In order to determine this quantity we make the assumption that all forms of energy are independent and therefore that the total energy is just the sum of contributions from translation, rotation, vibration and the electronic state. In the case of a solid, we can neglect the translational and rotational components, and furthermore since we are considering forcefield methods the electronic energy is not directly calculated. Hence, the problem reduces down to one of determining the vibration energy levels, and subsequently the corresponding partition function.

The vibrational energy levels for a harmonic oscillator for the mth mode are simply given by;

$$U^{\text{vib}}_m(n,k) = \left(n + \frac{1}{2}\right) h\omega(m,k),$$

if the frequency is specified in Hz. By summing over all energy levels it is possible to show that the vibrational partition function is equal to;

$$Z^{\text{vib}} = \sum_m \sum_k \frac{\exp\left(-\frac{h\omega}{2k_B T}\right)}{\left(1 - \exp\left(-\frac{h\omega}{k_B T}\right)\right)}.$$

Here, and subsequently, the explicit dependence of the frequency on mode and k-point has been dropped for brevity. Substituting this expression into the various statistical mechanical expressions for thermodynamic quantities, we obtain the following relationships:

$$U^{\text{vib}} = \sum_m \sum_k w_k \left(\frac{1}{2} h\omega + \frac{h\omega}{\left(\exp\left(-\frac{h\omega}{k_B T}\right) - 1\right)}\right)$$

$$A^{\text{vib}} = \sum_m \sum_k w_k \left(\frac{1}{2} h\omega + k_B T \ln\left(1 - \exp\left(-\frac{h\omega}{k_B T}\right)\right)\right)$$

$$C_{\text{v}}^{\text{vib}} = \sum_m \sum_k w_k k_B \left(\frac{h\omega}{k_B T}\right)^2 \frac{\exp\left(-\frac{h\omega}{k_B T}\right)}{\left(\exp\left(-\frac{h\omega}{k_B T}\right) - 1\right)^2}.$$

Note that the first term in the internal energy is just the zero point vibrational energy. In the above equations, a sum over points in the Brillouin zone is included and the term w_k represents the weight of that particular point, such that the sum of all weights is equal to one. Formally speaking, this sum should be an integration over the phonon density of states. In practice, a discrete sum of points is typically used to numerically integrate the quantities.

There are a number of different choices possible for the set of points in the Brillouin zone for integration of the thermodynamic properties, just as there are for integration of the band structure in an electronic calculation. For high symmetry materials there are occasionally special point(s), such as those due to Baldereschi [89] and Chadi and Cohen [90]. However, the most appropriate in general is usually the Monkhorst–Pack scheme [91]. This involves an evenly spaced mesh of k-points, given by three so-called shrinking factors, one along each axis of the Brillouin zone. This still leaves the issue of the translational position of the mesh relative to the origin. The optimal choice is to maximize the distance of all points from the origin along each axis by using an offset of half a mesh spacing from there to the first mesh point. The benefits of this choice are that it increases the rate of convergence as a function of increasing shrinking factor, as well as avoiding the issue of the non-analytic correction to the LO/TO splitting at the Γ-point.

If the space group symmetry has been specified for the material, then it is possible to determine

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

the Patterson group, which is the equivalent of the space group in reciprocal space. Note that all systems have at least inversion symmetry (also known as time-reversal symmetry) in the Brillouin zone—a fact that is always used to accelerate the Ewald sum. By using the Patterson group, it is only necessary to sum over *k*-points that lie within the asymmetric wedge of the Brillouin zone, with appropriate weighting of points situated on the boundary [92]. This can lead to substantial computational savings for systems with small unit cells (and therefore high shrinking factors) and high symmetry.

When calculating thermodynamic quantities it is important to ensure convergence to the desired precision with respect to the shrinking factors. Because the mesh is in reciprocal space, the shrinking factor required for a given precision is inversely proportional to the unit cell length along that axis.

Frequency-dependent Dielectric Constants and Reflectivity

We have already seen that the limiting dielectric properties in the static and high-frequency limits can be readily calculated for a system. However, the dielectric constant may also be calculated as an explicit function of the frequency of applied field, ω_f. In order to determine this, first it is necessary to calculate the oscillator strength, Ω, for each vibrational mode based on the Born effective charges and the eigenvector, e, for that mode:

$$\Omega_{\alpha\beta} = \left(\sum_{i=1}^{N} \frac{q_{i\alpha\gamma}^{\text{Born}} e_{i\gamma}}{m_i^{\frac{1}{2}}} \right) \left(\sum_{i=1}^{N} \frac{q_{i\beta\gamma}^{\text{Born}} e_{i\gamma}}{m_i^{\frac{1}{2}}} \right).$$

The dielectric constant at the applied field is then given by:

$$\epsilon_{\alpha\beta}(\omega_f) = \epsilon_{\alpha\beta}(\infty) + \frac{4\pi}{V} \left(\sum_{m=1}^{\text{modes}} \frac{\Omega_{\alpha\beta}^{m}}{\left(\omega_m^2 - \omega_f^2 \right)} \right).$$

From the form of the above relationship, it can be readily seen that there is a singularity in the dielectric constant when the applied frequency exactly matches that of one of the vibrational modes of the system. The limiting behavior of the above expression is such that the results are equivalent to the previously calculated values for the dielectric constant as the frequency of the applied field tends to both zero and infinity.

One further property can be readily extracted from the above data, which is the bulk reflectivity. This measures the ratio of the reflected to absorbed power at a given frequency. While the frequency-dependent dielectric constant is a tensor, the reflectivity, R, has a value for each specific direction of incidence, k^{in} and reflection, k^{out}, typically specified as vectors in reciprocal space:

$$R(\omega_f) = \left(\frac{\sqrt{\varepsilon(\omega_f)} - 1}{\sqrt{\varepsilon(\omega_f)} + 1} \right)^2$$

$$\varepsilon(\omega_f) = k^{\text{in}} \epsilon(\omega_f) k^{\text{out}}.$$

Calculation of Surface Properties

The properties of the surfaces of materials are every bit as important as the bulk properties, since they control the interaction between the substance and the external environment. At the most obvious level, the very shape of the particles or crystallites formed is determined by the properties of the surface relative to the bulk, while catalysis and reactions of the material also predominantly occur at the surface. From the earlier discussion of electrostatics, it is apparent that the surface structure is a key factor in determining the bulk polarization and net dipole of the material, which has consequences even for the bulk region.

Due to the interest in surface phenomena, the history of modeling surfaces using atomistic simulation is also a long one spanning a number of different computer codes. Of the recent era, the dominant computer code for surface simulation of ionic materials has been MARVIN [93], which has been developed alongside GULP for many years. However, to ensure the maximum integration of functionality between bulk and surface simulations, both in terms of forcefield models, as well as accessible properties, it was decided to incorporate much of the functionality of MARVIN into GULP to produce a single code capable of simulating systems with any type of boundary condition from zero-dimensional (0-D), through to 3-D. Here we describe the methodology employed for surface calculations.

There are two phases to any surface calculation, namely the creation of the surface from the bulk material and the subsequent calculation of its optimized structure and properties. Each surface is specified by at least two pieces of information. Firstly, there are the Miller indices (*hkl*) of the plane that defines the orientation of the bulk cleavage. Secondly, there is the so-called shift—i.e. the displacement of the plane relative to the unit cell origin. For simple cases, such as the (0 0 1) surface of a rock salt structure there is only one unique choice of shift. However, for more complex cases there may be several shifts for a given plane that lead to distinct surfaces. When cleaving surfaces there are also other important considerations to take into account, in particular the type of surface. As shown in Fig. 1, there are three basic types of surface. In type 1, the atomic structure consists of charge neutral sheets of ions parallel to the surface plane and thus all shifts

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

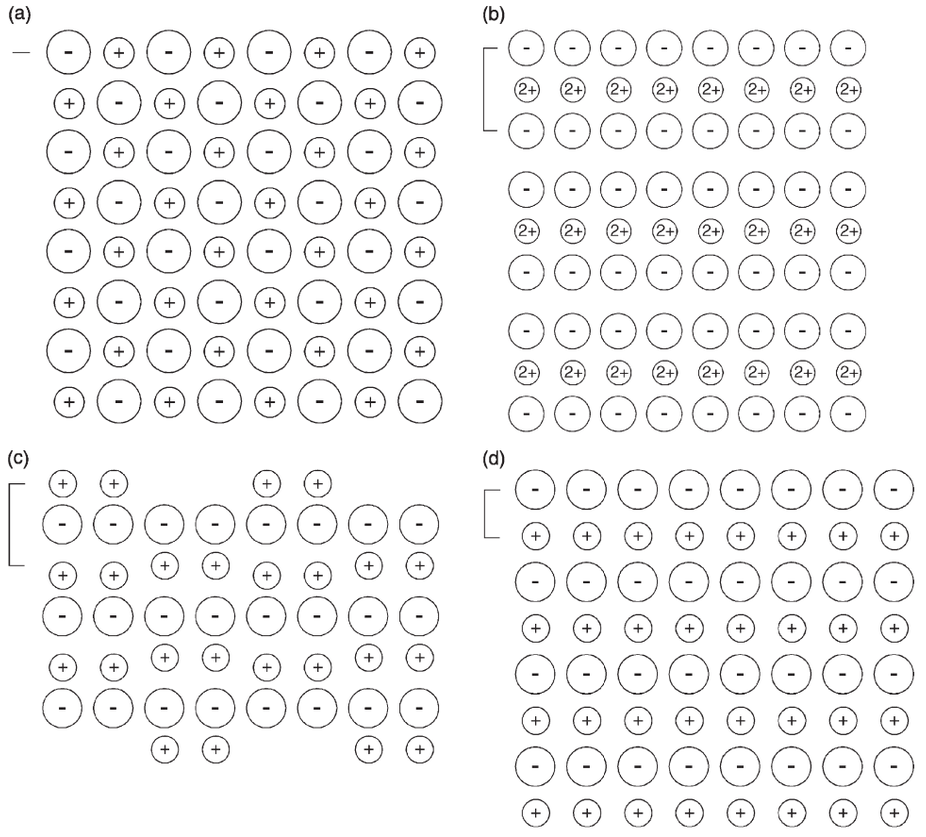


FIGURE 1 The three types of surface as categorized by Tasker (a) type I, where each layer consists of coplanar anions and cations, and all planar cuts lead to a non-polar surface; (b) type IIa, where the anions and cations comprising the layers are not coplanar, but which allows for some surface cuts to split the layers in such a way as to produce no dipole; (c) type IIb, which is as per type IIa, except that some ions must be moved from the surface to the bottom of region 2 in order to achieve zero dipole; and (d) type III, where there are alternating layers of cations and anions, and all possible planar cuts result in a surface with a dipole moment.

are guaranteed to yield a surface with no dipole moment in the direction of the surface normal. For type 2 surfaces, there are combinations of layers of cations and anions that possess zero net dipole in the appropriate direction. Hence, for well-chosen values of the shift a non-dipolar surface can be obtained. Finally, for type 3, all cleavage planes result in a dipolar surface, which is therefore likely to be less stable. While dipolar surfaces can exist, this normally leads to twinning of crystals or strong solvation in order to anul the dipole. Alternatively, the surfaces can reconstruct in order to remove the dipole. This typically involves the creation of cation or anion vacancies at the surface or chemical modification. When an ion is conceptually removed, in practice it is moved to the bottom of the surface slab in a calculation in order to maintain charge neutrality [94]. Real examples of the three types of surface are presented for polymorphs of calcium carbonate in Fig. 2.

From the above, it can be seen that often the creation of the surface is a significant task in its own right and can be a complex process. In previous programs for surface calculations, the structure manipulation has usually been performed via the input deck of the code. Clearly, when reconstructions are involved this can become rather unwieldy. As a result, a different strategy has been adopted for use with GULP. All construction of the surface and structural manipulation is performed independently from the main forcefield engine by graphical means. This is achieved through an interface to the freely available program GDIS developed by Dr Sean Fleming (http://gdis.seul.org/). This interface allows surfaces to be specified by their Miller indices, valid shifts to be searched for, and the geometries then to be manipulated, if necessary. Once the desired surface structure has been generated, then the necessary GULP calculation can be performed.

Surface Energy

The thermodynamic penalty for cleaving a surface from a bulk material is measured according to the surface energy. Given a bulk energy of U_{bulk} and an energy for the same system with a surface created of U_{surface}, then the surface energy, ΔU_{SE}, is defined as

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

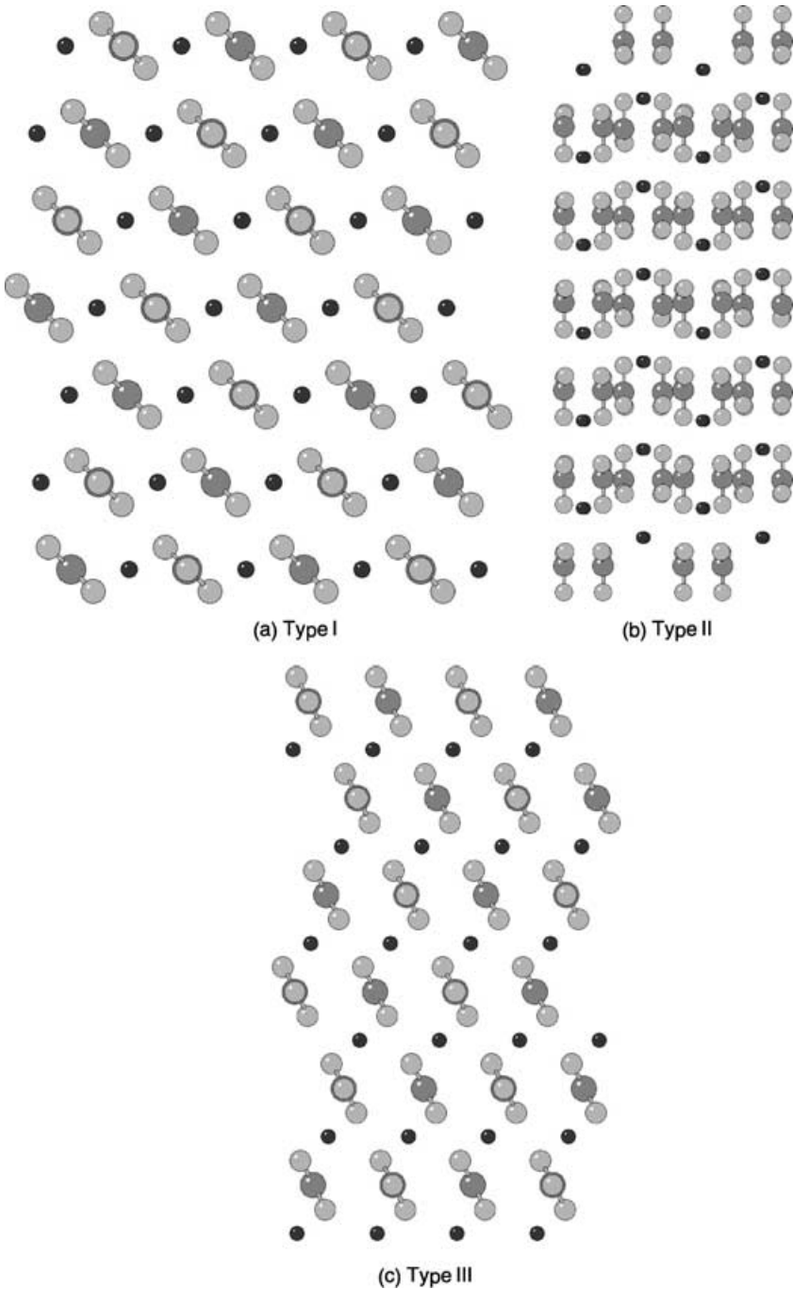


FIGURE 2 The examples of the three types of surface illustrated using polymorphs of calcium carbonate; (a) type I (b) type II and (c) type III. The type I and III surfaces illustrated are from the crystal structure of calcite, which due to its high symmetry has no type II surfaces present. The type II example is from aragonite.

an intensive quantity according to:

$$\Delta U_{\mathrm{SE}} = \frac{(U_{\mathrm{surface}} - U_{\mathrm{bulk}})}{A},$$

where A is the resulting surface area. By definition, for any stable material the surface energy will be endothermic. A calculated negative surface energy implies that a material should dissociate, i.e. the crystal should disperse into the surrounding medium.

There are two practical approaches that are widely used to determine the surface energy by computational means. In the first, a 2-D slab of material is created from the bulk, thus creating two surfaces overall. This method has the advantage that it can be used within programs that only allow for 3-D boundary conditions through the introduction of a sufficiently large vacuum gap between the images, such that the surfaces do not interact across the vacuum. In addition, it becomes necessary to assess

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

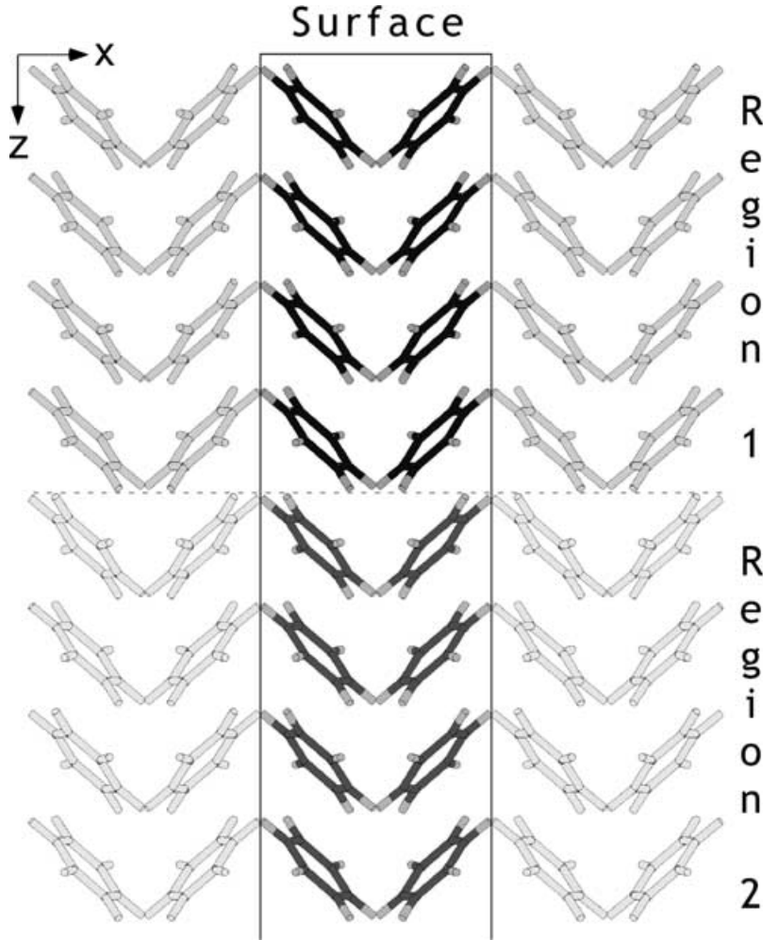


FIGURE 3 The two region surface simulation cell viewed at right angles to the surface normal, where solid vertical lines denote the boundaries between 2-D periodic images of the cell and the dash line indicates the boundary between region 1 and the frozen region 2.

whether the slab is also thick enough since the properties must converge to those of the bulk at the center of the slab. In the second method, a single surface is created by employing a two region strategy, as shown in Fig. 3. Here the solid is divided into region 1, which contains the surface and all layers of atoms below it that exhibit a significant atomic relaxation, and region 2, which contains the rest of the bulk material where it is assumed that no displacements from the 3-D crystal structure are induced. In practice, only the atoms of region 2 that have an interaction with region 1 need to be explicitly considered, and so the depth of region 2 is controlled by the cut-offs of the forcefield and the Parry sum used for the electrostatics. This second method is the most efficient and precise for atomistic techniques. However, it is considerably harder to extend to quantum mechanical methods since the electronic perturbations may extend further into the bulk and embedding, typically via Green's functions, is required to determine the influence of the electronic structure of region 2 on region 1. Through the GDIS interface, it is possible to automatically estimate the required region 1 and 2 sizes needed to converge the surface energy (with a default tolerance of $0.001\,\mathrm{J/m^2}$) based on the unrelaxed surface energies. However, if there are strong relaxations of the surface, it may be necessary to further verify that the relaxed surface energy is sufficiently converged.

The total energy for a surface calculation comprising two regions can be written in terms of the interaction energies within, and between, the different regions:

$$U_{\mathrm{tot}} = U_{11} + U_{12} + U_{22}.$$

The energy of region 2 with itself, U_{22}, is not particularly meaningful, since the region 2 is just a partial representation of the effectively infinite bulk material below, and any given particle within this region will not experience the full set of interactions that it should. However, this term is just an additive constant that is unaltered on energy minimization, or any other displacement of region 1. Consequently, it can be ignored in calculations. In the two-region model, the energy required to determine the surface

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

energy is given by:

$$U_{\text{surface}} = U_{11} + \frac{1}{2}U_{12}.$$

Note that while the above energy only includes half of the region 1–region 2 interaction energy, the objective quantity for energy minimization is the total energy, which includes the full value of U_{12}. This is because the energy change of region 2 must be allowed for when optimizing.

Attachment Energy

The surface energy, described above, provides a measure of the thermodynamic stability of a cleavage plane. However, there is a widely used alternative criterion which is the attachment energy. This is the energy associated with the addition of a stoichiometric layer of material on to the surface cut:

$$U_{\text{attachment}} = U_{\text{tot}}^{n+1} - U_{\text{tot}}^{n} - U_{\text{tot}}^{1},$$

where U_{tot}^{n} represents the total internal energy of a surface model consisting of n growth layers, and U_{tot}^{1} is the energy of the growth layer alone. For any stable material, this implies that the attachment energy must be an exothermic quantity. In practice, the calculation of the attachment energy is obtained from the energy of interaction of the growth layer at the surface with the rest of the underlying material. This benefits from the fact that the attachment energy can be obtained from a single calculation, just as is the case for the surface energy, rather than by performing the actual addition of a layer as part of a multistage process.

Although the attachment energy is also a strictly thermodynamic quantity, it is often regarded as representing the kinetics of crystal growth because of the conceptual link between the ease of the addition of a growth layer and the rate at which a surface is added to. Consequently, those faces where the attachment energy is most exothermic will tend to grow most rapidly.

Morphology

The morphology of a crystal is the macroscopic shape that it adopts. Because this can be readily observed for nearly all materials, either under an electron microscope or, in the case of many naturally occurring minerals, by visual inspection with the naked eye, it should provide a ready means to test the validity of a simulation model. Of course, the reality is rather more complex, since the morphology is sensitive to the presence of impurities, the nature of the solvent used as the growth medium, and many other factors relating to the sample preparation. Consequently, there are materials where many different morphologies can be observed for the same compound. A classic example, is that of calcite ($CaCO_3$), where there are both several different polymorphs of the bulk material and several hundred different known morphologies. Many of these variations result from biomineralization by different species. Despite this, for many pure inorganic materials morphological prediction using atomistic techniques is surprisingly successful.

Crystal morphologies can be calculated based on either the surface energy or attachment energy, which are typically taken to represent growth under conditions of thermodynamic and kinetic control, respectively. In order to do this, it is first necessary to determine the objective quantity for all significant faces. Given that dipolar faces are usually unstable, the number of likely cleavage planes for most materials is actually considerably smaller than initially might be conceived of based on permutations of the Miller indices. Furthermore, where there is space group symmetry present for the bulk, many surface planes are equivalent, thus reducing the number of unique faces. Finally, only those faces with the largest interplanar spacings are likely to appear in the morphology [95]. The actual morphology is generated as a 3-D Wulff plot [96]. Here the ratio of the surface normal distances of all planes from the center of the polyhedron are determined according to either the surface or attachment energy. The final shape of the polyhedron is then determined by the intersection of the cleavage planes. Unstable surfaces lie outside the polyhedron and never intersect. Morphological plots can also be produced through the GDIS interface to GULP.

In the equilibrium morphology approach, the contribution of a given plane to the total surface area is inversely proportional to its surface energy [97]. For the growth morphology methodology [98], the surface area contribution is again inversely proportional, but now to the negative of the attachment energy. This is because surfaces with a highly exothermic attachment energy will rapidly grow out of the morphology to leave the slow growing bounding faces.

Surface Phonons

The calculation of the phonons of a 2-D slab is exactly analogous to that for the 3-D case, except that the Brillouin zone is also now 2-D leading to dispersion only in the xy-plane (taking z to be the direction of the surface normal).

Turning to consider the standard two region surface model, there are some important issues to consider. Because region 2 is quasi-infinite, it is only possible to determine the phonons for the particles in region 1. Therefore, the dynamical matrix is constructed based on region 1 alone, but with

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

contributions to the on-diagonal matrix blocks from the potential experienced due to a rigid, non-vibrating region 2. Consequently, it is assumed that the vibrations of the two regions are completely decoupled. Since this is an approximation, the frequencies near the interface of the regions will be slightly in error, particularly in the low frequency regime where coupling is generally strongest. However, the surface modes, which are usually the ones of primary interest, will be less affected. As always, it is essential to monitor the convergence of all quantities with respect to increasing the region 1 depth. Finally, because region 1 is vibrating under the influence of an external potential, the first three frequencies at the Γ-point will no longer be zero, though they typically will be small.

Free Energy Minimization

One of the most important issues in solid state modeling is the variation of materials properties with the applied conditions. While isotropic external pressure is trivial to incorporate, as has already been shown, inclusion of the effect of temperature is more complex. This process is exacerbated by the fact that the majority of potentials have been derived through the use of some empirical (i.e. experimental) data, which has been measured under a given set of conditions. Hence, the forcefield itself can often already contain an implicit temperature, or worse, if multiple pieces of experimental data have been employed that were measured at different conditions, it can be a convolution of several temperatures. One solution to this is to extract forcefields from quantum mechanical methods so that everything is obtained explicitly at absolute zero. For now, we will assume that the forcefield has been derived so as to be free of implicit temperature effects, by whatever means.

There are several distinct approaches to the inclusion of temperature into simulations. Which one is most appropriate depends on the particular temperature and nature of the system. In the low temperature regime, the atoms of the crystal structure just execute vibrations about their lattice sites, which in the limit of absolute zero will be purely harmonic. This situation is best described through the use of lattice dynamics, which is the quasi-static approach to treating a vibrating lattice. As the temperature increases, the motions will become increasingly anharmonic. In principle, this can be handled, to a point, through the use of anharmonic corrections to the harmonic lattice dynamics, in order to allow for weak phonon–phonon coupling. However, when the temperature becomes sufficiently high that diffusion can occur it is necessary to switch to an alternative approach. Typically, one of two approaches can then be employed. If detailed information is required concerning the atomic motions and related properties, then the method of choice is molecular dynamics [99]. This propagates Newton's equations of motion through time using a finite difference formalism. Hence, it retains time correlation functions and the trajectories of all atoms. Its strength is that anharmonic effects are fully included. However, since it regards nuclei as classical particles, it is invalid at low temperatures due to the neglect of the quantization of vibration and zero point motion, unless the Path-Integral formalism is employed. It can be seen therefore that lattice dynamics and molecular dynamics are largely complementary in their regions of applicability. Although GULP also includes the capability to perform molecular dynamics, this topic will not be discussed here since this area, in regard to other codes, has been discussed elsewhere [16], and the more unique, static, features will be focused on here. Likewise, the effect of temperature can also be explored through Monte Carlo simulation [83], if the focus is integration over phase space, with no regard to timescale or kinetic factors, but this topic will not be discussed further because it is not one of the more novel features of GULP.

Concentrating now on the use of lattice dynamics for examining the temperature dependence of crystal properties, the dominant effect is the change in the crystal structure. When heated, most materials undergo thermal expansion, which correspondingly leads to softening of many of the mechanical properties. There are a few celebrated examples of materials that actually contract on heating, through the rotation of quasi-rigid polyhedra, which is technologically very important in the quest for zero thermal expansion composites.

It has already been shown that it is possible to calculate the Helmholtz free energy for a given structure as a function of temperature through the determination of the vibrational partition function from the phonon density of states. Hence, a natural approach to determine the dependence of structure on temperature is through free energy minimization. Here the key foundation is the quasi-harmonic approximation, which assumes that the vibrational frequencies can be determined as if the atoms are vibrating purely harmonically while the cell parameters are adjusted to minimize the free energy. Previous studies have indicated that this is a reasonable approximation until a temperature of approximately half the melting point has been reached.

The major barrier to free energy minimization is that we have already seen that efficient optimization requires at least the first derivatives of the quantity with respect to the structural variables. Hence, a number of approaches evolved for tackling this

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

problem. LeSar *et al.* [100] developed a method whereby the atoms where represented as Gaussian distributions, whose width represented the vibrational motion. The Gaussian exponent was then regarded as a variational parameter that minimized the free energy. Hence, the free energy could be obtained without direct recourse to a phonon density of states, making derivatives straight forward. While this approach is easy to apply to metals, the extension to more complex ionic systems is more demanding. A method that approximated the phonon density of states was introduced by Sutton [101], which involved taking moments of the dynamical matrix, as pioneered by Montroll [102] several decades earlier, and in the spirit of tight binding theory. This approach has the advantage that the derivatives are taken of the dynamical matrix elements, rather than of phonon frequencies, which are the product of a matrix diagonalization.

In the field of ionic materials, Parker and Price [8] pioneered the use of free energy minimization through the use of numerical derivatives. Here central finite differences were taken with respect to the cell strains, with the internal degrees of freedom being formally optimized at every point. This approach had the advantage over the other methods that no approximation was being introduced beyond the quasiharmonic one. However, because this requires $(2N + 1)$ optimizations and phonon evaluations per energy/gradient evaluation with respect to the free energy, the minimization was restricted to the strains alone in order to reduce the number of variables, N, to a maximum of six.

More recently, Kantorovich [103] has derived expressions for the analytical derivatives of the free energy, which were implemented contemporaneously in the program SHELL of Allan and co-workers [104], and in GULP [105]. If we consider the differential with respect to strain, though the expressions would be identical for any degree of freedom, then the first derivative of the free energy is given by:

$$\left(\frac{\partial A}{\partial \epsilon}\right) = \left(\frac{\partial U_{\text{static}}}{\partial \epsilon}\right) + \sum_k \sum_m \left\{ \frac{h}{2\omega} \left(\frac{1}{2} + \frac{1}{\exp\left(\frac{h\omega}{k_B T}\right) - 1} \right) \left(\frac{\partial \omega^2}{\partial \epsilon}\right) \right\},$$

where the expression is written in terms of the derivatives of the square of the frequencies, since these values represent the eigenvalues of the dynamical matrix. Their derivatives can be calculated through the application of perturbation theory and expressed as the derivatives of the dynamical matrix projected onto the corresponding eigenvectors:

$$\left(\frac{\partial \omega(k,m)^2}{\partial \epsilon}\right) = e_m^*(k) \left(\frac{\partial D(k)}{\partial \epsilon}\right) e_m(k).$$

In order to determine the derivatives of the phonon frequencies, we therefore require the phased third derivatives of the energy with respect to either three Cartesian coordinates, or two Cartesian coordinates and one strain for internal and external variables, respectively.

For the most efficient optimizers, based on Newton–Raphson techniques, we strictly need the second derivatives with respect to the free energy, which corresponds to the fourth derivatives of the internal energy. However, this is considerably more expensive and complex, so the Hessian matrix is usually approximated by the conventional internal energy Hessian by assuming that the free energy contribution to the curvature is small. Furthermore, the use of updating formulae will ultimately correct for the discrepancy given sufficient optimization steps.

With the advent of analytical derivatives, it is now possible to consider two types of free energy minimization. The first has been christened the Zero Static Internal Stress Approximation (ZSISA) by Allan and co-workers [106], which resembles the approach taken with numerical first derivatives (i.e. the unit cell is minimized with respect to the free energy, while maintaining the internal degrees of freedom at a minimum with respect to the internal energy). When working within this formalism there is an additional contribution to the strain derivatives, which corrects for the fact that internal degrees of freedom are not at a minimum with respect to the free energy:

$$\left(\frac{\partial A}{\partial \epsilon}\right)_{\text{ZSISA}} = \left(\frac{\partial A}{\partial \epsilon}\right) - \left(\frac{\partial^2 A}{\partial \epsilon \partial \alpha}\right) \left(\frac{\partial^2 A}{\partial \alpha \partial \beta}\right)^{-1} \left(\frac{\partial A}{\partial \beta}\right).$$

Here the approximation is again made that the second derivative matrices can be approximated by neglecting the free energy contribution; an approximation that should be enhanced by the cancellation that results from taking a ratio. The second type of optimization can be called full free energy minimization (FFEM), in which the internal degrees of freedom are also minimized with respect to the free energy. This is potentially appealing since sometimes it is the details of the internal changes that might be of interest, for example, the nature of an adsorption complex within a microporous material.

Results for silicate materials show a number of interesting features regarding the merits of both approaches. Firstly, in the case of α-quartz where accurate experimental data is available for comparison, it appears that the ZSISA approach

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

underestimates the thermal expansion, while the FFEM method is more accurate, though of course this is subject to the limitations of the specific potential model chosen [105]. However, for all silicates tried so far the full minimization approach goes catastrophically wrong at about ambient conditions. This is illustrative of a general problem with this approach, which tends to drive systems to instability. This can be readily understood, since the free energy is minimized by creating phonons that tend to zero frequency and hence the structure is motivated to create soft modes. This behavior does not tend to happen in the ZSISA approach where only the cell strains are directly coupled to the free energy and thus the relaxations tend to lead only to a scaling of modes, rather than more individual changes. Hence, the use of ZSISA is far more robust and generally recommended for most purposes.

Defect Calculations

While the simulation of bulk material properties is important, just as crucial is the study of both intrinsic and extrinsic defects. Many of the key applications of solid state systems, such as catalysis, electronic and ionic conductivity, ion exchange and waste immobilization, critically depend on the utilization of the characteristics of defect centers. Consequently, from the early days of the field of atomistic simulation defect studied have been one of the most vigorously pursued topics.

There are two widely used approaches for performing defect calculations on solids; the supercell and the cluster methods, with or without embedding in the latter case. Both approaches have their merits and demerits. Putting computational implementation factors aside, the use of embedded clusters is ideal for the infinitely dilute limit, while the supercell method is more appropriate for high concentrations of defects where there exists significant defect–defect interaction. In practice, the nature of the computational method often biases the method of choice. However, with atomistic techniques both approaches are feasible. Since the supercell method is simply the extension of a bulk calculation, we will focus here on the embedded cluster approach. In this particular context, the technique is generally referred to as the Mott–Littleton method [107], after the authors of the pioneering work in the field, though the implementational details differ a little from their original work.

The basis of the Mott–Littleton method is the so-called two region strategy. Here a point called the defect center is defined, which typically lies at a point concentric with the initial defect site, or, where there is more than one defect, at the mid-point of the ensemble of point defects. The crystal around the defect center is divided into two spherical regions, with the inner sphere being labeled region 1, and the outer spherical shell of ions being region 2a. Atoms outside of these spheres belong to region 2b, which then extends to infinity. The dimensions of these regions are typically specified either by their radii or the number of ions contained within them. The ions in region 1 are assumed to be strongly perturbed by the defect and therefore are relaxed explicitly with respect to their Cartesian coordinates. In contrast, the ions in region 2 are assumed to be weakly perturbed and therefore their displacements, with the associated energy of relaxation, can be approximated in some way. Clearly, as the region 1 radius is increased these approximations will become increasingly valid, and thus an important stage of a defect calculation is to ensure that the defect energy is sufficiently converged with respect to the region radii. In some cases the convergence of the absolute defect energy can be slow, in which case it may be sufficient to monitor the convergence of relative defect energies instead. As a guideline, the radius of region 1 and the difference of the radii of regions 1 and 2 should both be greater than the short-range potential cut-off to achieve convergence, though for charged defects this may not be adequate.

Within the Mott–Littleton scheme, we can express the total energy of the two region system as the sum of contributions from the energies within the regions and between them:

$$U_{\text{tot}}(x, \xi) = U_{11}(x) + U_{12}(x, \xi) + U_{22}(\xi),$$

where $U_{11}(x)$ represents the energy of region 1 as a function of the Cartesian coordinates, x, $U_{22}(\xi)$ represents the energy of region 2 as a function of the Cartesian displacements, ξ, and $U_{12}(x, \xi)$ is the energy of interaction between the two regions. At this stage we do not distinguish between regions 2a and 2b. If the forces acting on region 2 are small, then we can assume that the response of the atoms in this region will be purely harmonic. Hence, the energy of region 2 can be written as:

$$U_{22}(\xi) = \frac{1}{2}\xi^T H_{22}\xi,$$

where H_{22} is the Hessian matrix for region 2. If we now apply the condition that the displacements in region 2 will be the equilibrium values this yields the following condition:

$$\left(\frac{\partial U_{\text{tot}}(x, \xi)}{\partial \xi}\right)_x = \left(\frac{\partial U_{12}(x, \xi)}{\partial \xi}\right)_x + H_{22}\xi = 0.$$

Combining this equation, and the previous one, it is possible to eliminate the energy of region 2 from the total energy without direct recourse to the Hessian

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

matrix (which would be of infinite dimension):

$$U_{\mathrm{tot}}(x,\xi) = U_{11}(x) + U_{12}(x,\xi) - \frac{1}{2}\left(\frac{\partial U_{12}(x,\xi)}{\partial \xi}\right)_x \xi.$$

Thus, the problem of calculating the energy of region 1 in the potential of region 2 has been reduced to evaluating the energy of region 1 and its interaction with region 2, without having to evaluate the self energy of region 2. Furthermore, in order to lead to partial cancellation of terms, the quantity calculated is the defect energy—i.e. the difference between the energy of the perfect regions 1 and 2, U^p_{tot}, and the defective case, U^d_{tot}, rather than the individual contributions:

$$U_{\mathrm{defect}}(x,\xi) = U^d_{\mathrm{tot}}(x,\xi) - U^p_{\mathrm{tot}}(x,\xi).$$

It should be noted that it is assumed that the energy of any species removed or added during the formation of the defect has an energy of zero at infinite separation. If this is not the case, then the defect energy must be corrected *a posteriori* for this. However, such corrections have no influence on the outcome of the defect calculation itself.

There is one important consequence of the above formulation of the total defect energy, in that it becomes necessary to find the optimized energy with respect to the Cartesian coordinates of region 1 by force balance, rather than by energy minimization. The point at which the forces tend to zero is usually only slightly different from the minimum in the internal energy, depending on the degree of perturbation of region 2. Hence, during an optimization of the defect energy GULP initially minimizes the energy, as per a conventional Newton–Raphson procedure. Once the gradient norm falls below a specified tolerance, and region 1 lies within the harmonic region, then the harmonic displacements according to the Hessian and gradients are applied without the use of a line search until the forces drop below the specified tolerance.

A further important point relates to the state of the bulk crystal when performing defect calculations. Because of the use of displacements in region 2, it is crucial that the bulk structure is at an energy minimum with respect to the internal coordinates before performing a defect calculation. Hence, the bulk crystal must be relaxed to equilibrium at least at constant volume, if not at constant pressure. Furthermore, it is also important that there are no imaginary phonon modes within the Brillouin zone otherwise the displacements in region 2 may correspond to unstable harmonic equilibrium. The presence of such modes is the most common cause of unphysical results, for instance, obtaining negative defect energies for intrinsic defects. Because the presence of defects lowers the symmetry, a Mott–Littleton calculation may encounter instabilities not apparent for the bulk material.

Now it is necessary to consider the treatment of region 2 in more detail, and in particular the difference between regions 2a and 2b. In region 2a, the forces on the individual ions due to short-range interatomic potentials and the Coulomb term are explicitly calculated and the local displacement evaluated. While the forces on region 2a are technically due to all ions in region 1, a commonly used approximation is to evaluate the forces due to defect species only. This is the approach taken in the default calculation method for compatability with earlier results.

In contrast to the above situation for region 2a, for region 2b the energy of relaxation must be determined implicitly since this region extends to infinity. It is assumed that in region 2b the only force acting is that due to the Coulomb potential in order to simplify the problem. By choosing the radius of region 2a to always be greater than that of region 1 by the short-range potential cut-off this can always be arranged to be valid. Even having retained only the Coulomb interaction in the perturbation of region 2b, this still leaves many interactions to consider. To simplify the problem, the electrostatic potential due to defects in region 1 can be represented by the multipole moments of the deformation situated at the defect center. If region 2a is sufficiently large, then the only significant term will be the monopole moment of the defect (i.e. the net charge). Hence, the energy of region 2b is evaluated as the induced relaxation energy due to the net charge of the defect situated at the specified defect center. While this is a significant approximation, it usually works well in practice and again becomes valid with increasing region sizes. In the original Mott–Littleton method, the interaction with region 2b was described by a continuum approximation. However, in subsequent implementations a sum of the induced polarization energy is performed over the atomic sites. For the general case of an anisotropic dielectric constant tensor [108], the energy of region 2b is calculated as:

$$U_{2b} = -\frac{1}{2}Q^2\left(\sum_{i\in 2b}\sum_{\alpha\beta}\frac{M_i^{\alpha\beta}r_{ci}^{\alpha}r_{ci}^{\beta}}{r_{ci}^{6}}\right),$$

where Q is the net charge of the defect, r_{ci} is the distance of the ith atom from the defect center, and $M_i^{\alpha\beta}$ is a 3×3 matrix for each atomic site, in the Cartesian frame, that represents the on-site polarizability. The quantitative definition of the matrix elements $M_i^{\alpha\beta}$ is [4]:

$$M_i^{\alpha\beta} = \sum_{\gamma}\left(\left(D^{-1}\right)^{\alpha\gamma}q\right)_i\left(\epsilon^{-1}\right)^{\gamma\beta},$$

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

where D_i^{-1} represents the on-diagonal block of a modified inverse second derivative matrix and ϵ is the static dielectric constant tensor. Note that the inverse of the second derivative matrix is singular and therefore a further approximation must be made. Physically, this problem corresponds to the division of the polarization between the sub-lattices being arbitrary. The usual solution taken is to assume that the polarization is divided equally between the cation and anion sub-lattices, with the second derivative matrix being modified correspondingly. In the special case of a cubic crystal, the expression for the region 2b energy can be simplified to:

$$U_{2b} = -\frac{1}{2}Q^2\left(\sum_{i\in 2b}\frac{m_i}{r_{ci}^4}\right),$$

where m_i is the average of the diagonal elements of the matrix M (for the cubic case these will all be equal, but if the isotropic formula were to be applied in a non-cubic case this would not be so).

Because the expression for the region 2b energy is not particularly short-ranged, it is appropriate to evaluate it using lattice summation techniques analogous to those applied to the monopole–monopole term. Hence, the term is summed to convergence based on a perfect infinite lattice and then the contributions from ions in regions 1 and 2a are subtracted off. The distance dependent factor within the expression for the energy can be rewritten as:

$$\frac{r^\alpha r^\beta}{r^6} = \frac{1}{8}\left(\frac{\partial^2\left(\frac{1}{r^4}\right)}{\partial r^\alpha \partial r^\beta}\right) + \frac{\delta_{\alpha\beta}}{4}\left(\frac{1}{r^4}\right).$$

Hence, by evaluating the equivalent of the Ewald summation for the inverse fourth power of distance, and its second derivatives with respect to Cartesian displacements, it is possible to achieve a rapidly convergent expression for the energy of region 2b.

Having described how defect energies are calculated in theory, we mention a few practical points. There are three types of defect that can be specified within GULP:

- Vacancy
- Interstitial
- Impurity.

The last one of these three is clearly the combination of a vacancy and an interstitial at the same atomic position in the structure. The location of a vacancy or an impurity can be specified by reference to either a spatial position or an atom position by referencing the site. An interstitial, by its very nature, must be specified by coordinates. When an atom is designated to be vacant, then both the core and shell will be removed automatically since it would not be sensible to leave one or the other present in the system. It is also possible to specify that a whole molecule be removed from the system, when the connectivity has been defined.

Most types of calculation, that are logically applicable, can also be applied to defect runs, such as optimization to a minimum or a transition state. In the case of vibrations, the frequencies are calculated for a dynamical matrix based on region 1. It must be noted that this is a large approximation, since any modes that are coupled between regions 1 and 2 will not be properly described. Only localized modes relating to atoms near the center of region 1 will be correctly predicted. Consequently, such calculations of vibrations must be interpreted cautiously.

Finally, a degree of point group symmetry may be utilized during defect calculations. An automatic search for common symmetry elements is performed about the defect center, including mirror planes and C_2 axes that are aligned with, or between, the Cartesian axes. Such symmetry is used to reduce the number of regions 1 and 2 atoms stored, thereby reducing the number of degrees of freedom for optimization, as well as taking advantage of symmetry adapted algorithms to accelerate the calculation of the energy and its derivatives.

Derivation of Interatomic Potentials

One of the major challenges facing anyone wishing to use forcefield methods is to determine the functional form and parameters required to calculate the energy and properties. In the field of organic chemistry and biomolecules, there are a number of well-established forcefields, such as MM3 [109], and those associated with the programs AMBER [14], and CHARMM [15], which are, in principle, capable of handling most elements typically found in these systems (C, H, O, N, S, P, etc.) in their common hybridization states. These are constructed around the molecular mechanics approach where the forcefield is connectivity driven and interactions are described in terms of two-, three-, and four-body bonding terms, plus long-range Coulomb and non-bonded interactions. While there have been several attempts to generate general forcefields that cover the entire periodic table, such as UFF [110], ESFF [111], etc., none have been completely successful. Because of the enormity of the amount of data required when spanning the whole range of elements, it is impractical to determine such a universal forcefield by empirical fitting. Instead general rules must be used to predict parameters based on extrapolations and intrinsic properties of the element that are known—for instance, the electronegativity. Not surprisingly, this leads to limitations in the quality of the results. For most

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

inorganic systems it is usually necessary to derive a forcefield for the specific material, or family of materials, of interest.

There are two means by which a forcefield can generally be derived, if we exclude rule based extrapolations. Firstly, it is possible to obtain parameters by fitting to a potential energy surface obtained from a non-empirical theoretical method. This would typically consist of results from *ab initio* calculations, ideally on the periodic solid [112], or perhaps on a gas phase cluster, or even better, both of the aforementioned sources. The potential energy surface can be fitted either as a sequence of geometries with their corresponding energies, or derivatives of the energy could also be included to maximize the amount of information from each higher level calculation. Many of the early forcefields for ionic materials were determined using electron gas methods [113], in which the energies of interaction between pairs of ions were determined by a density functional calculation using the overlapping ionic electron densities, where the anion is confined in an appropriate potential well. Secondly, parameters can be obtained by empirical fitting in which the normal process of using a forcefield to determine the properties of a material is inverted. This approach depends on the availability of a range of experimental data. Knowledge of the crystal structure is a definite prerequisite for this method, but is insufficient alone since information is required as to the curvature of the energy surface about the minimum. This later data may come from quantities such as elastic constants, bulk moduli, piezoelectric constants, dielectric constants or phonon frequencies.

In order to perform a fit, first it is necessary to define a quantity that measures the quality of the results, known as the sum of squares;

$$F = \sum_{i=1}^{N_{\text{obs}}} w_i \left(f_i^{\text{obs}} - f_i^{\text{calc}} \right)^2,$$

where N_{obs} is the number of observables, f_i^{obs} and f_i^{calc} are the fitted and calculated values of the observable, respectively, and w_i is the weighting factor for the given observable. Because of the weighting factor, there are an infinite number of possible solutions all of which are equally valid. Hence, one of the most important skills is knowing how to choose appropriate and sensible weighting factors. There are several criteria that can be used for guidance though. Firstly, if fitting experimental data, the weighting factor should be inversely proportional to the uncertainty in the measured value. Obviously, trusted, precise values should be given more priority than data where there are large error bars. Secondly, the weight factor should be inversely proportional to the magnitude of the observable squared. This ensures that all values are fitted on an equal footing, regardless of units. For example, fitted vibrational frequencies in wave numbers are typically two to three orders of magnitude larger than structural variables.

The fitting process itself involves minimizing the above function *F*. To do this, the default approach is similar to that used in optimization. For many terms, the evaluation of the derivatives of the sum of squares with respect to the variables is complex, and in some of the fitting algorithms that will be discussed subsequently it is even impossible. As a result, numerical derivatives are employed during fitting since it greatly simplifies the process. Because of this, the default optimizer is to use a BFGS update of an initial on-diagonal only Hessian, obtained by finite differences, in a Newton–Raphson process. However, for particularly difficult cases, where correlation between variables is strong, there is the option to use a full Hessian matrix, again obtained by finite differences.

Now we turn to specifically consider the fitting methodology for the case of empirical data. Traditionally, the experimental structure is fitted by varying the potential parameters so as to minimize the forces at this configuration, and this is the default strategy. Other observables, such as elastic constants etc, are then calculated at the experimental structure too. When working with the shell model, either dipole or breathing shell, there is an additional complication though in that while the cores are fully specified since they are equated with the nuclei of the ions, the positions/radii of the shells are undefined. One approach is to fit with the shells positioned to be concentric with the cores. However, this is unphysical since it implies that there is no ionic polarization, which defeats the object of including the model in the first place. A second approach might be to place the shell according to specified ion dipoles, but this information is almost impossible to come by. Only data about the total polarization of the unit cell is typically available and a unique atomic decomposition is not possible. In order to handle this issue, the simultaneous relaxation of shells approach was introduced into GULP [114]. Here the shells are allowed to move during fitting. Formally, the most correct approach is to allow the shells to be energy minimized at every evaluation of the fitting function. However, a simpler approach has been implemented in which the shell forces are added as observables and the shell positions become fitting variables. In this way, the shells are minimized directly within the fitting procedure. In the absence of any observables other than the structure, this is exactly equivalent to minimizing the shells at every step of fitting. When observables are present there will be small differences, but these are usually an acceptable price to pay for the greater ease of implementation.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

There is actually a more fundamental flaw with the approach of fitting forces at the experimental geometry as a method of fitting. Often we judge the quality of a fit by the percentage error in the structural variables rather than the forces. Although lowering the forces during a fit generally improves the optimized structure with respect to experiment, this is not guaranteed to be the case (and indeed we have found examples where it is not). This can be understood by making a harmonic estimate of the atomic displacements, x, based on the forces, f:

$$x = H^{-1}f.$$

It can be readily seen that the magnitude of the displacements also depends on the inverse of the Hessian matrix. Thus, if the forces improve, but the description of the curvature about the minimum deteriorates, then the errors can potentially increase. If curvature information is included in the fit, then this can tend to reduce this problem. However, there is a further difficulty here. Formally speaking, the expressions for the elastic constants and other properties are defined about a position of zero stress and zero internal derivative. Therefore, calculating the properties at the experimental structure when the forces are non-zero leads to errors also. The solution to both of these dilemmas is to use the so-called relax fitting methodology [114] in which the structure is fully optimized at every evaluation of the fitting function and the properties calculated at the optimized configuration. Obviously, this is a far more expensive procedure, but does yield the most reliable results. Also there is the requirement that the initial potential set is reasonable enough to actually give a valid minimization of the structure.

Having obtained an apparently successful fit, it is important to assess the quality of the results, since there are plenty of pitfalls and so convergence should not be taken to represent a good quality solution. Firstly, the potential model should be tested for all reasonable properties, not just those used in the fit. It could easily be the case that a forcefield reproduces a high symmetry structure and, say, a single curvature observable, such as the bulk modulus. However, examination of the full elastic constant tensor, dielectric properties and phonons might reveal that the system is unstable with respect to a lowering of symmetry which would not show up in the fit. Secondly, the forcefield could be transferred to a different material, not used in the original fit, to test whether the results are sensible. Finally, it is important to look at the potential parameters and assess whether the numbers are physically sensible. For instance, it is not uncommon for dispersion terms to become extremely large if allowed to fit unconstrained. While this might have improved the sum of squares for one particular system, it means all hope of transferability has been lost. Similarly, there can often be problems with fitting A and ρ of a Buckingham potential concurrently due to the strong correlation coefficient between the parameters.

The focus above has been primarily on empirical derivation of interatomic potentials. However, with the increasing ability to perform periodic *ab initio* calculations on solids an attractive alternative is to derive forcefields that attempt to reproduce the results of such methods. There are several reasons to take this approach. Firstly, by fitting the outcomes of a single Hamiltonian it is possible to guarantee that the training set is fully consistent (i.e. there are no differences in temperature, pressure, sample quality or variable uncertainties in the observables). Secondly, data can be obtained for materials were no experimental information exists or at geometries that are significantly perturbed from the equilibrium one. Thirdly, the data from quantum mechanical methods is free of statistical mechanical effects, such as thermal vibrations and zero point motion. Hence, if the aim is to perform free energy minimization then the interatomic potentials will represent a proper starting point for the inclusion of these quantities.

Fitting of quantum mechanical data can be performed in one of two ways, either by proceeding in the same fashion as for empirical derivation, or by use of an energy hypersurface. In the latter case, this is achieved by specifying a series of structures with their corresponding energies, and optionally first derivatives. Typically, the structures would include the equilibrium configuration and as many distinct distortions about this point in order to probe as many different interatomic distances between atoms as possible. Perhaps the most difficult decision is how to weight the configurations. Unless the forcefield is able to reproduce the *ab initio* data very accurately, then it is usually desirable to weight the fit in favor of configurations nearer the equilibrium structure. One approach that has been taken is to use a Boltzmann factor weighting based on the energy difference to the minimum energy configuration, with an appropriate choice of temperature to the task ultimately to be performed [115]. However, there are many other possible choices too. A further issue concerns the fitting of quantum mechanical energies. Unless these values have been converted to a binding or lattice energy with reference to the dissociated state of the species within the system then it is inappropriate to fit the absolute values of the energies. Consequently, the easiest solution is to include an additive energy shift parameter in the fit, such that only relative energies are actually fitted.

Finally, the option exists within GULP to perform fitting using GAs as well as via least squares techniques. This may be potentially useful in cases where a complex system is being fitted when there is no reasonable starting approximation to

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

the forcefield available or where there may be multiple local minima in the parameter space. However, to date we have yet to encounter a situation where this approach has proved beneficial over the more conventional methodology. This emphasizes that there is no substitute for making physically sensible choices for the functional form of the forcefield and the initial parameters.

Calculation of Derivatives

In order to be able to optimize structures efficiently and to calculate many properties requires the availability of derivatives. While finite difference methods can be used to determine these, this is both inefficient and inaccurate for forcefields, since numerical errors can cause problems, especially when potentials do not go exactly to zero at the cut-off distance. Consequently, all derivatives are determined analytically in GULP. All functional forms for the energy have up to analytic second derivatives available, while two-, three- and four-body interactions include third derivatives. In addition, analytic third derivatives can be calculated for the embedded atom method, but currently not for any other many-body potential functions.

Because the determination of derivatives is central to many quantities, a brief description of the approach to their calculation is given here, while fuller details can be found in both the original Harwell Report [116], and the subsequent publication of many of these details [117]. Ultimately, two types of derivative are required—those with respect to atomic degrees of freedom and those with respect to the unit cell. In GULP, all atomic coordinate derivatives are calculated in the Cartesian frame of reference and then transformed to the fractional coordinate derivatives, when appropriate to the periodicity. For the unit cell, strain derivatives are taken as previously described.

Both the Cartesian and strain derivatives can be related to a set of pivotal quantities, which are the first derivatives with respect to the interatomic distances, divided by the distance:

$$\frac{\partial U}{\partial \alpha} = \left(\frac{1}{r}\frac{\partial U}{\partial r}\right)_{\alpha}$$

$$\frac{\partial U}{\partial \epsilon} = \left(\frac{1}{r}\frac{\partial U}{\partial r}\right)_{\alpha\beta},$$

where, in the expression for the strain derivative, the components α and β are the appropriate Cartesian directions for the given strain (e.g. ϵ_4 implies $\alpha = y$ and $\beta = z$). It is implicit that all quantities are subscripted with ij to indicate that the terms refer to a specific pairwise interaction. Let us introduce the shorthand:

$$U_1 = \left(\frac{1}{r}\frac{\partial U}{\partial r}\right)$$

$$U_2 = \left(\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial U}{\partial r}\right)\right)$$

$$U_3 = \left(\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial U}{\partial r}\right)\right)\right).$$

The second derivatives can then be obtained by differentiating the above expressions once more and written as:

$$\frac{\partial^2 U}{\partial \alpha \partial \beta} = U_2 \alpha\beta + U_1 \delta_{\alpha\beta}$$

$$\frac{\partial^2 U}{\partial \alpha \partial \epsilon} = U_2 \alpha\beta\gamma + U_1(\beta\delta_{\alpha\gamma} + \gamma\delta_{\alpha\beta})$$

$$\frac{\partial^2 U}{\partial \epsilon_{\alpha\beta} \partial \epsilon_{\gamma\zeta}} = U_2 \alpha\beta\gamma\zeta + U_1(\delta_{\alpha\beta}\delta_{\gamma\zeta}) + \frac{1}{4}U_1(\delta_{\alpha\gamma}\beta\zeta + \delta_{\alpha\zeta}\beta\gamma + \delta_{\beta\gamma}\alpha\zeta + \delta_{\beta\zeta}\alpha\gamma).$$

Likewise, the third derivatives can be obtained by further differentiation:

$$\frac{\partial^3 U}{\partial \alpha \partial \beta \partial \gamma} = U_3 \alpha\beta\gamma + U_2(\alpha\delta_{\beta\gamma} + \beta\delta_{\alpha\gamma} + \gamma\delta_{\alpha\beta})$$

$$\frac{\partial^3 U}{\partial \alpha \partial \beta \partial \epsilon_{\gamma\zeta}} = U_3 \alpha\beta\gamma\zeta + U_2(\alpha(\gamma\delta_{\beta\zeta} + \zeta\delta_{\beta\gamma}) + \beta(\gamma\delta_{\alpha\zeta} + \zeta\delta_{\alpha\gamma}) + \gamma\zeta\delta_{\alpha\beta}) + U_1(\delta_{\alpha\gamma}\delta_{\beta\zeta} + \delta_{\alpha\zeta}\delta_{\beta\gamma}).$$

Note that for free energy minimization, which is were the third derivatives are required, only the derivatives with respect to two Cartesian coordinates and one strain, or three Cartesian coordinates are needed.

In both three- and four-body terms there exist derivatives with respect to either a trigonometric function of an angle, or the angle itself. These derivatives can be converted into the above forms through the use of the cosine rule:

$$\cos(\theta) = \frac{r_{12}^2 + r_{13}^2 - r_{23}^2}{2 r_{12} r_{13}},$$

where θ is the angle at the pivot atom 1, lying between the vectors r_{12} and r_{13}. Derivatives with respect to angles are therefore obtained through the expression:

$$\frac{\partial \theta}{\partial r} = -\frac{1}{\sin(\theta)}\frac{\partial \cos(\theta)}{\partial r}.$$

Care must be taken in handling the limit as θ tends to either 0° or 180°.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

Crystal Symmetry

An important topic, particularly in the context of optimization, is crystal symmetry. For periodic crystals, there is the option to specify the solid via the asymmetric unit atoms, the space group number/symbol, and the origin setting. It should be noted that the use of space group symbols is preferable since it distinguishes between different settings of the same space group. Apart from the building of the full unit cell from the asymmetric unit, the symmetry can be utilized to greatly accelerate the structural optimization through several different means. The first benefit of symmetry is that it lowers the number of independent geometric variables, since many atomic coordinates are related via a roto-translation operation. Furthermore, the existence of special positions implies that some coordinates are not allowed to vary, typically for sites such as $(1/2, 1/2, 1/2)$, and that in other cases coordinates are not independent and must be related by constraints. Even larger reductions in scale can be achieved by using unit cell centering to reduce the centered-cell down to the primitive representation, thereby reducing the number of atoms by a factor or between 2 and 4. All of these factors reduce the number of variables, and since the rate of convergence is usually correlated to the number of independent variables, depending on the algorithm, this can lead to fewer optimization steps being required. Furthermore, the amount of memory to store the Hessian matrix is reduced, as well as there being a large improvement in the cost of inverting this matrix.

The second gain from the use of symmetry is that it is possible to use new algorithms to calculate the energy and its derivatives that involve fewer computations, provided the symmetry is high enough [118]. Considering a two-body potential model, instead of looping over all possible pairs of atoms in order to compute the energy, when symmetry is present it is sufficient to only calculate the interactions between the atoms of the asymmetric unit and all other atoms. For a system of N^f atoms in the full unit cell and N^a atoms in the asymmetric unit, then the execution times for the algorithms, T^f and T^a, respectively, will be given by:

$$T^f \propto \frac{1}{2} N^f \left(N^f + 1\right)$$

$$T^a \propto N^a N^f.$$

It can therefore be seen that provided the number of atoms in the asymmetric unit is roughly less than half the number in the full unit cell then the symmetry adapted algorithm will be faster. The extension of symmetry to other types of potential is also straightforward. For example, for a three-body angular potential it is only necessary to calculate the terms that arise when an asymmetric unit atom lies within a valid triad of species.

While the symmetry-adapted evaluation of the energy is trivial, more care is required for the derivatives since these are vector/tensorial properties. For the first derivatives with respect to atomic positions, it is sufficient to again only evaluate the derivatives with respect to the asymmetric unit atoms and then to scale these terms by the number of symmetry-equivalent atoms in the full cell. Conversely, the first derivatives with respect to cell strain are more complex, since although the interaction of the asymmetric unit with all other atoms spans all possible unique derivatives, the orientation of the terms now matters. In short, the strain derivatives will violate the crystal symmetry if the terms are evaluated for the asymmetric unit interacting with the rest of the crystal alone. However, given that the crystal symmetry is specified and that the sum of the terms is correct, when again scaled by the site multiplicities, it is possible to obtain the correct first strain derivatives by appropriate averaging.

Turning now to the second derivatives, the symmetry-adapted generation of the Hessian matrix is more complex. Considering the process by which the Hessian is generated in the absence of symmetry, there are three steps. Firstly, the Cartesian derivatives that are initially generated have to be transformed into fractional space by multiplying each 3×3 block by the unit cell vector matrix and its transpose. This generates the matrix D_{ff}, where the subscript f signifies a dimension equal to number of atomic coordinates in the full unit cell. Secondly, D_{ff} is reduced to the smaller matrix D_{aa}, where the subscript a now signifies that the dimension is equal to the number of atomic coordinates in the asymmetric unit. This is achieved using the transformation matrix, T_{fa}, and its transpose:

$$D_{aa} = T_{af} D_{ff} T_{fa}.$$

The transformation matrix is sparse and contains 3×3 blocks between each asymmetric unit atom and its symmetry equivalent images, which are just the rotation matrices, R, that created those images. The third, and final, step is to reduce the matrix D_{aa} according to any constraints that are present between coordinates. It is possible to combine the second and third steps into one by pre-multiplying the transformation matrix by the constraint matrix.

All the symmetry unique information concerning the second derivatives is contained within the columns between the asymmetric unit and the full cell. Hence, it is more efficient just to calculate the sub-matrix D_{fa} instead. This can then be transformed

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

to the required matrix according to:

$$D_{aa} = S_{af} D_{fa}.$$

This not only reduces the computational effort in calculating the second derivatives, but also reduces the memory required and lowers the number of matrix multiplications required to form the Hessian. The 3×3 blocks of the required transformation matrix are given by:

$$S_{af} = \frac{1}{N^a_{\text{eqv}}} \sum_{n=1}^{N^a_{\text{eqv}}} R_n^{-1} R_{f'},$$

where N^a_{eqv} is the number of symmetry equivalent atoms to the asymmetric unit atom a, and f' is the atom to which the atom f maps under the transformation R_n^{-1}. Again the second derivatives with respect to strain–strain and strain–internal are more complex since they are initially generated such that symmetry is violated. However, resymmetrization by averaging symmetry related matrix elements solves this problem.

The use of crystal symmetry in reciprocal space is even more straightforward than in real space. Because the Ewald sum can be written to be a sum over one-centered terms in reciprocal space, instead of a pairwise expression, the only change necessary is to restrict the sum to the asymmetric unit with appropriate weighting for site multiplicity. The same strategy is used in the symmetrization of other one-center terms in real space, such as the Einstein model. Crystal symmetry is also exploited in the calculation of charges via the electronegativity equalization method, which is thereby accelerated, especially through the reduction of the size of the matrix to be inverted.

Code Details

The original version of GULP [118] was written in Fortran 77 since the more recent standards had yet to be released. This implied that memory was statically allocated via a series of parameters. Subsequently, non-standard extensions were introduced to allow the second derivative arrays to be dynamically declared, since they represented the dominant use of memory. For the current version Fortran 90 has now been adopted leading to full use of dynamic memory.

The program has been compiled and tested for most Unix-style operating systems, including Linux and Apple-Macintosh OSX, using most Fortran 90 compilers. While compilation under MS-DOS is in principle possible, this operating system is not supported since it is the only operating system that cannot be automatically catered for within a single standard Makefile.

The code has also been parallelized for the evaluation of the energy and first derivatives using MPI, based upon a replicated data paradigm. When performing calculations on sufficiently large systems that require the use of parallel computers, then the most appropriate types of calculation are usually either conjugate gradient optimization or molecular dynamics. Hence, the absence of second derivatives is less critical. However, a distributed data algorithm for the second derivatives and using Scalapack for matrix diagonalization/inversion would be feasible and may be implemented in the future. Because GULP is currently targeted primarily at crystalline systems, where unit cells are typically small, the distribution of parallel work does not use a spatial decomposition. Instead the Brode–Ahlrichs algorithm [119] is used for the pairwise loops in real space in order to try to ensure load balancing over the processors. A similar approach is used for the four-body potentials based on the first two atoms of the sequence of four. In the case of three-body potentials, the work is divided by a straight distribution of pivot atoms over the nodes. Parallelization in reciprocal space is achieved by a equal division of reciprocal lattice vectors over nodes. Given that the number of operations per k-vector is equal, this should guarantee load balancing as long as the number of reciprocal lattice vectors is large compared with the number of processors (which is almost always the case).

REVIEW OF APPLICATIONS

There are two possible measures to gauge the success of the GULP simulation code. The first is to look at the number of times it has been cited in the open literature. Since 1997, when the first paper describing the symmetry-adapted algorithms employed in GULP was published, GULP has been cited over 300 times. A more reliable measure of success is to survey the breadth of areas where GULP has been employed as the calculation engine of choice.

Solid-state Structures

In solid-state chemistry, atomistic simulations have been long utilized to study the structures of oxides. GULP has been extensively used in this area to study both oxides with a simple stoichiometry, but a complex arrangement of ions within the lattice, such as in the case of transition aluminas, and also oxides with a more complex stoichiometry, such as high T_c superconductors.

Paglia and co-workers [70] have successfully simulated the structure of κ-alumina. In this system, there is an ordered oxide sublattice with 666 different ways to arrange the aluminum ions between

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

the oxide layers that are consistent with the experimentally observed space group. The lowest energy structure obtained by minimization of these 666 structures was found to agree with the experimentally determined structure. Watson and Willock [71] have studied the more complex γ-alumina, where information from crystallography and solid state NMR reduces the possible number of possible structures to just under 736,000. GULP was used to calculate the single point energies of these structures, which were used to estimate thermodynamic probabilities. Subsequent energy minimization on the most probable structures allowed the influence of configurational entropy on the structure to be assessed.

Going on to consider materials with even less long-range order, Nesper and co-workers [120] have calculated possible models for V_2O_5 nanotubes as an approximation to the more complicated wall structures which have been determined for the real VO_x nanoscrolls. Huang *et al.* [121] studied amorphous silica using a unit cell containing 1296 atoms and verified that their model is a good approximation to experiment via specific heat calculations using the normal vibrational modes.

For binary oxides of transition metals, the situation is more complicated with the presence of *d*-electrons, and simple interatomic models often fail. One strategy to correcting these failings is to utilize a model where ligand field effects are included through the adaptation of the AOM, as described previously. A further problem arises when modeling systems containing the same element in more than one oxidation state, such as shear defect phases. One solution to this is to employ the electronegativity equalization method which determines partial charges that depend on the local environment. Consequently there is no need for a prior assignment of species to lattice sites. Swamy and Gale [122] have applied this methodology to titanium oxides, including some of the Magneli phases, and found that the model is successful for bulk materials across different stoichiometries, polymorphs and Ti oxidation states.

One very important complex oxide system is the spinel family of structures, where there are issues such as the nature of the cation ordering between the A and B sites. Atomistic simulation can rapidly offer insight into such structural questions. Even GAs have been employed to predict spinel structures, though in this case it was found to be crucial to include cation exchange as a possible random step in order to avoid being trapped in the wrong local minimum [82]. The properties of spinel materials have also been extensively studied using GULP. Some of these simulations are of industrial relevance with applications ranging from high-temperature superconductors [123], zinc-based spinels as heat resistant coatings [124] and solid solutions [125,126]. Other ternary oxides that have been successfully simulated include $LiCoO_2$ [127], which is of interest as a battery material, and $Zr_2W_2O_8$ [128], which is an example of a material that demonstrates isotropic negative thermal expansion over a wide temperature range. In the case of the zirconium tungstate, free energy minimization was used, in conjunction with Rigid Unit Mode analysis, to reproduce the experimental observation and to explain how the contraction occurs.

GULP has also been used to simulate the structure and properties of a diverse range of inorganic salt families including nitrides [129–132], fluorides [133,134], nitrates [135,136], cuprates [137], tetrachlorozincates [138], titanyl phosphates and arsenates [139].

Systems containing organic moieties have also been simulated using GULP. One obvious extension of the inorganic solid-state work discussed above is to look at salts of organic molecular ions. Barretto *et al.* [140] have derived potentials for the acetate anion, which are able to model all the phases of lithium acetate dihydrate and anhydrous sodium acetate. Another class of solid-state compounds where organic–inorganic interactions are important is the area of intercalation chemistry. Fogg *et al.* [141] have investigated the intercalation compound made when gibbsite (aluminum hydroxide) is mixed with lithium chloride. The resulting structure consists of double hydroxide layers of formula $LiAl_2(OH)_6$ with the chloride ions sitting between the layers. These chloride anions can be easily exchanged for other inorganic as well as organic ions. GULP was used to rationalize the known 3-D structures containing inorganic anions. In addition, predictions of the structures when organic ions such as oxalate were intercalated were made, and the results found to be consistent with the observed interlayer spacings. Breu and co-workers [142] have used GULP to study $[Ru(bpy)_3]^{2+}$ and $[Ru(phen)_3]^{2+}$ intercalated into smectites. They found that the pillars formed between the layers consisted of clusters of the cations if a racemic mixture was used, due to favorable π stackings. They used the results from these simulations to rationalize the chiral discrimination exhibited by some achiral clays.

Polymers are an important class of organic solid-state materials. Corish and co-workers [143–147] have made extensive use of GULP to understand the mechanism of thermochromism in a poly(3-alkylthiophene). The change of polymer color at the critical temperature is believed to be due to the twisting of the conjugated polymer backbone. However, the details of this twist depends critically on the nature of the packing of the polymer chains, which in turn is dependent on the side chain attached to the polymer backbone.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

Atomistic techniques have been utilized to study these relationships in great detail.

Finally Breu *et al.* [148] have performed calculations on the solid state structures of the organometallic series $[M(bpy)_3](PF_6)_2$ (M = Ru, Ni, Zn). They chose this system as these compounds represent a system where a small change in molecular geometry induces qualitatively different crystal packing. They found that atomistic simulations were able to reproduce the different structures for these three compounds, but that it is crucial to input the most accurate molecular structures, correcting for bond shrinkage due to librational motion. They suggested, from these results, that routine polymorph prediction may be more severely hampered by inaccuracies in molecular structures than by inadequacies in the intermolecular potentials.

Defect Studies

GULP has been widely used to study defects. A particularly nice body of work examining the location of defects in binary oxides has been produced by Berry and co-workers. They started by simulating the incorporation of magnesium in α-Fe_2O_3 [149]. The most stable defect was found to be where a linear cluster of three Mg^{2+} ions replaces two Fe^{3+} ions. This agreed with their Rietveld refinement of the X-ray diffraction data. Buoyed with this success, they studied [150] the defect structure of tin and titanium doped in α-Fe_2O_3 using GULP as two different models, based on Rietveld refinement of X-ray data, existed in the literature. They found that the most stable structure was one where the dopant partially substituted iron sites and partially occupied empty interstitial oxygen sites. The other proposed model, where the dopant was present only at empty interstitial sites was less energetically stable. They subsequently examined computationally how lithium could be inserted into these metal doped iron oxides [151]. They found the configuration where a Li^+ occupies an interstitial site balanced by the reduction of a Fe^{3+} to Fe^{2+} was more favorable than the substitution of Fe^{3+} with Li^+ accompanied by an O^{2-} vacancy. The former model led to a decrease in the lattice parameters, which has been observed experimentally. Similar calculations on α-Cr_2O_3 have been undertaken [152] and both modeling via GULP and Rietveld refinement shows that the tin and titanium dopants occupy interstitial sites and partially substitute on octahedral chromium sites in the α-Cr_2O_3 structure, which is isostructural to α-Fe_2O_3. In contrast to α-Fe_2O_3, doping with magnesium results in the formation of spinel-related $MgCr_2O_4$. Their GULP calculations accounted for this difference by showing that not only is the formation of $MgCr_2O_4$ more favorable than the formation of $MgFe_2O_4$, but also that the formation of Mg-doped α-Cr_2O_3 is less favorable than that of Mg-doped α-Fe_2O_3.

A significant area of defect research is simulating the processes at play in lithium rechargeable batteries. These batteries have cathodes of transition metal oxides into which reversible intercalation of lithium is possible. The resulting battery can store high charge densities and has excellent cycling properties and thus are used in mobile phones and laptop computers. One possible anode material is V_2O_5 and it has been studied computationally by Braithwaite and co-workers [153]. GULP was used to predict the stable sites occupied by inserted lithium ions and changes to the unit cell dimensions as lithium ions are incorporated. They also studied the diffusion pathways of lithium ions in this system and calculated the activation energy, which is required if a fundamental understanding of the recharging process in these materials is to be attained. Similar understanding has been gained via molecular modeling on another cathode material $LiMn_2O_4$, which often has Mn substituted with other transition metals as it improves the cycling properties [154–156].

Another technologically important area where defect diffusion is important is the use of zircon and related structures to dispose of plutonium. In this area, in contrast to lithium ion batteries, a slow diffusion rate is required. Meis and Gale [157] computationally determined the migration pathways for the diffusion of uranium and plutonium in zircon and were thus able to deduce the diffusion coefficients, which were found to be in good agreement with the experimental data. These diffusion coefficients are extremely low suggesting that the actinides are essentially immobile, a prerequisite for any material which will be used to store plutonium. Williford and co-workers [158] have examined the energetics of the diffusion of native defects within zircon as these could also be involved in any migration pathway of actinides immobilized in zircon. They went on to study the location of tri and tetravalent plutonium in zircon [159] as well as gadolinium zirconate, another possible plutonium immobilizer [160]. Another possible immobilizing material, particularly for cesium are apatites. Chartier *et al.* [161] have utilized free energy calculations to find the preferential site for cesium in the three apatites they studied. They found that the calculated activation energies for cesium migration suggested that cesium diffusion is mainly controlled by intersite jumps to adjacent vacancies. They found that the apatite $Ca_2La_8(SiO_4)_6O_2$ proves to have the higher immobilization capacity, because of the high activation energies characterizing all the possible lattice diffusion mechanisms.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

Atomistic methods have long been associated with simulations of molecular sieves, since their structure drives so much of their chemistry. There are many examples in the literature, but we shall focus on contemporary issues. GULP has been used to probe the interactions between templating molecules and zeolite frameworks in an effort to understand how the templates drive the structure of the zeolite that forms around it [162]. Girard and co-workers have had great success predicting the final calcined structures of molecular sieves from a knowledge of the structure containing the templating molecules [163–166]. Extensive work exists on framework substitution in aluminophosphates with silicon [167,168], vanadium [169] and iron [170].

Significant defect research utilizing GULP has been performed on many other oxide containing systems including ceria [171], zirconia [172], wustite [173], pyrochlores [174,175], spinel type structures [176,177] and materials with the perovskite structure [178–183]. In addition, defects in fluorides have also been examined. Two fluoride systems studied with possible applications are β-PbF_2 as a scintillating detector [184], and trivalent rare earth ions in $BaLiF_3$ as a laser material [185].

Other Calculations

GULP has proven to be a popular component in quantum mechanics/molecular mechanics (QM/MM) embedding schemes. In these calculations, the area of interest (usually the site of a reaction where electron transfer is important) is simulated using quantum mechanics and the rest of the structure using molecular mechanics. Particular attention has to be paid to the boundary between the two calculation methodologies, which leads to different schemes. GULP seems to be particularly popular for treating the molecular mechanics part of the system as it contains several atomistic polarization models that often significantly interact with the quantum mechanically treated embedded cluster. The first scheme employing GULP used the INDO semi-empirical method for the embedded cluster [186] and examined the geometric and electronic structures and electron excitation energies for the one-center hole polaron state in KI and several possible intermediate states. This scheme has been updated to include *ab initio* and density functional treatments of the embedded cluster [187] and has recently been validated via comparison with other theoretical methods [188]. Sauer and Sierka [189] have developed a more general computer code QMPOT which can use a variety of programs, both commercial and academic, for both the quantum mechanical and molecular mechanical parts. GULP has been used for the molecular mechanics component in calculations on zeolites, including calculating zeolite acidities [190], calculating the rates of proton jumps between neighboring oxygen atoms in chabazite, faujasite and ZSM-5 [191], and determining the siting and coordination of CuI pairs in zeolite ZSM-5 [192].

GULP has also been used in a few simulations of surfaces. GULP and MARVIN began their development at the same time and at the same place (the Royal Institution of Great Britain) and thus many surface simulations have used GULP to generate the interatomic potentials and the relaxed bulk structure, which were then fed to MARVIN to perform surface simulations. However, MARVIN did not keep up with GULP in terms of available potential and calculation types. Thus, some authors who required the extra features coded in GULP have simulated surfaces by employing a slab that repeats in the z-direction with a vacuum gap. Warschkow *et al.* [193] have investigated the energetics of oxygen vacancies near the (0 0 0 1) surface of hematite. They used the transition state features of GULP to calculate barriers for several pathways of migration of oxygen vacancies in the bulk and at varying depths relative to the surface. They found that the energy barriers of several transitions were lowered considerably close to the surface. Swamy and co-workers used GULP to simulate the (0 0 1) and (1 1 0) surfaces of TiO_2 with their variable charge model discussed above [194].

The above offers pointers to just a few of the areas where atomistic simulation may be used and in particular where GULP has been cited to date. However, there are still many more areas that lie within the scope of the technique yet to be studied.

RESULTS

In this section we present a few illustrations of the application of the new functionality within GULP, including validation studies to compare with previous implementations.

Mechanical Properties

The range of mechanical, and related, properties computed by GULP has been significantly extended for the present version. Since no article on simulations of ionic materials would be complete without a mention of the ubiquitous and evergreen perennial MgO, we choose to take this well-studied system as an example.

Magnesium oxide adopted the cubic rock salt structure and possesses the well-known characteristic of exhibiting a Cauchy violation in the elastic constants (i.e. $C_{12} \neq C_{44}$). No simple two-body forcefield is capable of reproducing this many body effect. Consequently, it is necessary to use a breathing shell model to describe this material accurately.

TABLE VI Breathing shell model for magnesium oxide

Species 1	*Species 2*	*A (eV)*	*ρ (Å)*	C_6 *(eVÅ^6)*	k_{cs} *(eVÅ^{-2})*	k_{BSM} *(eVÅ^{-2})*
Mg core	O bshel	28.7374	0.3092	0.0	–	–
O shel	O shel	0.0	0.3	54.038	–	–
O core	O shel	–	–	–	46.1524	–
O shel	O bshel	–	–	–	–	351.439

Here "shel" denotes a potential that acts on the central position of the shell, while "bshel" denotes an interaction that acts on the radius of the shell which was fixed at 1.2 Å. The charge on Mg is +2, while the core and shell charges for O are +0.8 and −2.8, respectively.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

While there have been previous breathing shell models for MgO [195], we choose to fit a new set of potentials here that reproduce the structure, elastic constants and high and low frequency dielectric constants under ambient conditions. The resulting potential model is described in Table VI, while the calculated properties are given in Table VII.

The calculated properties for magnesium oxide can be seen to be in excellent agreement with experiment under ambient conditions, with the exception of the Poisson ratio. Of course, this agreement is a consequence of fitting a model with the correct essential physics to a subset of the experimental data. The disagreement in the Poisson ratios is because the values are calculated using different expressions. If the Poisson ratio is evaluated based on the sound velocities according to:

$$\sigma = \frac{\left(\frac{V_p}{V_s}\right)^2 - 2}{2\left[\left(\frac{V_p}{V_s}\right)^2 - 1\right]},$$

then our calculated value becomes 0.182 in good agreement with the determination of Zha *et al.* [196].

To provide a test of the model, it is possible to calculate the variation of the elastic properties of magnesium oxide as a function of applied pressure.

TABLE VII Calculated and experimental properties for magnesium oxide under ambient conditions

Properties	*Experiment*	*Calculated*
a (Å)	4.212	4.2123
C_{11} (GPa)	297.0	297.1
C_{12} (GPa)	95.2	95.1
C_{44} (GPa)	155.7	155.7
ϵ^0	9.86	9.89
ϵ^∞	2.96	2.94
K (GPa)	162.5	162.4
G (GPa)	130.4	130.9
V_s (km/s)	6.06	6.05
V_p (km/s)	9.68	9.70
σ	0.18	0.24

All experimental elastic properties are taken from Ref. [196]. Note, for the calculated bulk (K) and shear (G) moduli the Hill value is taken, though the variation between definitions is small.

The variation of the elastic constants up to 60 GPa is shown in Fig. 4.

When compared to the experimental results of Zha *et al.*, the calculated trend in the value of C_{11} is in good agreement in that it consistently increases under pressure and approximately doubles in magnitude by the time that 60 GPa is reached. Unfortunately, the trends for the other elastic constants are not so good, since the curve for C_{12} flattens with increasing pressure, rather than becoming steeper, and the curve for C_{44} passes through a maximum which is not observed in the experimental data from the aforementioned group. However, the calculated trends do match the extrapolated behavior based upon the results of ultrasonic measurements [197].

Born Effective Charges

The Born effective charges represent a useful means of characterizing the response of a potential model to perturbations, particularly those that create an electric field. Increasingly values are becoming available from *ab initio* techniques for solids through the application of linear response methods. Hence, this quantity can provide a useful comparison between the results of formal-charge shell model calculations and more accurate first principles methods.

One of the first materials for which the Born effective charges were determined using planewave techniques is α-quartz. In Table VIII the values obtained from the shell model of Sanders *et al.* [198] are compared with those yielded by a planewave calculation using the Local Density Approximation and norm-conserving pseudopotentials [199].

The comparison of the Born effective charge tensors demonstrates that the oxygen shell model is surprisingly successful at reproducing the quantum mechanical data, especially in comparison to rigid ion models, which would yield a diagonal matrix with all components equivalent. Furthermore, the polarizability of the shell leads to the ions behaving as partially charged species with realistic magnitudes. Consequently, this explains why the seemingly unreasonable use of a formal charge for

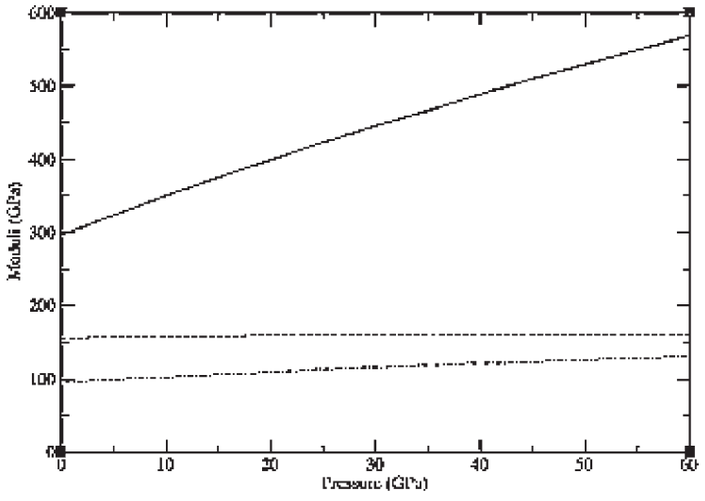


FIGURE 4 The variation of the elastic constants of magnesium oxide with applied pressure as determined by breathing shell model calculation. C_{11}, C_{12} and C_{44} are presented by a solid, dashed and dot-dashed line, respectively.

Si^{4+}actually works extremely well in practice. Similar observations have been previously made for perovskite materials [200].

Frequency-dependent Optical Properties

Ever since the early days of atomistic simulation within the shell model context it has been routine to calculate the static and high frequency dielectric constant tensors. Indeed this data has often been used during the fitting process as well. However, the values obtained from experiment will always be measured at a particular frequency and this will never exactly correspond to the limiting values determined by the direct means of calculation. As described in the background theory section, it is possible to evaluate the dielectric properties and refractive indices as a function of radiation frequency for the gamma point.

Here we present results for the frequency variation of the dielectric constant tensor of quartz, shown in Fig. 5, as calculated using the previously mentioned shell model potential of Sanders *et al.* [198]. Note that the limiting values are the same as the ones obtained without reference to the phonon frequencies.

In accord with experiment, the dielectric constant decreases slowly with the frequency of measurement until the highest phonon mode of quartz—the Si–O stretch-is approached. At frequencies below this the curve contains considerable variation caused by the discontinuities when a lattice phonon mode is reached. For simplicity, the curve shown is for the dielectric constant in the *ab* plane only. The corresponding curve for the principal tensor component parallel to the 0 0 1 direction is almost identical, except that the limiting values are slightly different and the positions of the discontinuities due to coincidence with phonons are displaced to higher frequency.

While the qualitative agreement with experimental data is good, there is a quantitative discrepancy in the dielectric constant variation. This is a result of the imperfection of the original fit to the dielectric data for quartz, though there are also issues concerning the variation with temperature since the calculations are formally performed at absolute zero, while the experiment data was measured at 293 K. However, the use of empirical fitting implies that the interatomic potentials do partially account for the temperature difference already. At the lowest measured frequency, the calculated values are

TABLE VIII Born effective charges (in a.u.) calculated according to the shell model of Sanders *et al.* [198] and from first principles techniques [199]

		Si			*O*	
Shell model	3.122	0.0	0.0	− 1.406	0.368	0.252
	0.0	3.530	0.292	0.364	− 1.920	− 0.517
	0.0	− 0.171	3.422	0.176	− 0.568	− 1.711
LDA	3.016	0.0	0.0	− 1.326	0.429	0.222
	0.0	3.633	0.282	0.480	− 1.999	− 0.718
	0.0	− 0.324	3.453	0.298	− 0.679	− 1.726

Values are shown for the asymmetric unit atoms with the approximate positions of the Si atom at (0.46, 0, 0) and the O atom at (0.41, 0.27, 0.11) in space group 154, with the origin set to (0, 0, 1/3).

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

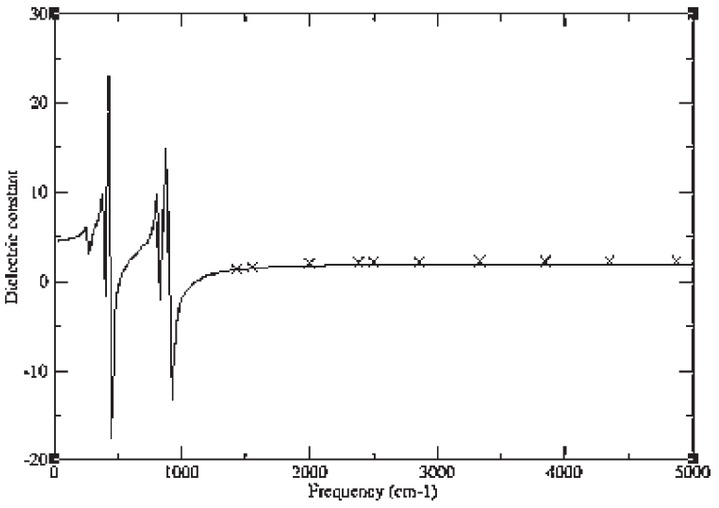


FIGURE 5 The on-diagonal component of the dielectric constant tensor for α-quartz in the *ab* plane as a function of frequency of measurement. The solid line represents the calculated shell model values, while the crosses represent values estimated from experimental measurements of the refractive index as a function of frequency.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

an almost perfect match due to the faster rate of decrease of the dielectric constant in the experiment data.

Surface Calculations

Before applying GULP to surfaces problems that were previously not possible, it must first be validated. Firstly, we focus on comparing our results to MARVIN, starting with the simple inorganic salt, corundum. In the original MARVIN paper [93], the twelve faces with the lowest interplanar spacings were identified and their surfaces relaxed using several different potential models. The calculations utilizing the QM5 model were repeated using both MARVIN with the BFGS minimizer and the same surfaces generated using GDIS and minimized in GULP using its BFGS minimizer. The unrelaxed surface and attachment energies were compared and both sets of calculations agreed to better than $0.0001\,\mathrm{J\,m^{-2}}$ for the surface energies and within $0.0001\,\mathrm{eV\,mol^{-1}}$ for the attachment energies. This indicates that the 2-D Ewald sum incorporated into GULP is correct. Next the relaxed surface and attachment energies were compared. Except for the (310) face, all the GULP calculations returned the same relaxed surface energy as the corresponding MARVIN run to within $0.0001\,\mathrm{J\,m^{-2}}$. The relaxed attachment energies agreed to within $0.003\,\mathrm{eV\,mol^{-1}}$. Excluding the (310) face, the MARVIN calculations took 138 s on a 1133 MHz Intel Pentium III CPU Linux system, whilst the GULP calculations took just 108 s. This difference in timing is primarily due to the faster energy calculation time in GULP since the MARVIN minimizer is based on the GULP algorithm and consequently the number of cycles to minimize a face in GULP and MARVIN only differed by at most one cycle, except for the (21-1) where GULP took 31 cycles versus the 24 for MARVIN. The (310) face is interesting as the relaxed surface energy calculated by GULP is the same as that reported in the original paper [93] and it is the minimized surface energy from the present calculation with MARVIN that is different. Although not stated in the MARVIN paper, the minimizations were done using a combination of minimizers; conjugate gradients until the gradient norm is unity, followed by a BFGS minimization. If the MARVIN calculations are repeated with this combination, all surface energies between MARVIN and GULP agree to within $0.0001\,\mathrm{J\,m^{-2}}$. In conclusion then, we can say that for this simple inorganic system, GULP and MARVIN produce essentially the same results.

As an example of the application of the new GULP surface capabilities, they have been recently utilized to search for surface reconstructions of the $(10\bar{1}4)$ cleavage plane of calcite [201]. The previous modeling studies have not found any evidence of surface reconstructions, despite the LEED results of Stipp [202]. A very efficient and assumption free way of finding reconstructions is to determine the surface phonon dispersion across the Brillouin zone, where the presence of any imaginary phonons will indicate that a reconstruction wants to occur. This calculation was performed on the cleavage plane of calcite, using a new calcite potential model that we have recently developed. An imaginary mode was found to be present at (1/2, 0) in reciprocal space, which indicates that the system is unstable within the (1×1) surface cell and that the reconstruction requires the formation of a (2×1) supercell. On creation of the surface supercell, the system was perturbed along the eigenvector of the imaginary mode and reminimized using the rational function optimization technique to ensure that

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

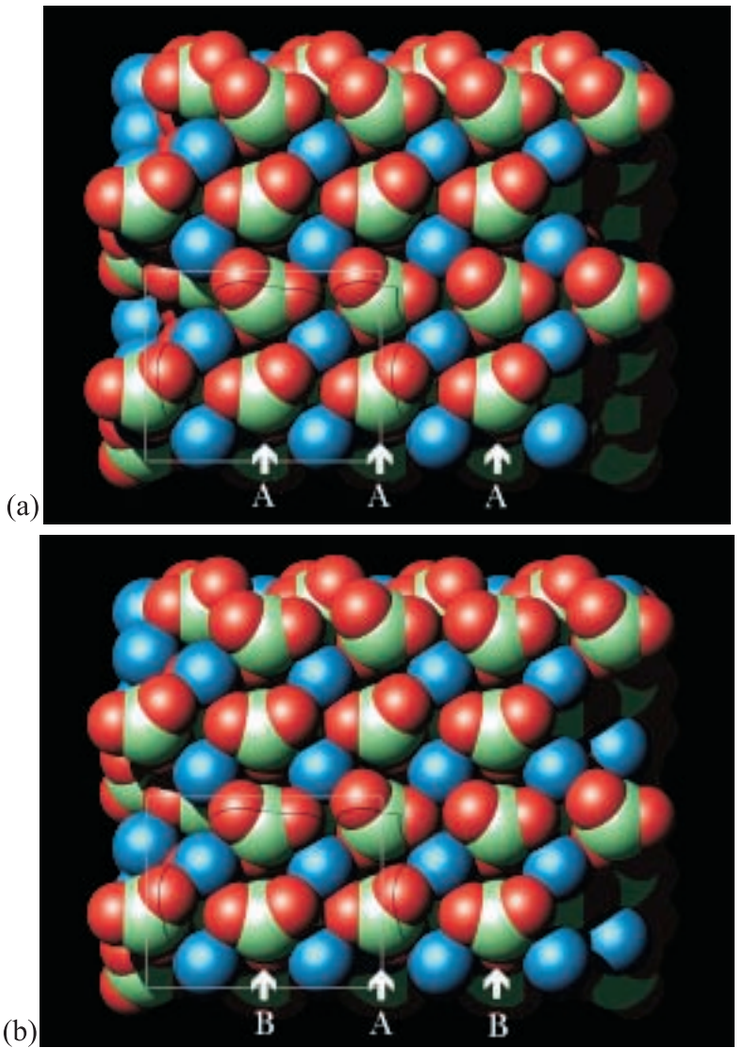


FIGURE 6 Comparison of the geometry for (a) a doubled unit cell of the (1 × 1) structure for the (10$\bar{1}$4) surface of calcite and (b) the optimized (2 × 1) reconstruction of the same surface. The reconstruction results in every second vertical row, labeled B, adopting a different configuration to the unreconstructed rows, labeled A.

the final Hessian matrix had the correct character for an energy minimum. Finally, the optimized (2 × 1) cell was examined to ensure that there were no imaginary phonon modes anywhere within the Brillouin zone. The calculations were repeated using other calcite models in the literature, which were found not to possess any imaginary modes.

In the reconstructed surface, as shown in Fig. 6, every second row of surface carbonate anions are found to rotate. The rotation moves the O atom of each carbonate such that it is 13° closer to the vertical. This is accompanied by significant change in the carbonate geometry with an increase of the O–C–O angle of 3°, where the two oxygen atoms are those pointing out of the surface and a compensating decrease in the other two O–C–O angles. Finally, we note that the reconstruction is confined to the top layer of the surface. In order to confirm the correct nature of the surface reconstruction, we have calculated the LEED pattern that would result, which is found to be an excellent match to the experimental pattern measured by Stipp under low pressure conditions.

Bond-order Potentials

Given that a new implementation of the Brenner model has been introduced into GULP we present here some results for diamond, Table IX, as previously studied in the original paper [63], in order to validate the code. For the second derivative properties, there is also the difference that the values obtained here are fully analytic whereas the published values are obtained via finite differences.

TABLE IX Calculated properties of diamond based on the model of Brenner *et al.* [63]

Property	*Experiment*	*Original value*	*GULP value*
Bond length (Å)	1.54	1.54	1.5441
Bond energy (eV)	3.68	3.68	3.684
C_{11} (GPa)	10.8	10.7	10.75
C_{12} (GPa)	1.3	1.0	1.26
C_{44} (GPa)	5.8	6.8	7.21

Note that the calculated values for the bond length and energy are not quoted in the original reference. However, since the experimental values were part of the fitted data, we take these values to be equal to the observed ones.

This may explain the small discrepancies in the elastic constants.

As stated in the background section, two algorithms have been implemented for the evaluation of the Brenner potential based on either pairwise atom looping or a spatial decomposition in order to determine the neighbor list. The computational cost of the two approaches for increasing supercells of diamond are shown in Fig. 7. The linear scaling behavior of the spatial decomposition can clearly be seen, as can the fact that this algorithm is effectively superior in performance for systems containing beyond a 100 atoms. Obviously, this trade off point is dependent on the density of the system, though there are few cases for hydrocarbons where the density is greater than that of diamond. For systems much below a 100 atoms the cost of evaluating the Brenner potential is so negligible in comparison to other tasks, such as a matrix inversion for property calculation, that the choice of algorithm is irrelevant.

CONCLUSIONS

The current state of the program GULP has been described above, including an extensive range of new facilities. Most notably the program has been extended to be able to treat systems of all dimensionalities from 0- to 3-D. Hence, it is now possible to model clusters, embedded defects, polymers, surfaces and bulk materials. A greater range of forcefields can now be used, including ones that allow for bond breaking and ligand field effects, and the number of properties that can be determined has significantly increased. However, there are many further developments in progress and the list of facilities available is nearly always advancing to reflect the advancements in the field. Internally the program has been greatly modified through the use of Fortran90 and all memory is now allocated dynamically for greater convenience.

Perhaps the greatest challenge in the area of atomistic simulation, and therefore also the most exciting prospect, is the development of reactive forcefields that are capable of correctly describing bond dissociation, at least qualitatively, if not semi-quantitatively. There has already been significant progress in this area with the bond order potentials of Brenner and co-workers, and more recently with the ReaxFF model from the group of Goddard [203]. Currently, parameters are only published for hydrocarbon systems, and so there is a long way to go before a full forcefield is available for the entire periodic table. However, they demonstrate that this is an important way forward, which will flourish through generally available software, so work is underway to implement such forcefields.

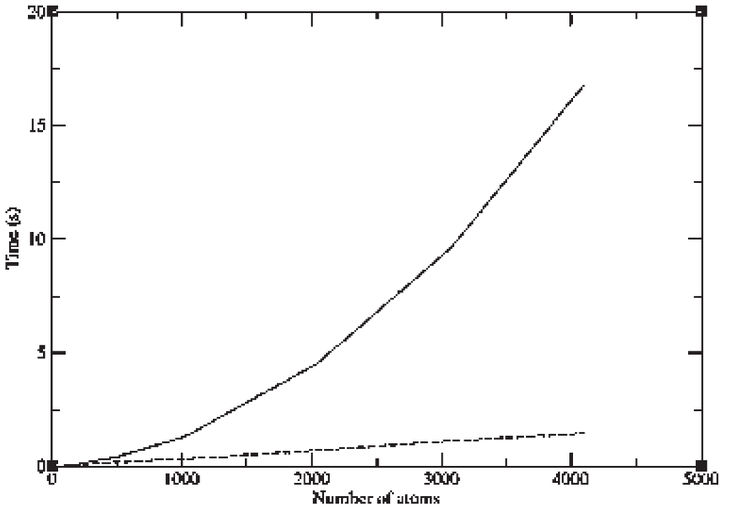


FIGURE 7 A comparison of the computer times required for a single energy-force evaluation using the Brenner model according to whether the algorithm used involves a pairwise sum (solid line) or a spatial decomposition (dashed line) to evaluate the neighbor list. Timings given are for a Mac PowerBook G4 laptop with a clock speed of 700 MHz.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

Acknowledgements

We would like to thank all our friends and collaborators who have contributed to both the work being reviewed here, and the field in general. In particular, we would like to thank Prof. Richard Catlow for his inspiration and guidance during the development of GULP. Special thanks must also go to Drs Sean Fleming and David Gay for their contributions to the surface functionality, Dr Scott Woodley for coding the AOM, and Dr Nilesh Raj who contributed to the parallelization of the program. Also we would like to thank Dr Kate Wright for suggestions regarding the manuscript. Funding from the Royal Society, EPSRC, and ARC is also gratefully acknowledged.

References

[1] Madelung, E. (1918) "Das elektrische feld in systemen von regelmaessig angeordneten punktladungen", *Phys. Z.* **19**, 524–532.
[2] Born, M. and Mayer, J.E. (1932) "Zur gittertheorie der ionenkristalle", *Z. Phys.* **75**, 1–18.
[3] Kapustinskii, A.F. (1956) "Lattice energy of ionic crystals", *Q. Rev. Chem. Soc.* **10**, 283–294.
[4] Norgett, M.J. (1972) A user's guide to HADES. Technical Report AERE-R7015, AERE Harwell Laboratory.
[5] Harding, J.H. (1988) A guide to MIDAS. Technical Report AERE-R13127, AERE Harwell Laboratory.
[6] Harding, J.H. (1982) A guide to the Harwell PLUTO program. Technical Report AERE-R10546, AERE Harwell Laboratory.
[7] Leslie, M. (1984) CASCADE. Technical Report DLSCITM36T, Daresbury Laboratory.
[8] Parker, S.C. and Price, G.D. (1989) "Computer modelling of phase transitions in minerals", *Adv. Solid State Chem.* **1**, 295–327.
[9] Willock, D.J., Price, S.L., Leslie, M. and Catlow, C.R.A. (1995) "The relaxation of molecular crystal structures using a distributed multipole electrostatic model", *J. Comp. Chem.* **16**, 628–647.
[10] Busing, W.R. (1981) WMIN, A computer program to model molecules and crystals in terms of potential energy functions. Technical Report ORNL-5747, Oak Ridge National Laboratory, 1981.
[11] Williams, D.E. (1972) "Molecular packing analysis", *Acta. Cryst. A* **28**, 629–635.
[12] Eckold, G., Stein-Arsic, M. and Weber, H.J. (1987) "UNISOFT—A program package for lattice dynamical calculations", *J. Appl. Cryst.* **20**, 134–139.
[13] van Gunsteren, W.F. and Berendsen, H.J.C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, Technical report (Biomos, Groningen).
[14] Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, III, T.E., DeBolt, S., Ferguson, D., Seibel, G. and Kollman, P. (1995) "AMBER, A package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules", *Comput. Phys. Commun.* **91**, 1–41.
[15] Brooks, B.R., Bruccoleri, R.E., Olfson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations", *J. Comp. Chem.* **4**, 187–217.
[16] Smith, W. and Forester, T.R. (1996) "Dl_poly_2.0: a general purpose parallel molecular dynamics simulation package", *J. Mol. Graph.* **14**, 136–141.
[17] Smith, W., Yong, C.W. and Rodger, P.M. (2002) "Dl_poly: application to molecular simulation", *Mol. Simul.* **28**, 385–471.
[18] Watson, G., Kelsey, E.T., de Leeuw, N.H., Harris, D.J. and Parker, S.C. (1996) "Atomistic simulation of dislocations, surfaces and interfaces in MgO", *J. Chem. Soc. Faraday Trans.* **92**, 433–438.
[19] Taylor, M.B., Barrera, G.D., Allan, N.L., Barron, T.H.K. and Mackrodt, W.C. (1998) "SHELL: A code for lattice dynamics and structure optimisation of ionic crystals", *Comp. Phys. Commun.* **109**, 135–143.
[20] Gale, J.D. (1997) "Semi-empirical methods as a tool in solid state chemistry", *Faraday Discuss.* **106**, 219–232.
[21] Evjen, H.M. (1932) "On the stability of certain heteropolar crystals", *Phys. Rev.* **39**, 675–687.
[22] Ewald, P.P. (1921) "Die berechnung optischer und elektrostatisher gitterpotentiale", *Ann. Phys.* **64**, 253–287.
[23] Tosi, M.P. (1964) "Cohesion of ionic solids in the born model", *Solid State Phys.* **16**, 1–120.
[24] Jackson, R.A. and Catlow, C.R.A. (1988) "Computer simulation studies of zeolite structure", *Mol. Simul.* **1**, 207–224.
[25] Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) "A smooth particle mesh Ewald method", *J. Chem. Phys.* **103**, 8577–8595.
[26] Greengard, L. and Rokhlin, V. (1987) "A fast algorithm for p@article simulations", *J. Comput. Phys.* **73**, 325–348.
[27] Petersen, H.G., Soelvason, D., Perram, J.W. and Smith, E.R. (1994) "A very fast multipole method", *J. Chem. Phys.* **101**, 8870–8876.
[28] de Leeuw, S.W., Perram, J.W. and Smith, E.R. (1980) "Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants", *Proc. R. Soc. Lond. Ser. A* **373**, 27–56.
[29] Leslie, M. and Gillan, M.J. (1985) "The energy and elastic dipole tensor of defects in ionic-crystals calculated by the supercell method", *J. Phys. C Solid State Phys.* **18**, 973–982.
[30] Parry, D.E. (1975) "The electrostatic potential in the surface region of an ionic crystal", *Surf. Sci.* **49**, 433–440.
[31] Parry, D.E. (1976) "Errata: the electrostatic potential in the surface region of an ionic crystal", *Surf. Sci.* **54**, 195–195.
[32] Wasserman, E., Rustad, J.R. and Felmy, A.R. (1999) "Molecular modeling of the surface charging of hematite I. The calculation of proton affinities and acidites on a surface", *Surf. Sci.* **424**, 19–27.
[33] Harding, J.H. (1988) A guide to MIDAS. Technical Report AERE R-13127, AERE Harwell Laboratory.
[34] Arnold, A. and Holm, C. (2002) "A novel method for calculating electrostatic interactions in 2D periodic slab geometries", *Chem. Phys. Lett.* **354**, 324–330.
[35] Saunders, V.R., Freyia-Fava, C., Dovesi, R. and Roetti, C. (1994) "On the electrostatic potential in linear periodic polymers", *Comp. Phys. Commun.* **84**, 156–172.
[36] Wolf, D., Keblinski, P., Philpot, S.R. and Eggebrecht, J. (1999) "Exact method for the simulation of Coulombic systems by spherically truncated, pairwise r^{-1} summation", *J. Chem. Phys.* **110**, 8254–8282.
[37] Mulliken, R.S. (1955) "Electronic population analysis on LCAO-MO molecular wave functions. I", *J. Chem. Phys.* **23**, 1833–1840.
[38] Baroni, S., de Gironcoli, S., Dal Corso, A. and Giannozzi, P. (2001) "Phonons and related crystal properties from density-functional perturbation theory", *Rev. Mod. Phys.* **73**, 515–562.
[39] Sanderson, R.T. (1951) "An interpretation of bond lengths and a classification of bonds", *Science* **114**, 670–672.
[40] van Genechten, K., Mortier, W. and Geerlings, P. (1987) "Intrinsic framework electronegativity: a novel concept in solid state chemistry", *J. Chem. Phys.* **86**, 5063–5071.
[41] Rappe, A.K. and Goddard, III, W.A. (1991) "Charge equilibration for molecular dynamics simulations", *J. Phys. Chem.* **95**, 3358–3363.
[42] Njo, S.L., Fan, J.F. and van de Graaf, B. (1998) "Extending and simplifying the electronegativity equalization method", *J. Mol. Catal. A* **134**, 79–88.
[43] Williams, D.E. (1989) "Accelerated convergence treatment of r^{-n} lattice sums", *Cryst. Rev.* **2**, 3–25.
[44] Tang, K.T. and Toennies, J.P. (1984) "An improved simple-model for the van der Waals potential based on universal damping functions for the dispersion coefficients", *J. Chem. Phys.* **80**, 3726–3741.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

[45] Applequist, J. (1985) "A multipole interaction theory of electric polarization of atomic and molecular assemblies", *J. Chem. Phys.* **83**, 809–826.

[46] Beyer, T., Day, G.M. and Price, S.L. (2001) "The prediction, morphology, and mechanical properties of the polymorphs of paracetamol", *J. Am. Chem. Soc.* **123**, 5086–5094.

[47] Madden, P.A. and Wilson, M. (1996) "'Covalent' effects in 'ionic' systems", *Chem. Soc. Rev.*, 339–350.

[48] Harding, J.H. and Pyper, N.C. (1995) "The meaning of the oxygen 2nd-electron affinity and oxide potential models", *Phil. Mag. Lett.* **71**, 113–121.

[49] Dick, B.G. and Overhauser, A.W. (1958) "Theory of the dielectric constants of alkali halide crystals", *Phys. Rev.* **112**, 90–103.

[50] Car, R. and Parinello, M. (1985) "Unified approach for molecular-dynamics and density-functional theory", *Phys. Rev. Lett.* **55**, 2471–2474.

[51] Lindan, P.J.D. and Gillan, M.J. (1993) "Shell-model molecular-dynamics simulation of superionic conduction in CaF_2", *J. Phys. Condens. Matter* **5**, 1019–1030.

[52] Schroeder, U. (1966) "A new model for lattice dynamics (breathing shell model)", *Solid State Commun.* **4**, 347–349.

[53] Axilrod, P.M. and Teller, E. (1943) "Interaction of the van der Waals type between three atoms", *J. Chem. Phys.* **11**, 299–300.

[54] Banerjea, A. and Smith, J.R. (1988) "Origins of the universal binding-energy relation", *Phys. Rev. B* **37**, 6632–6645.

[55] Sutton, A.P. and Chen, J. (1990) "Long-range Finnis–Sinclair potentials", *Phil. Mag. Lett.* **61**, 139–146.

[56] Finnis, M.W. and Sinclair, J.E. (1984) "A simple empirical n-body potential for transition metals", *Phil. Mag. A* **50**, 45–55.

[57] Cai, J. and Ye, Y.Y. (1996) "Simple analytical embedded-atom-potential model including a long-range force for fcc metals and their alloys", *Phys. Rev. B* **54**, 8398–8410.

[58] Abell, G.C. (1985) "Empirical chemical pseudopotential theory of molecular and metallic bonding", *Phys. Rev. B* **31**, 6184–6196.

[59] Tersoff, J. (1986) "New empirical-model for the structural-properties of silicon", *Phys. Rev. Lett.* **56**, 632–635.

[60] Tersoff, J. (1988) "New empirical-approach for the structure and energy of covalent systems", *Phys. Rev. B* **37**, 6991–7000.

[61] Pettifor, D.G., Oleinik, I.I., Nguyen-Manh, D. and Vitek, V. (2002) "Bond-order potentials: bridging the electronic to atomistic modelling hierarchies", *Comp. Mater. Sci.* **23**, 33–37.

[62] Brenner, D.W. (1990) "Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films", *Phys. Rev. B* **42**, 9458–9471.

[63] Brenner, D.W., Shenderova, O.A., Harrison, J.A., Stuart, S.J., Ni, B. and Sinnott, S.B. (2002) "A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons", *J. Phys.: Condens. Matter* **14**, 783–802.

[64] Che, J., Cagin, T. and Goddard, III, W.A. (1999) "Generalized extended empirical bond-order dependent force fields including nonbond interactions", *Theor. Chem. Acc.* **102**, 346–354.

[65] Jorgensen, C.K., Pappalarado, R. and Schmidtke, H.H. (1963) "Do the ligand field parameters in lanthanides represent weak covalent bonding?", *J. Chem. Phys.* **39**, 1422–1430.

[66] Woodley, S.M., Battle, P.D., Catlow, C.R.A. and Gale, J.D. (2001) "Development of a new forcefield that reproduces the John–Teller disorted structure of $LaMnO_3$", *Chem. Commun*, B **105**, 1879–1880.

[67] Woodley, S.M., Catlow, C.R.A., Battle, P.D. and Gale, J.D. (2002) "Development of a new interatomic potential for the modeling of ligand field effects", *J. Phys. Chem. B* **105**, 6824–6830.

[68] Einstein, A. (1907) "Die Plancksche theorie der strahlung und die theorie der spezifischen waerme", *Ann. Phys.* **22**, 180–190.

[69] Frenkel, D. and Ladd, A.J.C. (1981) "New Monte Carlo method to compute the free energy of arbitrary solids. Application to the fcc and hcp phases of hard spheres", *J. Chem. Phys.* **81**, 3188–3193.

[70] Paglia, G., Rohl, A.L., Buckley, C.E. and Gale, J.D. (2001) "A computational investigation of the structure of κ-alumina using interatomic potentials", *J. Mater. Chem.* **11**, 3310–3316.

[71] Watson, G.W. and Willock, D.J. (2001) "The enumeration of structures for γ-alumina based on a defective spinel structure", *Chem. Commun.*, 1076–1077.

[72] Allan, N.L., Barrera, G.D., Lavrentiev, M.Y., Todorov, I.T. and Purton, J.A. (2001) "*Ab initio* calculation of phase diagrams of ceramics and minerals", *J. Mater. Chem.* **11**, 63–68.

[73] van Laarhoven, P.J.M. and Aarts, E.H.L. (1987) *Simulated Annealing: Theory and Applications* (Reidel, Dordrecht).

[74] Holland, J.H. (1975) *Adaption in Natural and Artificial Systems* (The University of Michigan Press, Ann Arbor).

[75] Bush, T.S., Catlow, C.R.A. and Battle, P.D. (1995) "Evolutionary programming techniques for predicting inorganic crystal-structures", *J. Mater. Chem.* **5**, 1269–1272.

[76] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1986) *Numerical Recipes in Fortran: The Art of Scientific Computing* (Cambridge University Press, Cambridge, MA).

[77] Fletcher, R. and Powell, M.J.D. (1964) "A rapidly convergent descent method for minimization", *Comput. J.* **6**, 163–168.

[78] Shanno, D.F. (1970) "Conditioning of quasi-Newton methods for function minimization", *Math. Comput.* **24**, 647–656.

[79] Mills, G., Jonsson, H. and Schenter, G.K. (1995) "Reversible work transition-state theory-application to dissociative adsorption of hydrogen", *Surf. Sci.* **324**, 305–337.

[80] Banerjee, A., Adams, N., Simons, J. and Shepard, R. (1985) "Search for stationary points on surfaces", *J. Phys. Chem.* **89**, 52–57.

[81] Voter, A.F., Montalenti, F. and Germann, T.C. (2002) "Extending the time scale in atomistic simulation of materials", *Ann. Rev. Mater. Res.* **32**, 321–346.

[82] Woodley, S.M., Battle, P.D., Gale, J.D. and Catlow, C.R.A. (1999) "The prediction of inorganic crystal structures using a genetic algorithm and energy minimization", *Phys. Chem. Chem. Phys.* **1**, 2535–2542.

[83] Frenkel, D. and Smit, B. (2001) *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, New York).

[84] Nye, J.F. (1957) *Physical Properties of Crystals* (Oxford University Press, Oxford).

[85] Akitt, J.W. (1983) *NMR and Chemistry* (Chapman and Hall, London).

[86] Bader, R.F.W., Nguyendang, T.T. and Tal, Y. (1981) "A topological theory of molecular-structure", *Rep. Prog. Phys.* **44**, 893–948.

[87] Cochran, W. and Cowley, R.A. (1962) "Dielectric constants and lattice vibrations", *J. Phys. Chem. Solids* **23**, 447–450.

[88] Seddon, J.M. and Gale, J.D. (2002) *Thermodynamics and Statistical Mechanics* (Royal Society of Chemistry, London).

[89] Baldereschi, A. (1973) "Mean-value point in the Brillouin zone", *Phys. Rev. B* **7**, 5212–5215.

[90] Chadi, D.J. and Cohen, M.L. (1973) "Special points in the Brillouin zone", *Phys. Rev. B* **8**, 5747–5753.

[91] Monkhorst, H.J. and Pack, J.D. (1976) "Special points for Brillouin-zone integrations", *Phys. Rev. B* **13**, 5188–5192.

[92] Ramirez, R. and Boehm, M.C. (1988) "The use of symmetry in reciprocal space integrations—asymmetric units and weighting factors for numerical-integration procedures in any crystal symmetry", *Int. J. Quantum Chem.* **34**, 571–594.

[93] Gay, D.H. and Rohl, A.L. (1995) "MARVIN: A new computer code for studying surfaces and interfaces and its application to calculating the crystal morphologies of corundum and zircon", *J. Chem. Soc. Faraday Trans.* **91**, 926–936.

[94] Tasker, P.W. (1979) "The stability of ionic crystal surfaces", *J. Phys. C* **12**, 4977–4984.

[95] Donnay, J.D.H. and Harker, D. (1937) "A new law of crystal morphology extending the law of Bravais", *Am. Miner.* **22**, 446–467.

[96] Wulff, G. (1901) "Zur frage der geschwindigkeitb des wachsthums und der auflosung der krystallflachen", *Z. Kryst.* **34**, 449–530.

[97] Gibbs, J.W. (1928) *Collected Works* (Longman, New York).

[98] Hartman, P. and Bennema, P. (1980) "The attachment energy as a habit controlling factor 1. Theoretical considerations", *J. Crystal Growth* **49**, 145–156.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

[99] Allen, M.P. and Tildesley, D.J. (1987) *Computer Simulation of Liquids* (Oxford University Press, Oxford).

[100] LeSar, R., Najafabati, R. and Srolovitz, D.J. (1991) "Thermodynamics of solid and liquid embedded-atom-method metals—a variational study", *J. Chem. Phys.* **94**, 5090–5097.

[101] Sutton, A.P. (1992) "Direct free-energy minimization methods—application to grain-boundaries", *Phil. Trans. R. Soc. Lond. A* **341**, 233–245.

[102] Montroll, E.W. (1942) "Frequency spectrum of crystalline solids", *J. Chem. Phys.* **10**, 218–229.

[103] Kantorovich, L.N. (1995) "Thermoelastic properties of perfect crystals with nonprimitive lattices 1. General-theory", *Phys. Rev. B* **51**, 3520–3534.

[104] Taylor, M.B., Barrera, G.D., Allan, N.L. and Barron, T.H.K. (1997) "Free-energy derivatives and structure optimization with quasiharmonic lattice dynamics", *Phys. Rev. B* **56**, 14380–14390.

[105] Gale, J.D. (1998) "Analytical free energy minimization of silica polymorphs", *J. Phys. Chem. B* **102**, 5423–5431.

[106] Allan, N.L., Barron, T.H.K. and Bruno, J.A.O. (1996) "The zero static internal stress approximation in lattice dynamics, and the calculation of isotope effects on molar volumes", *J. Chem. Phys.* **105**, 8300–8303.

[107] Mott, N.F. and Littleton, M.J. (1938) "Conduction in polar crystals. I. Electrolytic conduction in solid salts", *Trans. Faraday Soc.* **34**, 485–499.

[108] Catlow, C.R.A., James, R., Mackrodt, W.C. and Stewart, R.F. (1982) "Defect energetics in (α-Al_2O_3 and rutile TiO_2", *Phys. Rev. B* **25**, 1006–1026.

[109] Allinger, N.L., Zhou, X.F. and Bergsma, J. (1994) "Molecular mechanics parameters", *J. Mol. Struct. (Theochem.)* **118**, 69–83.

[110] Rappe, A.K., Casewit, C.J., Colwell, K.S., Goddard, III, W.A and Skiff, W.M. (1992) "Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations", *J. Am. Chem. Soc.* **114**, 10024–10035.

[111] Barlow, S., Rohl, A.L., Shi, S.G., Freeman, C.M. and O'Hare, D. (1996) "Molecular mechanics study of oligomeric models for poly(ferro cenylsilanes) using the extensible systematic forcefield (ESFF)", *J. Am. Chem. Soc.* **118**, 7578–7592.

[112] Gale, J.D., Catlow, C.R.A. and Mackrodt, W.C. (1992) "Periodic *ab initio* determination of interatomic potentials for alumina", *Model. Simul. Mater. Sci. Eng.* **1**, 73–81.

[113] Gordon, R.G. and Kim, Y.S. (1972) "Theory for the forces between closed-shell atoms and molecules", *J. Chem. Phys.* **56**, 3122–3133.

[114] Gale, J.D. (1996) "Empirical potential derivation for ionic materials", *Phil. Mag. B* **73**, 3–19.

[115] Grey, T.J., Gale, J.D. and Nicholson, D. (2000) "Parameterization of a potential function for the Ca^{2+}–Ne and Ca^{2+}–N_2 interactions using high-level *ab initio* data", *Mol. Phys.* **98**, 1565–1573.

[116] Catlow, C.R.A. and Norgett, M.J. (1976) Lattice structure and stability of ionic materials, Technical Report AERE-M2936. AERE Harwell Laboratory.

[117] Catlow, C.R.A. and Mackrodt, W.C. (1982) "Theory of simulation methods for lattice and defect energy calculations in crystals", *Lect. Notes Phys.* **166**, 3–20.

[118] Gale, J.D. (1997) "GULP: a computer program for the symmetry-adapted simulation of solids", *J. Chem. Soc. Faraday Trans.* **1**, 629–637.

[119] Brode, S. and Ahlrichs, R. (1986) "An optimized MD program for the vector computer Cyber-205", *Comp. Phys. Commun.* **42**, 51–57.

[120] Patzke, G.R., Krumeich, F. and Nesper, R. (2002) "Oxidic nanotubes and nanorods—anisotropic modules for a future nanotechnology", *Angew. Chem. Int. Ed.* **41**, 2446–2461.

[121] Huang, M.Z., Ouyang, L.Z. and Ching, W.Y. (1999) "Electron and phonon states in an ideal continuous random network model of (α-SiO_2 glass", *Phys. Rev. B* **59**, 3540–3550.

[122] Swamy, V. and Gale, J.D. (2000) "Transferable variable-charge interatomic potential for atomistic simulation of titanium oxides", *Phys. Rev. B* **62**, 5406–5412.

[123] Green, M.A., Dalton, M., Prassides, K., Day, P. and Neumann, D.A. (1997) "Lattice vibrations of the superconducting oxide spinels (Li, $Mg)_{1+x}$ Ti_{2-x} O_4", *J. Phys. Condens. Matter* **9**, 10855–10865.

[124] Pandey, R., Gale, J.D., Sampath, S.K. and Recio, J.M. (1999) "Atomistic simulation study of spinel oxides: zinc aluminate and zinc gallate", *J. Am. Ceram. Soc.* **82**, 3337–3341.

[125] Woodley, S.M., Catlow, C.R.A., Piszora, P., Stempin, K. and Wolska, E. (2000) "Computer modeling study of the lithium ion distribution in quaternary Li–Mn–Fe–O spinels", *J. Solid State Chem.* **153**, 310–316.

[126] Wolska, E., Catlow, C.R.A., Piszora, P. and Woodley, S.M. (2000) "Structure refinement of quaternary spinel oxides—experiments and modelling", *Comput. Chem.* **24**, 603–607.

[127] Hart, F.X. and Bates, J.B. (1998) "Lattice model calculation of the strain energy density and other properties of crystalline $LiCOO_2$", *J. Appl. Phys.* **83**, 7560–7566.

[128] Pryde, A.K.A., Dove, M.T. and Heine, V. (1998) "Simulation studies of $Zr_2W_2O_8$ at high pressure", *J. Phys.: Condens. Matter* **10**, 8417–8428.

[129] Ching, W.Y., Xu, Y.N., Gale, J.D. and Ruhle, M. (1998) "*Ab initio* total energy calculation of α and β silicon nitride and the derivation of effective pair potentials with application to lattice dynamics", *J. Am. Cream. Soc.* **81**, 3189–3196.

[130] Marian, C.M., Gastreich, M. and Gale, J.D. (2000) "Empirical two-body potentials for solid silicon nitride, boron nitride, and borosilazane modifications", *Phys. Rev. B* **62**, 3117–3124.

[131] Marian, C.M. and Gastreich, M. (2000) "A systematic theoretical study of molecular Si/N, B/N, and Si/B/N(H) compounds and parameterisation of a force-field for molecules and solids", *Theochem.—J. Mol. Struct.* **506**, 107–129.

[132] Zapol, P., Pandey, R. and Gale, J.D. (1997) "An interatomic potential study of the properties of gallium nitride", *J. Phys. Condens. Matter* **9**, 9517–9525.

[133] Valerio, M.E.G., Jackson, R.A. and de Lima, J.F. (2000) "Derivation of potentials for the rare-earth fluorides, and the calculation of lattice and intrinsic defect properties", *J. Phys. Condens. Matter* **12**, 7727–7734.

[134] Ayala, A.P. (2001) "Atomistic simulations of the pressure-induced phase transitions in BaF_2 crystals", *J. Phys. Condens. Matter* **13**, 11741–11749.

[135] Mort, K.A., Wilde, P.J. and Jackson, R.A. (1999) "Computer modelling of ammonium nitrate: I. Development of potentials and calculation of lattice properties", *J. Phys. Condens. Matter* **11**, 3967–3972.

[136] Sorescu, D.C. and Thompson, D.L. (2001) "Classical and quantum mechanical studies of crystalline ammonium nitrate", *J. Phys. Chem. A* **105**, 720–733.

[137] Eremin, N.N., Leonyuk, L.I. and Urusov, V.S. (2001) "Interatomic potentials for simulation of alkaline-earth cuprates", *J. Solid State Chem.* **158**, 162–168.

[138] Ferrari, E.S., Roberts, K.J., Thomson, G.B., Gale, J.D. and Catlow, C.R.A. (2001) "Interatomic potential parameters for potassium tetrachlorozincate and their application to modelling its phase transformations", *Acta Cryst. A* **57**, 264–271.

[139] Glasser, L. and Catlow, C.R.A. (1997) "Modelling phase changes in the potassium titanyl phosphate system", *J. Mater. Chem.* **7**, 2537–2542.

[140] Barreto, L.S., Mort, K.A., Jackson, R.A. and Alves, O.L. (2000) "Computer modelling of solid alkali metal carboxylates", *J. Phys. Condens. Matter* **12**, 9389–9394.

[141] Fogg, A.M., Rohl, A.L., Parkinson, G.P. and O'Hare, D. (1999) "Predicting guest orientations in layered double hydroxide intercalates", *Chem. Mater.* **11**, 1194–1200.

[142] Breu, J., Raj, N. and Catlow, C.R.A. (1999) "Chiral recognition among trisdiimine-metal complexes. Part 4. Atomistic computer modeling of $[Ru(bpy)_3]^{2+}$ and $[Ru(phen)_3]^{2+}$ intercalated into low charged smectites", *J. Chem. Soc. Dalton Trans.*, 835–845.

[143] Corish, J., Feeley, D.E., Morton-Blake, D.A., Beniere, F. and Marchetti, M. (1997) "Atomistic investigation of thermochromism in a poly(3-alkylthiophene)", *J. Phys. Chem. B* **101**, 10075–10085.

[144] Xie, H.W., Corish, J., Morton-Blake, D.A. and Aasmundtveit, K. (1999) "An atomistic simulation of thermochromic

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012

distortions in poly(3-butylthiophene)", *Radiat. Eff. Defects Solids* **151**, 261–266.
[145] Xie, H.W., Corish, J., Ali, S.G., Morton-Blake, D.A. and Aasmundtveit, K. (1999) "Lattice simulations of thermochromic distortions in poly(alkylthiophene)s", *Synth. Met.* **101**, 318–319.
[146] Xie, H.W., Corish, J. and Morton-Blake, D.A. (2000) "Thermochromic distortions in a poly(3-alkylthiophene): an atomistic simulation investigation", *Synth. Met.* **113**, 65–72.
[147] Xie, H.W., O'Dwyer, S., Corish, J. and Morton-Blake, D.A. (2001) "The thermochromism of poly(3-alkylthiophene)s: the role of the side chains", *Synth. Met.* **122**, 287–296.
[148] Breu, J., Domel, H. and Norrby, P.O. (2000) "Chiral recognition among trisdiimine-metal complexes, 7. Racemic compound formation versus conglomerate formation with $[M(bpy)_3](PF_6)_2$ (M = Ni, Zn, Ru): lattice energy minimisation and implications for structure prediction", *Eur. J. Inorg. Chem.* **11**, 2409–2419.
[149] Berry, F.J., Bohorquez, A., Greaves, C., McManus, J., Moore, E.A. and Mortimer, M. (1998) "Structural characterization of divalent magnesium-doped α-Fe_2O_3", *J. Solid State Chem.* **140**, 428–430.
[150] Berry, F.J., Bohorquez, A. and Moore, E.A. (1999) "Rationalisation of defect structure of tin- and titanium-doped α-Fe_2O_3 using atomistic simulation", *Solid State Commun.* **109**, 207–211.
[151] Moore, E.A., Widatallah, H.M. and Berry, F.J. (2002) "Prediction of defect structure of tin- and titanium-doped α-Fe_2O_3 using atomistic simulation", *J. Phys. Chem. Solids* **63**, 519–523.
[152] Ayub, I., Berry, F.J., Johnson, C., Johnson, D.A., Moore, E.A., Ren, X.L. and Widatallah, H.M. (2002) "Tin-, titanium-, and magnesium-doped α-Cr_2O_3: characterisation and rationalisation of the structures", *Solid State Commun.* **123**, 141–145.
[153] Braithwaite, J.S., Catlow, C.R.A., Harding, J.H. and Gale, J.D. (2001) "A theoretical study of lithium intercalation into V_6O_{13}—a combined classical and quantum mechanical approach", *Phys. Chem. Chem. Phys.* **3**, 4052–4059.
[154] Braithwaite, J.S., Catlow, C.R.A., Harding, J.H. and Gale, J.D. (2000) "A computational study of the high voltage $Li_xCO_yMn_{4-y}O_8$ cathode material", *Phys. Chem. Chem. Phys.* **2**, 3841–3846.
[155] Ammundsen, B., Burns, G.R., Islam, M.S., Kanoh, H. and Roziere, J. (1999) "Lattice dynamics and vibrational spectra of lithium manganese oxides: a computer simulation and spectroscopic study", *J. Phys. Chem. B* **103**, 5175–5180.
[156] Ammundsen, B., Islam, M.S., Jones, D.J. and Roziere, J. (1999) "Local structure and defect chemistry of substituted lithium manganate spinels: X-ray absorption and computer simulation studies", *J. Power Sources* **82**, 500–504.
[157] Meis, C. and Gale, J.D. (1998) "Computational study of tetravalent uranium and plutonium lattice diffusion in zircon", *Mat. Sci. Eng. B* **57**, 52–61.
[158] Williford, R.E., Weber, W.J., Devanathan, R. and Cormack, A.N. (1999) "Native vacancy migrations in zircon", *J. Nucl. Mater.* **273**, 164–170.
[159] Williford, R.E., Begg, B.D., Weber, W.J. and Hess, N.J. (2000) "Computer simulation of Pu^{3+} and Pu^{4+} substitutions in zircon", *J. Nucl. Mater.* **278**, 207–211.
[160] Williford, R.E. and Weber, W.J. (2001) "Computer simulation of Pu^{3+} and Pu^{4+} substitutions in gadolinium zirconate", *J. Nucl. Mater.* **299**, 140–147.
[161] Chartier, A., Meis, C. and Gale, J.D. (2001) "Computational study of Cs immobilization in the apatites $Ca_{10}(PO_4)_6F_2$, $Ca_4La_6(SiO_4)_6F_2$ and $Ca_2La_8(SiO_4)_6O_2$", *Phys. Rev. B* **64**, 085110.
[162] Sabater, M.J. and Sastre, G. (2001) "A computational study on the templating ability of the trispyrrolidinium cation in the synthesis of ZSM-18 zeolite", *Chem. Mater.* **13**, 4520–4526.
[163] Girard, S., Mellot-Draznieks, C., Gale, J.D. and Ferey, G. (2000) "A predictive computational study of $AlPO_4$-14: crystal structure and framework stability of the template-free $AlPO_4$-14 from it's as-synthesised templated form", *Chem. Commun.*, 1161–1162.
[164] Girard, S., Gale, J.D., Mellot-Draznieks, C. and Ferey, G. (2002) "Computational prediction of nanoporous inorganic structures upon template extraction and calcination: a theoretical study of gallophosphate polymorphs", *J. Am. Chem. Soc.* **124**, 1040–1051.
[165] Girard, S., Tuel, A., Mellot-Draznieks, C. and Ferey, G. (2002) "Computational prediction of the phase transformation of two as-synthesised oxyfluorinated compounds into the zeotype CHA forms", *Angew. Chem. Int. Ed.* **41**, 972–975.
[166] Loiseau, T., Mellot-Draznieks, C., Sassoye, C., Girard, S., Guillou, N., Huguenard, C., Taulelle, F. and Ferey, G. (2001) "Chemistry-structure-simulation or chemistry-simulation structure sequences? The case of MIL-34, a new porous aluminophosphate", *J. Am. Chem. Soc.* **123**, 9642–9651.
[167] Sastre, G. and Lewis, D.W. (1998) "Modelling of Broensted acidity in AFI and CHA zeotypes", *J. Chem. Soc. Faraday Trans.* **94**, 3049–3058.
[168] Sastre, G., Lewis, D.W. and Catlow, C.R.A. (1997) "Modeling of silicon substitution in SAPO-5 and SAPO-34 molecular sieves", *J. Phys. Chem. B* **101**, 5249–5262.
[169] Gulin-Gonzalez, J., Alcaz, J.D., Nieto, J.M.L. and de las Pozas, C. (2000) "A computer simulation study of vanadium substitution in the $AlPO_4$-5 framework", *J. Mater. Chem.* **10**, 2597–2602.
[170] Gulin-Gonzalez, J. and del Rio, C.D. (2002) "A computational study of the symmetry of an aluminophosphate microporous material. Incorporation of iron defects in aluminium tetrahedral sites", *Chem. Mater.* **14**, 2817–2825.
[171] Balducci, G., Islam, M.S., Kapar, J., Fornasiero, P. and Graziani, M. (2000) "Bulk reduction and oxygen migration in the ceria-based oxides", *Chem. Mater.* **12**, 677–681.
[172] Khan, M.S., Islam, M. and Bates, D.R. (1998) "Cation doping and oxygen diffusion in zirconia: a combined atomistic simulation and molecular dynamics study", *J. Mater. Chem.* **8**, 2299–2307.
[173] Haavik, C., Stolen, S., Hanfland, M. and Catlow, C.R.A. (2000) "Effect of defect clustering on the high-pressure behaviour of wustite. High-pressure X-ray diffraction and lattice energy simulations", *Phys. Chem. Chem. Phys.* **2**, 5333–5340.
[174] Chartier, A., Meis, C., Weber, W.J. and Corrales, L.R. (2002) "Theoretical study of disorder in Ti-substituted $La_2Zr_2O_7$", *Phys. Rev. B* **65**, 134116.
[175] Williford, R.E., Weber, W.J., Devanathan, R. and Gale, J.D. (1999) "Effects of cation disorder on oxygen vacancy migration in $Gd_2Ti_2O_7$", *J. Electroceram.* **3**, 409–424.
[176] De Souza, S.S., Ayres, F. and Blak, A.R. (2001) "Simulation models of defects in $MgAl_2O_4$: Fe^{2+}, Fe^{3+} spinels", *Radiat. Eff. Defects Solids* **156**, 311–316.
[177] Moore, E.A., Bohorquez, A., Berry, F.J., Helgason, O. and Marco, J.F. (2001) "Investigation of defect structures formed by doping tetravalent ions into inverse spinel-related iron oxides using atomistic simulation calculations", *J. Phys. Chem. Solids* **62**, 1277–1284.
[178] Davies, R.A., Islam, M.S. and Gale, J.D. (1999) "Dopant and proton incorporation in perovskite-type zirconates", *Solid State Ion.* **126**, 323–335.
[179] De Souza, R.A., Islam, M.S. and Ivers-Tiffee, E. (1999) "Formation and migration of cation defects in the perovskite oxide $LaMnO_3$", *J. Mater. Chem.* **9**, 1621–1627.
[180] Islam, M.S., Davies, R.A., Fisher, C.A.J. and Chadwick, A.V. (2001) "Defects and protons in the $CaZrO_3$ perovskite and $Ba_2In_2O_5$ brownmillerite: computer modelling and exafs studies", *Solid State Ion.* **145**, 333–338.
[181] Khan, M.S., Islam, M.S. and Bates, D.R. (1998) "Dopant substitution and ion migration in the $LaGaO_3$-based oxygen ion conductor", *J. Phys. Chem. B* **102**, 3099–3104.
[182] Davies, R.A., Islam, M.S., Chadwick, A.V. and Rush, G.E. (2000) "Cation dopant sites in the $CaZrO_3$ proton conductor: a combined EXAFS and computer simulation study", *Solid State Ion.* **130**, 115–122.
[183] Williford, R.E., Armstrong, T.R. and Gale, J.D. (2000) "Chemical and thermal expansion of calcium-doped lanthanum chromite", *J. Solid State Chem.* **149**, 320–326.
[184] Jiang, H.T., Costales, A., Blanco, M.A., Gu, M., Pandey, R. and Gale, J.D. (2000) "Theoretical study of native and rare-earth defect complexes in β-PbF_2", *Phys. Rev. B* **62**, 803–809.

[185] Jackson, R.A., Valerio, M.E.G. and de Lima, J.F. (2001) "Computer modelling of rare-earth dopants in $BaLiF_3$", *J. Phys. Condens. Matter* **13**, 2147–2154.
[186] Shluger, A.L. and Gale, J.D. (1996) "One-center trapping of the holes in alkali halide crystals", *Phys. Rev. B* **54**, 962–969.
[187] Sushko, P.V., Shluger, A.L. and Catlow, C.R.A. (2000) "Relative energies of surface and defect states: *ab initio* calculations for the MgO (0 0 1) surface", *Surf. Sci.* **450**, 153–170.
[188] Braithwaite, J.S., Sushko, P.V., Wright, K. and Catlow, C.R.A. (2002) "Hydrogen defects in forsterite: a test case for the embedded cluster method", *J. Chem. Phys.* **116**, 2628–2635.
[189] Sauer, J. and Sierka, M. (2000) "Combining quantum mechanics and interatomic potential functions in *ab initio* studies of extended systems", *J. Comput. Chem.* **21**, 1470–1493.
[190] Brandle, M. and Sauer, J. (1998) "Acidity differences between inorganic solids induced by their framework structure. A combined quantum mechanics molecular mechanics *ab initio* study on zeolites", *J. Am. Chem. Soc.* **120**, 1556–1570.
[191] Sierka, M. and Sauer, J. (2001) "Proton mobility in chabazite, faujasite, and ZSM-5 zeolite catalysts, comparison based on *ab initio* calculations", *J. Phys. Chem. B* **105**, 1603–1613.
[192] Spuhler, P., Holthausen, M.C., Nachtigallova, D., Nachtigall, P. and Sauer, J. (2002) "On the existence of Cu–I paris in ZSM-5—a computational study", *Chem. Eur. J.* **8**, 2099–2115.
[193] Warschkow, O., Ellis, D.E., Hwang, J.H., Mansourian-Hadavi, N. and Mason, T.O. (2002) "Defects and charge transport near the hematite (0 0 0 1) surface: an atomistic study of oxygen vacancies", *J. Am. Ceram. Soc.* **85**, 213–220.
[194] Swamy, V., Muscat, J., Gale, J.D. and Harrison, N.M. (2002) "Simulation of low index rutile surfaces with a transferable variable-charge Ti–O interatomic potential and comparison with *ab initio* results", *Surf. Sci.* **504**, 115–124.
[195] Catlow, C.R.A., Faux, I.D. and Norgett, M.J. (1976) "Shell and breathing shell model calculations for defect formation energies and volumes in magnesium oxide", *J. Phys. C: Solid State Phys.* **9**, 419–429.
[196] Zha, C.-S., Mao, H.-K. and Hemley, R.J. (2000) "Elasticity of MgO and a primary pressure scale to 55 GPa", *Proc. Natl Acad. Sci.* **97**, 13494–13499.
[197] Jackson, I. and Niesler, H. (1982) *High Pressure Research in Geophysics* (Center for Academic Publishing, Tokyo).
[198] Sanders, M.J., Leslie, M. and Catlow, C.R.A. (1984) "Interatomic potentials for SiO_2", *J. Chem. Soc. Chem. Commun.*, 1271–1273.
[199] Gonze, X., Allan, D.C. and Teter, M.P. (1992) "Dielectric tensor, effective charges, and phonons in α-quartz by variational density-functional perturbation theory", *Phys. Rev. Lett.* **68**, 3603–3606.
[200] Ghosez, Ph., Michenaud, J.-P. and Gonze, X. (1998) "Dynamical atomic charges: the case of ABO_3 compounds", *Phys. Rev. B* **58**, 6224–6240.
[201] Rohl, A.L., Wright, K. and Gale, J.D. (2003) "Evidence from surface phonons for the (2×1) reconstruction of the $(10\bar{1}4)$ surface of calcite from computer simulation", *Am. Miner.*, in press.
[202] Stipp, S.L.S. (1999) "Toward a conceptual model of the calcite surface: hydration, hydrolysis, and surface potential", *Geochim. Cosmochim. Acta* **63**, 3121–3131.
[203] van Duin, A.C.T., Dasgupta, S., Lorant, F. and Goddard, III, W.A. (2001) "ReaxFF: a reactive force field for hydrocarbons", *J. Phys. Chem. A* **105**, 9396–9409.

Downloaded by [University of Saskatchewan Library] at 23:59 11 September 2012