Chapter 2

Methodology

2.1 The Perfect Lattice

The description of the perfect lattice used in this work is derived from the classical Born model of solids [54, 55], in which the lattice is constructed as an effectively infinite array of charged, spherical ions.

Atoms, within a crystal, experience interactions with all other atoms within the structure. The general form of this interaction may be written as the series summation:

$$\Psi(r_1 ... r_n) = \sum_{i,j=1}^{n} \Psi_2(r_{ij}) + \sum_{i,j,k=1}^{n} \Psi_3(r_{ijk}) + \sum_{i,j,k,l=1}^{n} \Psi_4(r_{ijkl}) \tag{2.1}$$

where $\Psi_2(r_{ij})$ is the interaction between pairs of ions $\{ij\}$, $\Psi_3(r_{ijk})$ the interaction between triplets of ions $\{ijk\}$ etc. The nature of the expansion in equation 2.1 requires that it be truncated so as to be computationally tractable. In highly ionic materials pair interactions dominate and consequently the expansion is truncated after the first term [56]. This is the pair potential approximation and ensures that the interactions between ions can be described via a central force. The approximate total interaction is now

$$\Psi_a(r_1 ... r_n) = \sum_{i,j=1}^{n} \Psi_2(r_{ij}) \tag{2.2}$$

The species $\Psi_2(r_{ij})$ in equation 2.2 is the potential energy between ions i and j, and is subsequently written $E(r_{ij})$. The form of this pair potential is an important factor in deciding the effectiveness of the model. The charged nature of the ionic species give rise to a Coulombic interaction; the relatively slow decay of $\frac{1}{r}$ with increasing r means that this forms the long range component of the potential. The total potential is written:

$$E(r_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} + \phi_{sr} \tag{2.3}$$

where ϵ_0 is the permittivity of free space, q_i and q_j are the charges on ions i and j respectively and r_{ij} is their separation. The second term, ϕ_{sr}, describes any remaining, short-range interaction. This interaction is shown schematically in figure 2.1 for the case of two oppositely charged ions. Having defined the interaction between any two ions these can be summed to give the lattice energy, E_L:

$$E_L = \sum_{j>i} \sum_{i} \left(\frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} + \phi_{sr} \right) \tag{2.4}$$

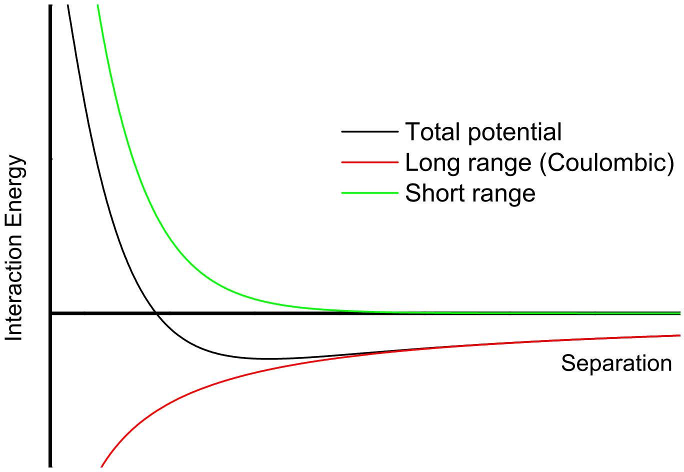

Figure 2.1: Total interaction energy between two oppositely charged ions showing the breakdown into long and short range components.

2.2 Ewald Method

The long range nature of the Coulomb term in equation 2.3 is highly inconvenient from a modelling perspective. The spatial interaction arising from this term typically falls off no faster than $\sim r^{-d}$ [57, 58] where d is the dimensionality of the system under consideration. This relatively slow decay precludes the truncation of the first term in equation 2.4 and is instead here overcome via applying the method due to Ewald [59].

Ewald derived a technique which sums the interactions between an ion and its array of periodic images. The original derivation is mathematically complex and so the following treatment is a simplification owing much to Kittel [60] (and therein attributed to an unpublished paper of Shockley and Ewald).

In the following derivation the lattice is assumed to consist of non-overlapping, spherical ions of positive or negative charge. The total potential at a given lattice point is written as:

$$\psi = \psi_1 + \psi_2 \tag{2.5}$$

where the potential ψ_1 is that of a lattice with a Gaussian charge distribution at each lattice point, with signs as in the real lattice. The potential ψ_2 is that of a lattice of point charges with an additional Gaussian charge distribution of opposite sign superimposed upon these point charges. When these two components are combined as in equation 2.5 they reduce to the original set of point charges. The point of splitting ψ in this way is that a parameter, η, can be optimised to determine the width of the Gaussian peaks such that both parts converge rapidly and independently. An optimum value for this width, determined by Catlow and Norgett [61], is:

$$\eta = \left(\frac{N\pi^3}{V^2} \right) \tag{2.6}$$

where N is the total number of charges and V is the unit cell volume. The definition of the Madelung constant dictates that the charge distribution at the reference point does not contribute to potentials ψ_1 and ψ_2, that is to say, ions do not experience their own electrostatic potential. It is therefore convenient to describe ψ_1 as the difference:

$$\psi_1 = \psi_a - \psi_b \tag{2.7}$$

between ψ_a, the potential of a continuous series of Gaussian distributions and ψ_b the potential of a single charge distribution at the reference point.

ψ_a and its associated charge density, ρ can be expanded in terms of Fourier series:

$$\psi_a = \sum_{\mathbf{G}} c_{\mathbf{G}} e^{i(\mathbf{G}.\mathbf{r})} \tag{2.8}$$

$$\rho = \sum_{\mathbf{G}} \rho_{\mathbf{G}} e^{i(\mathbf{G}.\mathbf{r})} \tag{2.9}$$

where $c_{\mathbf{G}}$ and $\rho_{\mathbf{G}}$ are coefficients and $\mathbf{G}$ is 2π times the set of reciprocal lattice vectors. The series converge as $\mathbf{G}$ increases and the magnitude of the coefficients decrease. The electrostatic potential is related to its charge distribution through Poisson's equation:

$$\nabla^2 \psi_a = -4\pi\rho \tag{2.10}$$

and so:

$$c_{\mathbf{G}} = 4\pi\rho_{\mathbf{G}}/G^2 \tag{2.11}$$

The Gaussian charge distribution for an ion, j, with charge q_j and half width $\left(\frac{ln2}{\eta}\right)^{\frac{1}{2}}$ is:

$$\rho = q_j \left(\frac{\eta}{\pi}\right)^{\frac{3}{2}} e^{-\eta r^{-2}} \tag{2.12}$$

To evaluate ρ_k now multiply both sides of equation 2.9 by $e^{-i(\mathbf{G}.\mathbf{r})}$ and integrate over the volume V of a single unit cell. In this case the charge distribution is that originating on the lattice points within the cell as well as the tails of distributions arising from all other lattice points. The integral of the total charge times $e^{-i(\mathbf{G}.\mathbf{r})}$ over a single unit cell is equal, however, to the integral of the charge density associated with a single lattice point times $e^{-i(\mathbf{G}.\mathbf{r})}$ over all space, so that:

$$\rho_{\mathbf{G}} V = \rho_{\mathbf{G}} \int_V e^{i(\mathbf{G}.\mathbf{r})} e^{-i(\mathbf{G}.\mathbf{r})} d\mathbf{r} = \left(\frac{\eta}{\pi}\right)^{\frac{3}{2}} \int_{V_\infty} \sum_t q_t e^{-\eta(r-r_t)^{-2}} e^{-i(\mathbf{G}.\mathbf{r})} d\mathbf{r} \tag{2.13}$$

This integral evaluates to:

$$\rho_{\mathbf{G}} V = \left(\sum_t e^{-i(\mathbf{G}.\mathbf{r})}\right) e^{-\frac{G^2}{4\eta}} = S(\mathbf{G}) e^{-\frac{G^2}{4\eta}} \tag{2.14}$$

where $S(\mathbf{G}) = \left(\sum_t e^{-i(\mathbf{G}.\mathbf{r})}\right)$ is the structure factor of the unit cell under consideration.

Substituting the above result into equation 2.8 via 2.11 gives:

$$\psi_a = \frac{4\pi}{V} \sum_{\mathbf{G}} S(\mathbf{G}) G^{-2} e^{i(\mathbf{G}.\mathbf{r}) - \frac{G^2}{4\eta}} \tag{2.15}$$

The contribution to ψ_b, the potential at the reference point due to the central Gaussian charge distribution is:

$$\psi_b = \int_0^\infty 4\pi r^2 \frac{\rho(\mathbf{r})}{\mathbf{r}} d\mathbf{r} = \frac{2q_i}{\epsilon_0}\left(\frac{\eta}{\pi}\right)^{\frac{1}{2}} \tag{2.16}$$

Combining equations 2.15 and 2.16 yields:

$$\psi_1 = \frac{4\pi}{V}\sum_{\mathbf{G}} S(\mathbf{G})G^{-2}e^{i(\mathbf{G}.\mathbf{r})-\frac{G^2}{4\eta}} - \frac{2q_i}{\epsilon_0}\left(\frac{\eta}{\pi}\right)^{\frac{1}{2}} \tag{2.17}$$

The potential ψ_2 consists of contributions from the tails of all Gaussians, other than that one centred on the reference point itself. The potential may be partitioned into three separate contributions per lattice point: the point charge associated with the ion j, situated at that lattice point, the part of the Gaussian charge distribution centred on j and falling within a sphere of radius r_j about this lattice point and finally that part of the Gaussian charge distribution centred on j that falls outside this sphere.

$$\begin{aligned} \psi_2 &= \frac{1}{4\pi\epsilon_0}\sum_j q_j\left[\frac{1}{r_j} - \frac{1}{r_j}\int_0^{r_j}\rho(\mathbf{r})d\mathbf{r} - \int_{r_j}^\infty \frac{\rho(\mathbf{r})}{r}d\mathbf{r}\right] && (2.18) \\ &= \frac{1}{4\pi\epsilon_0}\sum_j \frac{q_j}{r_j} erfc(\eta^{\frac{1}{2}} r_j) && (2.19) \end{aligned}$$

where $erfc(x)$, the complimentary error function, is of the form:

$$erfc(x) = \frac{2}{\pi^{\frac{1}{2}}}\int_x^\infty e^{t^2} dt \tag{2.20}$$

Finally we can write:

$$\psi = \frac{4\pi}{V}\sum_{\mathbf{G}} S(\mathbf{G})G^{-2}e^{i(\mathbf{G}.\mathbf{r})-\frac{G^2}{4\eta}} - \frac{2q_i}{\epsilon_0}\left(\frac{\eta}{\pi}\right)^{\frac{1}{2}} + \frac{1}{4\pi\epsilon_0}\sum_j \frac{q_j}{r_j} erfc(\eta^{\frac{1}{2}} r_j) \quad (2.21)$$

2.3 Short Range Interactions

While at intermediate and long ranges the interaction between charged ions may effectively be modelled as though they were static point charges this approximation breaks down with reducing interionic separation. Consequently we must define a potential ϕ_{sr} to account for the discrepancy. Within this work the potential form used is that of the Buckingham potential model [62]:

$$\phi_{sr}(r_{ij}) = A_{ij}\exp\left(\frac{-r_{ij}}{\rho_{ij}}\right) - \frac{C_{6,ij}}{r_{ij}^6} \quad (2.22)$$

where A_{ij}, ρ_{ij} and C_{ij} are adjustable parameters. The Buckingham potential model has previously proved successful in modelling a variety of oxide ceramics [67–70] This short-range interaction is composed of both attractive and repulsive components. As the components of $\phi_{sr}(r)$ fall off rapidly with increasing r, a cut-off of 20Å can safely be imposed beyond which the short range forces are neglected.

The first, repulsive term in equation 2.22 accounts for interactions between ions close enough that their electron clouds overlap. This overlap causes two effects. Increased electron-electron repulsion decreases the energy density between the ions. The shielding of the two ion's nuclei from one another is reduced, increasing the potential energy of the system. The second effect is a result of Pauli's exclusion principle, which requires that the electrons must occupy orthogonal orbitals. Within the linear combination of atomic orbitals approximation excess energy density within the overlap must occupy previously unoccupied orbitals, if the ions are considered closed shell then these unoccupied orbitals are at higher energy. This increases the potential, i.e. corresponds to a repulsive force between the ions.

The second term in equation 2.22 is attractive and describes the London dispersion forces [71–74]. Even within a system whose time averaged charge distribution is zero there may exist, at any instant, a net dipole moment.

If the ion i has an instantaneous dipole moment p_i then at a distance r from the ion this moment generates a field proportional to $\frac{p_i}{r^3}$. This will in turn induce a dipole moment in all other ions, such as ion j, proportional to the field:

$$p_j \sim \frac{\alpha p_i}{r_{ij}^3} \tag{2.23}$$

where α is the polarisability of the ion and r_{ij} the separation of the two ions. Since the energy of interaction of two dipoles is proportional to the product of the dipole moments divided by the cube of their separations the energy of interaction, E_d is of the order:

$$E_d \sim \frac{p_i p_j}{r_{ij}^3} \sim \frac{\alpha p_i^2}{r_{ij}^6} = -\frac{C_{6,ij}}{r_{ij}^6} \tag{2.24}$$

The parameter $C_{6,ij}$ may be calculated via the Slater-Kirkwood formulae [75]:

$$C_{6,ij} = \frac{3\alpha_i\alpha_j}{2[(\alpha_i/K_i)^{\frac{1}{2}} + (\alpha_j/K_j)^{\frac{1}{2}}]} \tag{2.25}$$

where α_i is the static dipole polarisability of ion i and K_i is its effective electron number, that is the number of electrons contributing to the polarisability. Values of α and K have been calculated by Grimes and Grimes and can be found in [76, 77].

2.4 Derivation of Model Parameters

The availability of good model parameters is critical to the success of any simulation. Approaches for fitting model parameters can be partitioned into two categories: empirical and non-empirical methods. Non-empirical methods (such as used in [78]) typically involve performing rigorous quantum mechanical calculations to evaluate the ion - ion interaction energy as a function of separation. Parameters can then be chosen so as reproduce this energy surface. A significant advantage of non-empirical techniques are that they may be employed even where there is a paucity of the experimental data.

Empirical methods involve adjusting the available parameters so as to best reproduce a set of the experimentally known properties of the material. These properties will generally contain the lattice parameter and ionic positions and may extend to include elastic and dielectric constants. A significant problem with fitting in this way is that the potential form is tested only at a single point, consequently once defects are introduced to the lattice the potentials may fail as they are now required to reproduce the interaction at a variety of interionic radii. This limitation may be reduced by fitting a potential to several structures, ensuring a degree of transferability. This method has been used throughout this study and is illustrated for the O^{2-}-O^{2-} case in figure 2.2

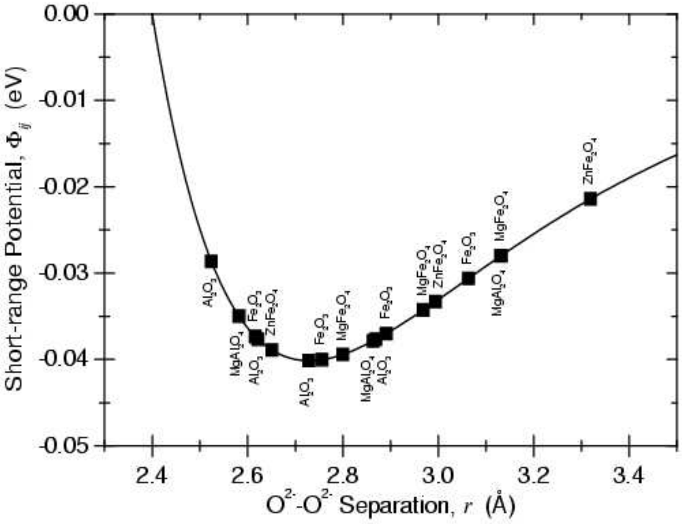

Figure 2.2: Showing the range of interionic separations in the series of compounds used in the fitting of the O^{2-}-O^{2-} potential.

2.5 Electronic Polarisability

Ionic polarisability is introduced via the shell model of Dick and Overhauser [79]. Within this scheme the rigid ion core is replaced by two species, a core and a rigid spherical shell. The shell has an effective mass of zero and charge Y|e| and is connected to the massive core of charge X|e| by a harmonic force of constant k (though in lattice statics simulations the mass plays no part). The total charge of the ion is then (X + Y)|e|.

The shell is allowed to move relative to the core, forming a dipole, thus simulating the dielectric response of the lattice. The ease by which such a dipole can be formed i.e. the polarisability of the species is controlled by a combination of the force constant k and the magnitude of the charges X and Y. An isolated species has a polarisability α_e given by,

$$\alpha_e = \frac{1}{4\pi\epsilon_0}\frac{Y^2}{k} \tag{2.26}$$

The model parameters (X, Y and k) are fitted empirically to the dielectric and elastic properties of the crystal. In particular the high frequency dielectric, ϵ_∞ is important as it arises as a consequence of the electronic polarisability alone, whereas the static dielectric constant also has a contribution from the ionic polarisation of the lattice.

The advantage of this model is that short range interactions are coupled to the shell rather than to the core, as in the rigid ion approach. This more complicated interaction introduces extra degrees of freedom into the relaxation allowing the lattice to find a more stable state. Some limitations still exist however, in particular the shell model is unable to reproduce Cauchy violation [80]. That is, calculations using this model (as well as the rigid ion model) inevitably find, for materials with the rock salt structure, that the elastic constants c_{12} and c_{44} are equal, in contrast with experimental evidence. More elaborate shell models exist and they are able to correct this, although are not used in this study. The breathing shell model which allows the shells radius to vary can reproduce cases where $c_{12} < c_{44}$ [81]. Values $c_{12} > c_{44}$ can be achieved through the use of an elliptically distorted shell [82].

2.6 Energy Minimisation

In practice, however well a potential model may be expected to reproduce the observed lattice structure, the predicted lattice must first be subjected to a process of energy minimisation. Within this scheme the system is minimised iteratively by adjusting the positions of ions within the lattice until the force upon them is reduced to zero:

$$\frac{\partial U_L}{\partial r} = 0 \tag{2.27}$$

where U_L is the lattice energy. The energy of the lattice consisting of N ions with new coordinates, r', can be defined in terms of its original set, r by:

$$U(r') = U(r) + \underline{g}^T.\underline{\delta} + \frac{1}{2}\underline{\delta}^T.\mathbf{W}.\underline{\delta} \tag{2.28}$$

where $\underline{\delta}$ is a generalised strain vector with 3N orthogonal displacement components, $\underline{\delta r}$ and 6 bulk strain components $\underline{\delta \epsilon}$. The vector $\underline{g}$ corresponds to the first derivatives of the lattice energy with respect to the ion displacements and strain components:

$$\underline{g} = \left(\frac{\partial U_L}{\partial \underline{r}}, \frac{\partial U_L}{\partial \underline{\epsilon}}\right) \tag{2.29}$$

The matrix $\mathbf{W}$ contains the corresponding second derivatives,

$$\mathbf{W} = \begin{pmatrix} \frac{\partial^2 U_L}{\partial r \partial r} & \frac{\partial^2 U_L}{\partial r \partial \epsilon} \\ \frac{\partial^2 U_L}{\partial \epsilon \partial r} & \frac{\partial^2 U_L}{\partial \epsilon \partial \epsilon} \end{pmatrix} \tag{2.30}$$

The new coordinates r' are related to the original via:

$$r' = \Delta\epsilon(r + \delta r) \tag{2.31}$$

where $\Delta\epsilon$ is the Voight matrix representation of the vector $\underline{\delta\epsilon}$:

$$\Delta\epsilon = \begin{pmatrix} \delta\epsilon_1 & \frac{1}{2}\delta\epsilon_6 & \frac{1}{2}\delta\epsilon_5 \\ \frac{1}{2}\delta\epsilon_6 & \delta\epsilon_2 & \frac{1}{2}\delta\epsilon_4 \\ \frac{1}{2}\delta\epsilon_5 & \frac{1}{2}\delta\epsilon_4 & \delta\epsilon_3 \end{pmatrix} \tag{2.32}$$

The system is relaxed iteratively, allowing the ion coordinates to adjust in such a way as to reduce the lattice strain. Energy can be minimised by varying only the internal cell ion coordinates, such that volume is constrained or the lattice can be relaxed under constant pressure by allowing the cell vectors to vary.

Minimising the lattice energy at constant volume requires that equation 2.28 be differentiated with respect to r. Now, when the lattice energy is minimised:

$$\begin{aligned}
\frac{\partial U_L}{\partial r'} &= 0 && (2.33)\\
&= \frac{\partial U_L}{\partial r'} + \frac{\partial g^T}{\partial r} && (2.34)\\
&= g + \mathbf{W}\delta && (2.35)
\end{aligned}$$

so the optimum ion displacements to give minimum energy are:

$$\delta = -\mathbf{W}^{-1} g \tag{2.36}$$

To perform a constant pressure calculation the bulk strain must also be minimised by relaxing the cell vectors. The bulk strain are defined such that they transform each vector r in the lattice to r' so that:

$$r' = (\mathbf{I} + \epsilon).r \tag{2.37}$$

where $\mathbf{I}$ is the identity matrix and ϵ the symmetric strain tensor defined by equation 2.32. New positions and lattice vectors can then be calculated by combining equation 2.37 with 2.31.

In practice the minimisation requires a number of iterations and typically continues until the change in the total energy of the system is smaller that some predetermined value. Several minimisation procedures can be applied, the Newton-Raphson method [83] being one of these.

2.7 The Defective Lattice

The inclusion of a defect into the minimised perfect lattice will cause a perturbation to the surrounding ions, hence the lattice must undergo further relaxation and obtain a new minimum energy configuration. Although the potential field associated with the defect will fall away slowly as $\frac{1}{r}$ (assuming that it is charged) it is found that the relaxation of ions close to the defect acts to shield the outer ions (polarization). The amount of relaxation therefore falls away rapidly with increasing separation from the defect. This allows the lattice to be described by the multi-region approach of Mott and Littleton [84].

Here, the lattice is partitioned into concentric spherical regions with the defect, or cluster of defects placed at the centre. The defect is surrounded by a sphere of ions, region I, initially arranged as in the perfect lattice geometry. These ions are treated explicitly, that is their adjusted lattice positions are shifted subject to the inter-atomic potentials defined in section 2.1. Beyond region I lies region II, as lattice relaxation is considerably smaller further from the defect a more approximate method can be used here. Within the Mott-Littleton approach region II is further split into regions IIa and IIb. The ions of region IIa are displaced according to forces described by the Mott-Littleton approximation but interactions with ions in region I are calculated by explicit summation. The positions of ions within region IIb, which extends to infinity, are not altered in response to the defect, the energy of interaction with the defect is determined directly from the Mott-Littleton approximation.

Within the Mott-Littleton approximation the crystal polarization P is described in terms of the distance r at which a defect of charge q is situated:

$$P = \frac{q}{4\pi r}\left(1 - \frac{1}{\epsilon}\right) \tag{2.38}$$

where $\epsilon = \epsilon_s \epsilon_0$. Individual displacements and their contribution to the total energy can then be summed by applying this equation over the whole lattice (or, rather that part of it treated within the Mott-Littleton approximation).

The total energy introduced to the lattice through the incorporation of a defect, E_d, is then written as a sum of contributions from the different regions such that:

$$E_d = E_1(r) + E_2(r, \zeta) + E_3(\zeta) \tag{2.39}$$

where E_1 is the energy of region I, due to ions displaced to positions r within this region. E_2 is the energy of interaction between region I and region II and is a function of both r and ζ the vector of coordinate displacements within region II. E_3 is the energy of region II.

E_3 cannot be calculated exactly because it is the sum of an infinite number of displacements. Relaxation in this region is limited so the harmonic approximation is acceptable and so E_3 is a quadratic function of the displacements,

$$E_3(\zeta) = \frac{1}{2}\zeta . \mathbf{A} . \zeta \tag{2.40}$$

where $\mathbf{A}$ is the force constant matrix. Substituting this 2.38 and differentiating with respect to the displacements in region II, ζ gives

$$\left.\frac{\partial E_2(r, \zeta)}{\partial \zeta}\right|_{\zeta=\zeta'} = -\mathbf{A}.\zeta' \tag{2.41}$$

where ζ' are the displacements in region II at equilibrium. Equation 2.39 can now be written as

$$E_d = E_1(r) + E_2(r, \zeta) + \frac{\zeta}{2}. \left.\frac{\partial E_2(r, \zeta)}{\partial \zeta}\right|_{\zeta=\zeta'} \tag{2.42}$$

Thus the dependence on E_3 is removed.

2.8 Thermal Equilibrium Monte Carlo Simulation

The Mott-Littleton method detailed in section 2.7 allows us to evaluate the structure and energy of point defects and small clusters. At equilibrium, however, the defects in a material cannot simply be assumed to exist in their lowest energy state. Instead the system will occupy the state μ with a probability [86]

$$p_\mu = \frac{1}{Z} e^{\frac{-E_\mu}{kT}} \tag{2.43}$$

where E_μ is the energy of state μ, T is the system temperature and k is Boltzmann's constant. The quantity $(kT)^{-1}$ is often denoted by the symbol β and this convention shall be used from now on. p_μ is known as the Boltzmann probability and the distribution formed by all the p_μ's is called the Boltzmann distribution. The normalisation constant Z is called the partition function and is defined by:

$$Z = \sum_\mu e^{-\beta E_\mu} \tag{2.44}$$

where the sum is over all possible states of the system. The following will be a general discussion of the key points behind Monte Carlo simulation, a much more complete analysis can be found in many books including [85].

A system, in state μ will transform into a new state ν within a time dt with probability $R(\mu \to \nu)dt$. The quantity $R(\mu \to \nu)$ it the transition rate for the transition from state μ to state ν.

We can define a set of weights, $w_\mu(t)$ to represent the probability that the system will be in the state μ at a given time. The evolution of $w_\mu(t)$ is then described by a master equation:

$$\frac{dw_\mu}{dt} = \sum_\nu [w_\nu R(\nu \to \mu) - w_\mu R(\mu \to \nu)] \tag{2.45}$$

i.e., evolution is the difference between the rate at which the system is undergoing transition into state μ and the rate at which it is transitioning out of state μ.

These weights are important as they allow us to relate the transition states to some macroscopic quantity Q. The expectation value of Q, written $\langle Q \rangle$, is given by:

$$\langle Q \rangle = \sum_\mu Q_\mu w_\mu(t) \tag{2.46}$$

where Q_μ is the value of Q in state μ. $\langle Q \rangle$ can be thought of as a time average of Q, i.e. that it is to be expected, that if values of Q over a period of time are recorded and averaged, they would after sufficient time reproduce the real mean value of Q. In reality this is problematic as it may not be known whether the system has had long enough to pass through a representative series of states.

If after some time, t, the system is in a state such that the two terms on the right hand side of equation 2.45 are equal, then $\frac{dw_\mu}{dt}$ will vanish and the weights become constant, i.e. the system will have reached equilibrium. The first order nature of equation 2.45 (as well as the fact the values of $w_\mu(t)$ are between 0 and 1) ensure that given enough time the system will eventually reach equilibrium. This allows us to redefine p_μ:

$$p_\mu = \lim_{0\to\infty} w_\mu(t) \tag{2.47}$$

and so for a system at equilibrium the expectation value of a quantity Q is:

$$\langle Q\rangle = \sum_\mu Q_\mu p_\mu = \frac{\sum_\mu Q_\mu e^{-\beta E_\mu}}{\sum_\mu e^{-\beta E_\mu}} \tag{2.48}$$

Unfortunately calculating the value of $\langle Q\rangle$ is only possible for small systems. Monte Carlo techniques involve the random selection of some subset of states $M = \{\mu_1, \mu_2 ... \mu_M\}$, from the probability distribution p_μ. The best estimate of Q which can be achieved by considering only this subset of states is given by:

$$Q_M = \frac{\sum_{i=1}^{M} Q_{\mu_i} p_{\mu_i}^{-1} e^{-\beta E_{\mu_i}}}{\sum_{j=1}^{M} p_{\mu_j}^{-1} e^{-\beta E_{\mu_j}}} \tag{2.49}$$

Q_M is called the estimator of Q, it provides a more accurate estimate as the number of sampled states increases and in the limit of $M \to \infty$ it recovers exactly $\langle Q\rangle$. The problem which now arises is how to choose M so as to yield an accurate estimate of $\langle Q\rangle$. It we knew which states contributed significantly to the sums in equation 2.48 then we could pick M from only those states, ignoring all the others. In this way we could gain an accurate measure of $\langle Q\rangle$ with a minimum number of terms in the summation. The selection of an appropriate set of states in this way is called importance sampling.

The simplest way of selecting states would be to choose each state with equal probability, however such a choice may well be a very poor one because clearly the system will occupy lower energy states with greater frequency. In fact we know that states occur with a Boltzmann probability as defined in equation

2.43 so if states, M, are selected such that the probability of it being selected is $p_\mu = \frac{1}{Z}e^{-\beta E_\mu}$ then a representative sample will be obtained (provided that the sample is sufficiently large). Substituting 2.43 into 2.49 reduces the estimator to the form:

$$Q_M = \frac{1}{M}\sum_{i=1}^{M} Q_{\mu_i} \tag{2.50}$$

The selection of states according to their Boltzmann probabilities is difficult and typically relies on what is known as a Markov process. A Markov process is a system through which a new state ν is generated from its predecessor μ. The probability that such an evolution occurs is defined by $P(\mu \to \nu)$, this is the transition probability. These transition probabilities must be time independent and be a function only of the states μ and ν. Additionally the sum of all possible transitions from the state μ must be unity (though $P(\mu \to \mu)$ need not be zero) i.e.

$$\sum_{\nu} P(\mu \to \nu) = 1 \tag{2.51}$$

If this process is used repeatedly the chosen are a Markov chain of states. In order to ensure that this chain will eventually contain states with appropriate probabilities two further conditions must be applied, these are ergodicity and detailed balance.

The condition of ergodicity requires that every state must be attainable from the initial state if the chain is continued for a sufficiently long period. This means that, though in practice it may be convenient to set some transition probabilities to zero in order to simplify a calculation, we must ensure that there is at least one non-zero transition path available.

The condition of detailed balance ensures that it is the Boltzmann probability distribution which is generated after the system has come to equilibrium. A defining condition is that, at equilibrium, the rate at which the system enters the state μ must be equal to the rate at which it leaves this state (this is analogous to equation 2.45 when $\frac{dw_\mu}{dt} = 0$):

$$\sum_\nu p_\mu P(\mu \to \nu) = \sum_\mu p_\nu P(\nu \to \mu) \tag{2.52}$$

substituting 2.51 this becomes:

$$p_\mu = \sum_\mu p_\nu P(\nu \to \mu) \tag{2.53}$$

This is sufficient to ensure that p_μ is the equilibrium probability distribution, this is not however, sufficient to ensure that the probability distribution will tend to p_μ from any initial state. This is because it does not prevent the formation of limit cycles. These occur when the probability distribution becomes trapped in a dynamic equilibrium such that it rotates cyclically through a series of different values. This problem is eliminated by further requiring:

$$p_\mu P(\mu \to \nu) = p_\nu P(\nu \to \mu) \tag{2.54}$$

Thus we have arrived at the condition of detailed balance. Rearranging gives:

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{p_\nu}{p_\mu} = e^{-\beta(E_\nu - E_\mu)} \tag{2.55}$$

All that remains is to define the algorithm controlling the evolution of the Markov process, however it is useful to first introduce the concept of acceptance ratios. If, in equation 2.54 we set $\mu = \nu$ then it is clear that the condition is fulfilled regardless of the value taken by $P(\mu \rightarrow \mu)$. This gives some freedom in the definition of the probabilities $P(\mu \rightarrow \nu)$ as any adjustment to the value of $P(\mu \rightarrow \nu)$ can be compensated for by an equal and opposite adjustment in $P(\mu \rightarrow \mu)$ ensuring that the sum rule is fulfilled. This can be exploited by breaking the transition probability into two parts:

$$P(\mu \rightarrow \nu) = g(\mu \rightarrow \nu)A(\mu \rightarrow \nu) \tag{2.56}$$

$g(\mu \rightarrow \nu)$ is the selection probability, it is the probability that while the system is in state μ the Markov process algorithm will generate the state ν. $A(\mu \rightarrow \nu)$ is the acceptance ratio, this states that given that the state ν has been generated from initial state μ the transition should only be accepted some fraction of the time. If we write:

$$\frac{P(\mu \rightarrow \nu)}{P(\nu \rightarrow \mu)} = \frac{g(\mu \rightarrow \nu)A(\mu \rightarrow \nu)}{g(\nu \rightarrow \mu)A(\mu \rightarrow \nu)} \tag{2.57}$$

Then $\frac{A(\mu\rightarrow\nu)}{A(\nu\rightarrow\mu)}$ lies anywhere between zero and infinity, giving the freedom to choose any value for the selection probabilities $g(\mu \rightarrow \nu)$ and $g(\nu \rightarrow \mu)$. It is clear that although $\frac{A(\mu\rightarrow\nu)}{A(\nu\rightarrow\mu)}$ may be any real positive number it is beneficial for the individual acceptance ratios to be as close as possible to 1 as any rejected swaps equate to wasted computations. The best possible choice is for the larger of the two ratios to equal 1 and the other to be adjusted in accordance with equation 2.54.

An algorithm is now required that will generate new states ν from initial state μ with probability $g(\mu \rightarrow \nu)$, and, that will accept this change with a probability $A(\mu \rightarrow \nu)$. The most widely used example is the Metropolis algorithm [87]. An important facet of this algorithm is that only transitions which cause the state ν to vary from state μ by the smallest possible amount are permitted. Later in this thesis Monte Carlo analysis will be applied to the problem of cation inversion in $MgAl_2O_4$ spinel, in this case only transitions involving the exchange of a single pair of cations are allowed. In the Metropolis algorithm the $g(\mu \to \nu)$ are chosen to be equal for all allowed transitions and to be zero for all others. If there are N allowed transitions then:

$$g(\mu \to \nu) = \frac{1}{N} \tag{2.58}$$

The condition of detailed balance now reads:

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{A(\mu \to \nu)}{A(\mu \to \nu)} = e^{-\beta(E_\nu - E_\mu)} \tag{2.59}$$

As mentioned above the most efficient algorithm is one in which the larger acceptance ratio takes the highest possible value, namely 1. Of two states μ and ν one, for example ν will have higher energy, i.e. $E_\nu > E_\mu$. In this case the larger of the acceptance ratios is $A(\nu \to \mu) = 1$ thus $A(\mu \to \nu) = e^{-\beta(E_\nu - E_\mu)}$. We can now write the acceptance ratio as:

$$A(\mu \to \nu) = \begin{cases} e^{-\beta(E_\nu - E_\mu)} & \text{if} E_\nu > E_\mu \\ 1 & \text{otherwise} \end{cases} \tag{2.60}$$

This is the Metropolis algorithm.

2.9 Density Functional Theory

Density Functional Theory, (DFT), is a quantum mechanical technique originating from work by Hohenberg, Kohn and Sham [88,89]. They showed that the ground state energy of an interacting inhomogeneous electron gas in a static potential $v_{ext}(\mathbf{r})$ can be written in the form:

$$E = \int v_{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + G[\rho(\mathbf{r})] \tag{2.61}$$

where E is the total ground state energy of the electron gas, $\rho(\mathbf{r})$ is the charge density and $G[\rho(\mathbf{r})]$ is a universal functional of the charge density. Equation 2.61 is at a minimum for the correct density function $\rho(\mathbf{r})$. $G[\rho(\mathbf{r})]$ can be written as:

$$G[\rho(\mathbf{r})] \equiv T_s[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] \tag{2.62}$$

where $T_s[\rho(\mathbf{r})]$ is the kinetic energy of a system of noninteracting electrons with charge density $\rho(\mathbf{r})$ and $E_{xc}[n]$ is the exchange-correlation term. $T_s[\rho(\mathbf{r})]$ can be computed from,

$$T_s[n] \equiv \langle \Psi_n^0 | \widehat{T} | \Psi_n^0 \rangle \tag{2.63}$$

where Ψ_n^0 is the ground state wavefunction of the system and the kinetic energy operator $\widehat{T} = \frac{\nabla^2}{2}$. Up to this point the methodology is exact, however an analytical form for $E_{xc}[\rho(\mathbf{r})]$ is not known. For an electron gas, a system of interacting particles, the effects of exchange and correlation are important for an accurate description of the behaviour. In a non-interacting system, the antisymmetry of the wave-function requires that particles with the same spin occupy distinct orthogonal orbitals, thus the particles become spatially separated. In an electron gas, in which all the atoms repel one another, exchange will therefore lead to a lowering of the energy. Interactions also result in the motion of the particles becoming correlated to further reduce the interaction energy. Kohn and Sham showed that the local density approximation,

$$E_{xc}^{LD} = \int d\mathbf{r}\rho(\mathbf{r})\epsilon_{xc}[\rho(\mathbf{r})] \tag{2.64}$$

could be applied to the case of an inhomogeneous electron gas in the limit of a slowly varying density. In this ϵ_{xc} is the known exchange-correlation energy of the uniform electron gas [90]. Although this limit is not approached in the case of atoms, molecules or solids this approximation has been seen to perform well, reproducing accurately qualities such as ground state geometry and energetics, vibration and phonon frequencies [91]. This model can be further improved to take into account the inhomogeneity of the electron gas via a density gradient expansion [92], i.e. to treat E_{xc} as $\text{E}_{xc}[\rho(\mathbf{r}), \frac{d(\rho)}{d\mathbf{r}}]$. In addition to calculations utilising the local density approximation this thesis contains calculations in which E_{xc} has been approximated by the gradient corrected functional of Perdew and Burke and Ernzerhof (PBE) [93].

The charge density is constructed from a wavefunction, Ψ. The wavefunction is taken to be a Slater determinant of one-particle wavefunctions

$$\Psi = A(n)|\phi_1\phi_2...\phi_n|. \tag{2.65}$$

When the molecular orbitals, ϕ_i are orthonormal (i.e. $\langle\phi_i|\phi_j\rangle = \delta_{ij}$) the charge density is given by the sum

$$\rho(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2 \tag{2.66}$$

The energy is determined by optimizing variations in the total energy, defined in equation 2.61, with respect to variations in the charge density, subject to the orthogonality conditions stated above.

$$\frac{\delta E}{\delta \rho} - \sum_i \sum_j \epsilon_{ij} \langle \phi_i | \phi_j \rangle = 0 \tag{2.67}$$

where ϵ_{ij} are Lagrangian multipliers.

This process leads to the Kohn-Sham equations [89]:

$$\left[-\frac{1}{2} \nabla^2 + V_{KS}(\mathbf{r}) \right] \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \tag{2.68}$$

where the Kohn-Sham potential is

$$V_{KS}(\mathbf{r}) = \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + v_{xc}(\mathbf{r}) + v_{ext}(\mathbf{r}) \tag{2.69}$$

$v_{ext}(\mathbf{r})$ is a static potential, external to the electon gas and $v_{xc}(\mathbf{r})$, the exchange-correlation potential is

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} \tag{2.70}$$

Since $V_{KS}(\mathbf{r})$ depends upon the density $\rho(\mathbf{r})$ these equations must be solved self consistently; starting with an assumed $\rho(\mathbf{r})$, solve equation 2.68 to obtain a set of molecular orbitals $\{\phi_i(\mathbf{r})\}$ from which a new density is constructed.

This process is repeated until the input and output densities are the same, to within some chosen criteria.

The periodic nature of crystals means that periodic boundary conditions are appropriate. Bloch's theorem [94, 95] shows that within a periodic system each wavefunction may take the form

$$\Psi(\mathbf{r}) = u_k(\mathbf{r})e^{i\mathbf{k}.\mathbf{r}} \tag{2.71}$$

these Bloch Functions can be thought of as plane waves ($e^{i\mathbf{k}.\mathbf{r}}$) modulated by a periodic function, $u_k(\mathbf{r})$, provided that $u_k(\mathbf{r})$ satisfy the relation

$$u_k(\mathbf{r} + \mathbf{T}) = u_k(\mathbf{r}) \tag{2.72}$$

where the vector $\mathbf{T}$ represents a translation to an identical lattice point. The $u_k(\mathbf{r})$ can be used to construct a solution by summing the contributions from the terms with wave vectors $\mathbf{k} + \mathbf{G}$ where $\mathbf{G}$ is any reciprocal lattice vector. In principle there are an infinite number of such vectors, however, those which have small kinetic energies $|\mathbf{k} + \mathbf{G}|^2$ are more important than those with large kinetic energies, the plane-wave basis set is thus truncated to include only those waves with kinetic energies less than some cut-off energy. This cut-off energy must be high enough to allow sufficient plane-waves to accurately describe the reciprocal lattice, hence it is necessary to ensure that the calculations are converged with respect to changes in this value. In order to make computation possible the wave function is only calculated at a subset of these k-points. The process of reducing the k-point set is called k-point sampling. The scheme used for the quantum mechanical calculations in this study is due to Monkhorst and Pack [96] in which the chosen points form a uniform grid in k space.