Computational studies of point defects and dislocations in forsterite ($\text{Mg}_2\text{SiO}_4$) and some implications for the rheology of mantle olivine

Andrew Mark Walker

Davy-Faraday Research Laboratory, The Royal Institution of Great Britain

Department of Earth Sciences, University College London

Thesis submitted for the degree of Doctor of Philosophy

to the

University of London

2004
Abstract

Convection in the Earth’s solid mantle, accompanied by long time-scale viscous creep, is the mechanism operating on the global scale that allows internally generated heat to escape from the Earth’s deep interior. This process drives plate tectonics and, ultimately, almost every other geological process occurring at the Earth’s surface. On a small scale, the mantle deforms by plastic deformation of its constituent crystals – a process that involves the movement of a range of crystal defects (point defects, dislocations and grain boundaries). An understanding of the defect properties of mantle phases is thus essential to any unified, multi-scale view of the behaviour of the Earth. This thesis describes research on a range of defect phenomena in olivine ([Mg,Fe]$_2$SiO$_4$) undertaken using atomic scale computational modelling techniques.

Following an outline of the theoretical background, the structure, energetics and properties of a range of point defects are studied using an inter-atomic parameterised potential model and the Mott-Littleton method. A combined electronic structure – parameterised potential embedded cluster approach has also been employed for selected defects. Bulk diffusion occurring by point defect motion is then considered using static lattice techniques with the parameterised model. The diffusion of oxygen is treated in some detail and an explanation of a range of experimental observations is proposed by considering the formation and motion of defects on the oxygen sub-lattice coupled to the redox behaviour of iron.

The second part of the thesis concentrates on modelling line defects. First a methodology for studying screw and edge dislocations in structurally complex ionic materials is developed from an earlier approach limited to studies of ionic materials with the rock salt structure. This methodology is then applied to MgO, zeolite A (a technologically significant material) and finally forsterite, the pure magnesium end member of the olivine solid solution.
Acknowledgments

I am deeply indebted to my supervisors, Drs Kate Wright and Ben Slater, for their help, support and encouragement over the last three years. I am also extremely grateful that Kate and Ben gave me considerable freedom to dictate the direction of the research project leading to the production of this thesis, making the research so much more fulfilling.

In addition, I acknowledge useful discussions with a many people – both within the DFRL and beyond. Any list will inevitably be incomplete (apologies in advance all the people left out), but I particularly want to thank Drs Béjina, Cordier and Dr Sushko and Profs. Gale, Harding, Jaoul and Price. I also thank all in the DFRL, in particular Prof. Catlow and Drs Sokol and Woodley. Last, and not least, I thank all the members of the “Hydrospec” consortium, an EU funded training network which investigated hydrogen speciation in the mantle. Discussions with all of these people have enriched my experience of producing this thesis and I particularly enjoyed visits to the research groups of Dr Cordier in Lille, Prof. Jaoul in Toulouse and Prof. Gale in Perth and Dr Demouchey’s visit to the DFRL.

The research would not have been possible without finding from a range of sources, principally the Engendering and Physical Sciences Research Council who provided financial support via a studentship and access to various computer resources including local facilities and access to the UK Capability Computing resource (HPCx). Further computing resources were provided by NERC through their e-Science program. Additional travel funds were provided by the UCL Graduate School and Mineralogical Society of Great Britain.

Finally, I wish to thank my all my friends in London and my officemates, Dan, Dave and Kat for making the last the few years so much fun. I also thank my family for all their support – this is dedicated to them.
### 3.2 Parameterised potential models

- **3.2.1 Coulomb summation**
- **3.2.2 Pair-potentials**
- **3.2.3 Three and four body interactions**
- **3.2.4 Polarisation**

### 3.3 Energy minimisation and parameter fitting

- **3.3.1 Local optimisation algorithms**
- **3.3.2 Derivation of parameters**

### 3.4 Calculation of properties

- **3.4.1 Elastic properties**
- **3.4.2 Dielectric properties**
- **3.4.3 Lattice vibrations**

### 3.5 Accounting for the effect of pressure and temperature

### 3.6 Modelling point defects

- **3.6.1 The super cell approach**
- **3.6.2 Cluster based methods**

### 3.7 Extended defects

### Chapter 4: Models of point defects in olivine

- **4.1 Review of previous work**
- **4.1.1 Bulk properties**
- **4.1.2 Defect properties**

### 4.2 Calculation of bulk properties

### 4.3 Modelling point defects - methodology

- **4.3.1 Mott-Littleton calculations**
- **4.3.3 Embedded cluster calculations**

### 4.4 Modelling point defects - Results
List of figures

Figure 1.1: Divisions of the Earth’s interior .............................................................. 16
Figure 1.2: Mineralogy of pyrolite mantle as a function of depth ......................... 18
Figure 1.3: The olivine unit cell ................................................................................. 19
Figure 1.4: The structure of wadsleyite ................................................................. 20
Figure 1.5: (Mg,Fe)\textsubscript{2}SiO\textsubscript{4} phase diagram ........................................... 21
Figure 2.1: Types of point defects ........................................................................... 24
Figure 2.2: Example of solid-state diffusion ............................................................ 28
Figure 2.3: Arrhenius diagram for diffusion ............................................................ 31
Figure 2.4: Activation energy associated with diffusion .......................................... 31
Figure 2.5: Atomic scale mechanism for diffusion in crystalline solids ................. 33
Figure 2.6: Plastic deformation of perfect and dislocated crystal ......................... 36
Figure 2.7: The extra half plane ................................................................................. 37
Figure 2.8: Definition of the Burgers vector ........................................................... 37
Figure 2.9: Glide of edge dislocation by the propagation of kinks ......................... 38
Figure 2.10: Movement of the dislocation line out of the slip plane ....................... 39
Figure 3.1: Schematic potential energy between two atoms as a function of separation ..................................................................................................................... 58
Figure 3.2: Charge summations ................................................................................. 60
Figure 3.3: Graphical representation of the shell model and breathing shell model ........................................................................................................................................ 65
Figure 3.4: Schematic representation of the Mott-Littleton and super cell approaches to the modelling of defects .................................................................................. 77
Figure 3.5: Embedded cluster model ........................................................................ 85
Figure 3.6: Charge distribution at edge of cluster .................................................... 86
Figure 4.1: DFT calculations of equation of state of forsterite ............................. 100
Figure 4.2: Convergence in defect energy of a Mg\textsubscript{1} vacancy in forsterite with basis set size ..................................................................................................................... 107
Figure 4.3: Structure of silicon vacancy ................................................................... 109
Figure 4.4: Structure of O\textsubscript{3} vacancy ................................................................ 111
Figure 4.5: Structure of magnesium split interstitial defect ..................................... 113
Figure 4.6: Convergence in defect energy of the magnesium split interstitial in forsterite with basis set size ........................................................................................................ 114
Figure 4.7: Structure of $\{V_{Mg}OH_{O_2}\}^*$ and $\{V_{Mg}^2OH_{O_2}\}^*$ defects ........................................ 119

Figure 4.8: Structure of the $OH_{O_2}^*$ and $\{V_{Si}OH_{O_2}OH_{O_2}OH_{O_3}OH_{O_3}\}^*$ defects ...... 120

Figure 5.1: Idealised cartoon of the single atom RMS displacement ..................... 134
Figure 5.2: 2-dimensional contour representation of a potential energy surface, with surface overlay ................................................................. 135
Figure 5.3: Variation of oxygen diffusivity in olivine (m/s, y-axis) with temperature (°C, x-axis), direction and oxygen fugacity................................. 139
Figure 5.4: Possible routes for oxygen diffusion in forsterite assuming a vacancy mechanism where only O3 vacancies are involved.............................. 144
Figure 5.5: Possible routes for oxygen diffusion in forsterite assuming an interstitial mechanism where only I2 sites are preferred over I1 sites............. 147
Figure 5.6: Graphical representation of the 2000 data points used to estimate the location of the transition states for magnesium interstitial diffusion........ 156
Figure 5.7: Magnesium vacancy diffusion mechanisms ......................................... 157
Figure 6.1: Cartoon showing the process used to introduce a screw dislocation into a perfect crystal by Voltar's procedure ....................................... 169
Figure 6.2: Cell setup for the evaluation of $E(r)$ for a dislocation...................... 179
Figure 6.3: Three locations used to centre the 1D simulation cell and the elastic displacement field for models of the $<100>$ screw dislocation in MgO......... 183
Figure 6.4: Convergence of dislocation energy with thickness of the fixed portion of the simulation cell................................................................. 183
Figure 6.5: Simulation cell containing $<100>$ screw dislocation in MgO.......... 185
Figure 6.6: Energy of MgO screw dislocation .................................................. 186
Figure 6.7: Relaxation from elastic to atomistic solution for the $<100>$ screw dislocation in MgO................................................................. 187-8
Figure 6.8: AFM image of growth spirals on the $\{100\}$ surface of synthetic zeolite A ................................................................................................. 190
Figure 6.9: Topology of zeolite A based on the silica framework....................... 191
Figure 6.10: Two possible screw dislocations in the zeolite A ......................... 195
Figure 6.11: Views of three 8-ring structures in zeolite A................................. 196-7
Figure 6.12: Displacements perpendicular to and along the dislocation line during atomic relaxation................................................................. 198
Figure 6.13: Channel systems in perfect and dislocated zeolite A.................... 200
Figure 7.1: View of core of a[100] screw dislocation in forsterite..................... 208
Figure 7.2: Simple model of core of the a[100] screw dislocation in forsterite showing central octahedra................................................................. 209
Figure 7.3: Relaxation of the atomistic model after introduction of the a[100] screw dislocation................................................................. 209-10
Figure 7.4: View of core of c[001] screw dislocation in forsterite.................211

Figure 7.5: Core displacement field associated with the c[001] screw dislocation in forsterite..........................................................212

Figure 7.6: Model for the structure of the core of a c[001] screw dislocation in forsterite.................................................................213

Figure 7.7: Simulation cell used to model the a[100](010) edge dislocation in forsterite.................................................................215

Figure 7.8: View of core region of [100] edge dislocation .........................217

Figure 7.9: Difference between calculated linear elastic displacement field and total displacement field determined by experiment and the modelling ........218

Figure A1.1: General approach to generating simulation cells for dislocation simulations.........................................................224
List of tables

Table 1.1: Composition of pyrolite ................................................................. 17
Table 3.1: Major differences between atom-centred and plane wave basis functions ......................................................................................... 46
Table 4.1: Parameters of the potential model used in this study ................... 94
Table 4.2: Comparison of properties of forsterite, wadsleyite and ringwoodite calculated using the potential model with experimental data .................. 95
Table 4.3: Electronic structure studies of forsterite, including available computational details, cell parameters and errors .................................. 98
Table 4.4: Parameters of fitted Mumaghan athermal equation of state .......... 99
Table 4.5: Estimated transition pressures for the two phase transitions in the Mg$_2$SiO$_4$ system ........................................................................ 101
Table 4.6: Parameters of the potential model used for the molecular mechanics part of the embedded cluster study ........................................... 104
Table 4.7: Defect energies for a range of possible intrinsic defects in forsterite calculated via the Mott-Littleton and embedded cluster methods .......... 106
Table 4.8: Defect energies for the Mg$^+$ vacancy ........................................... 108
Table 4.9: Defect energies for the Si vacancy .................................................. 109
Table 4.10: Defect energies for O vacancies .................................................... 111
Table 4.11: Defect energies for Mg split interstitial ........................................ 115
Table 4.12: Defect energies of OH defects ..................................................... 118
Table 4.13: Calculation of OH frequencies ..................................................... 121
Table 4.14: OH - cation vacancy binding energies ........................................ 123
Table 4.15: Lattice energies ............................................................................ 126
Table 5.1: Migration energies for vacancy and interstitial diffusion hops ...... 145
Table 5.2: Energies of various hydroxyl defects and hydroxyl - oxygen defect clusters in forsterite ................................................................. 148
Table 5.3: Migration energies for magnesium vacancy and interstitial diffusion in forsterite ................................................................. 158
Table 6.1: Potential parameters used for modelling MgO .............................. 180
Table 6.2: Calculated and experimental properties of MgO .......................... 181
Table 6.3: Potential parameters used for the atomistic model of zeolite A ...... 192
Table 6.4: Calculated structure and properties for $\alpha$-quartz ....................... 193
Table 6.5: Pore dimensions, bond distance and geometry of the screw dislocation in zeolite A ................................................................. 199
Table 7.1: Parameters of the modified potential model used to study some dislocations ................................................................................................................206

Table 7.2: Comparison of the performance of two atomistic models used to model dislocations in forsterite ................................................................................206

Table A2.1: Total energies of embedded clusters using the smaller cell .......... 227
Table A2.2: Total energies of embedded clusters using the larger cell .......... 227
Table A2.3: Total energies of isolated ions............................................................ 228
Chapter 1: Crystal defects and the mantle

The simple conceptual view of a crystal is an infinite periodic array of points making up a Bravis lattice populated with motifs of atoms called the basis. This picture is of great utility for understanding and calculating a wide range of crystal properties. For example, it describes diffraction of electrons, x-rays and neutrons, it is sufficient for calculating the elastic and dielectric constants of a crystal and it allows static compression, thermal conduction and thermal expansion to be rationalised. But the picture is not complete, clearly the crystal cannot be infinitely periodic and all crystals must end somewhere, either at a surface or a grain boundary. Occasionally some atoms may be missing, displaced from their correct position or replaced by an atom of the wrong species. These are examples of two classes of crystal defects; surfaces and grain boundaries are types of planar or two-dimensional defect and vacancies, interstitials and impurities are types of point defects. There is a third class of defects, one-dimensional or line defects, which include dislocations corresponding to the termination of extra half planes of atoms. To understand many properties of crystals of geophysical significance these defects are crucial. Non-recoverable plastic deformation of crystals and polycrystalline aggregates can be controlled by the motion of point, line or planar defects; merely considering a perfect crystal results in a very large overestimate of its resistance to plastic deformation. Many other crystal properties are also controlled by defects. Examples include electrical and thermal conductivity, recoverable but non-elastic deformation (called anelasticity) responsible for the attenuation of seismic waves and, sometimes, the nucleation of new phases as the crystal passes through a solid state, or solid to liquid phase transition.

Clearly defect properties of crystals are of interest to the geophysicist, the Earth is, after all, solid from the surface to the Gutenberg discontinuity at the base of the mantle at 2889 km depth, beneath which is the outer core consisting of liquid iron and other elements (e.g. Alfé et al., 2000, 2002). Although solid, the mantle is slowly convecting, like a hot viscous fluid, because heat generated internally, largely by radiogenic processes within the mantle, causes a
temperature (and thus density and buoyancy) gradient in the mantle, which is too large to be removed by any other process (see Davies, 1999). On a smaller scale the convective flow occurs by plastic deformation of the crystals that make up the mantle, this in turn is controlled by the movement of defects. Mantle convection drives plate tectonics on the surface so, in a sense, all of the major surface characteristics of the Earth are controlled by the nature of defects in the mantle, and it should be possible to ascribe the difference in morphology between the Earth and the other terrestrial planets to differences in defects in their interiors. (It has been suggested by Regenaur-Lieb and co-workers (2001; 2003) that this difference in defect population is due to a difference in water content and this accounts for the very different styles of tectonics on the Earth, Mars and Venus.) In addition to controlling mantle flow defects in the crystals forming the mantle are important in other mantle processes. As defects control electrical conduction in insulating crystals they are important for understanding the mantle's electrical conductivity profile, which can be extracted from observations of the Earth's magnetic field. Defects are also responsible for the attenuation of seismic energy as it passes through the Earth (Tan et al., 1997, 2001) and allow mantle minerals to accommodate trace elements so defects also influence various geochemical cycles.

This thesis describes some attempts to understand a number of defect processes in the major upper mantle mineral olivine by using computational techniques. In the rest of this introductory chapter the composition, structure and dynamics of the Earth's interior is outlined for the benefit of the reader without a geological background. In chapter 2 the processes that allow plastic deformation are described in detail. Chapter 3 gives a background to the various computational methods available to model perfect crystals and point defects, chapter 4 puts these techniques to work in studies of the structure and energetics of stationary point defects in olivine. In chapter 5 computational models of oxygen and magnesium diffusion are presented, along with a review of the available methodologies for the atomic scale modelling of diffusion. Chapter 6 describes methodological advances in the simulation of dislocations in complex ionic materials and illustrates this with studies of dislocations in a number of complex systems. Chapter 7 focuses on the simulation of some
defects in olivine and chapter 8 outlines plans for future work and some general conclusions that can be drawn from the study.

1.1 The Earth's deep interior

Direct evidence regarding composition and structure of the Earth's interior comes only from geophysical observation; primarily travel time data for teleseismic waves. Interpretation of this data without additional information is extremely difficult and only yields the Earth's density profile (encapsulated in models such as PREM), but additional information allows this to be interpreted to reveal the composition of the Earth's interior. Despite the sparsity of sources of information, there is general agreement of the structure and composition (if not the detailed dynamical behaviour) of the Earth's deep interior. The inter-relationships between the various models based on geophysical and geochemical data have been explored by Poirier (2000), in the following section the consensus view of the structure of the Earth is briefly outlined.

1.1.1 Structure of the Earth's interior

Figure 1.1 shows the major divisions of the Earth's interior. The primary distinction is between the iron rich metallic core (at depths below 2889 km) and the silicate mantle (above 2889 km) capped by a thin heterogeneous crust. The crust varies in thickness between only ~7 km for the oceanic crust (occupying over two thirds of the surface) and up to ~60 km for particularly thick parts of the continental crust.

__________________________

1 Geophysical observations of the Earth's deep interior are limited to studies of the magnetic and gravitational field and the properties of seismic waves that have propagated through the Earth. These observations are supplemented by studies of the geochemistry of magmas and phenocrysts and expectations of the Earth's bulk composition derived from models of the early Solar System and from the composition of meteorites.
**Figure 1.1:** Divisions of the Earth’s interior using the nomenclature recommended in figure 5.2 of Davies (1999). Note that this figure is not to scale.

The mantle is conveniently divided into three sections defined by the presence of four discontinuities in seismic wave speeds calculated from travel time data. The Mohorovicic discontinuity (or Moho) marks the base of the crust where feldspars become unstable resulting in a major increase in density and seismic wave speed. The base of the mantle is marked by the even more dramatic Gutenberg discontinuity, beyond which S-waves are not propagated indicating the presence of a liquid outer core. In between these two discontinuities the seismic properties of the mantle can be ascribed to a constant chemical composition shown in table 1.1 and the rest of the seismic discontinuities can be
explained by solid state phase transitions which result in changes in density (and thus wave speeds). Of particular importance are the discontinuities at 410 and 660 km which are believed to be due to phase transitions associated with the most common upper mantle mineral, olivine (see figure 1.2), described in the following section. The major density changes indicated by the two discontinuities are due to density changes between an olivine-pyroxene-garnet upper mantle, a wadsleyite-ringwoodite-pyroxene-garnet transition zone and a perovskite-magnesowustite lower mantle. These assemblages are separated by the 440 km discontinuity and the 660 km discontinuity (Griffiths and Turner, 1998). The wadsleyite – ringwoodite transition is sometimes seen as a mid transition zone discontinuity at about 520 km (Kennett and van der Hilst, 1998), but this is complicated by the breakdown of pyroxenes at about the same depth (Deuss and Woodhouse, 2001). Before finishing the discussion of the structure of the Earth, two final discontinuities should be mentioned. Firstly there is a very deep discontinuity that separates the lower mantle from the D’’ layer, this has received much attention recently and may be due to a solid state post perovskite phase transition (Litaka et al., 2004; Murakami et al., 2004; Oganov and Ono, 2004). The second is a discontinuity that is sometimes observed at about 220 km and does not coincide with any known phase transition but may be due to a change in the deformation mechanism of olivine (Deuss and Woodhouse, 2002; Woodhouse and Deuss, 2003).

**Table 1.1**: Composition of pyrolite, a synthetic mixture that approximates the composition of the mantle (from Davies, 1999).

<table>
<thead>
<tr>
<th>Component</th>
<th>Weight %</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO₂</td>
<td>45.0</td>
</tr>
<tr>
<td>MgO</td>
<td>38.0</td>
</tr>
<tr>
<td>FeO</td>
<td>7.8</td>
</tr>
<tr>
<td>Fe₂O₃</td>
<td>0.3</td>
</tr>
<tr>
<td>Al₂O₃</td>
<td>4.4</td>
</tr>
<tr>
<td>CaO</td>
<td>3.5</td>
</tr>
</tbody>
</table>
1.1.2 Crystal chemistry of olivine and its polymorphs

As we have seen, the majority of the Earth’s upper mantle consists of olivine: a solid solution between forsterite (Mg$_2$SiO$_4$, Fo$_{100}$) and fayalite (Fe$_2$SiO$_4$, Fo$_0$). Structurally, olivine can be described as either an orthosilicate or a hexagonally close packed (HCP) lattice of oxygen ions with half of the octahedral sites occupied by iron or magnesium ions and one eighth of the tetrahedral sites occupied by silicon ions (Bragg and Brown, 1929). These ions distort the HCP lattice and give olivine orthorhombic symmetry (throughout this thesis the space group setting $Pbnm$\footnote{Use of this non standard space group (which sets $b > c > a$) has the advantage of aligning the $a$-axis with the pseudo-sixfold rotation axis in the HCP oxygen sub lattice.} is used – however, no symmetry is applied to constrain any of the calculations). The olivine unit cell, shown in figure 1.3, contains three symmetry inequivalent oxygen sites (O1, O2 and O3), two metal sites (M1 and M2) and one silicon site. Two unoccupied octahedral sites (I1 and I2) are related to M1 and M2. I1 and M1 are on inversion centres and I2 and M2 lie on
a mirror plane. The M sites can also host other elements, such as cobalt, nickel and manganese. At high temperature, these elements disorder over both M sites but tend to order to one or the other site on cooling. The tendency to order is hindered at low temperatures by sluggish kinetics; an effect that freezes in disorder in rapidly cooled olivine and can be used as a “geospeedometer” to determine the cooling, and thus exhumation rate (Redfern et al., 1996).

Figure 1.3: The olivine unit cell looking along [100] with oxygen in red, magnesium in green and silicon in yellow, shown bonded to the oxygen. The oxygen atoms O1 and O2, silicon and magnesium M2 lie on the mirror plane, magnesium M1 is on an inversion centre and oxygen O3 is in a general position. The two interstitial positions I1 (obscured by magnesium M1 ions in this view) and I2 are also indicated.

On increasing pressure, forsterite undergoes a pressure induced phase transition to wadsleyite (figure 1.4), a sorosilicate with a body centred cubic (BCC) oxygen sub lattice. The structure is orthorhombic (space group Imma), but there is also a monoclinic hydrous polymorph (space group I2/m) recently identified by Smyth et al. (1997). Wadsleyite is characteristic of the upper part of Earth’s transition zone (13-18 GPa, 1300-1400°C Fei et al., 1990), but, as natural wadsleyite samples from the Earth’s mantle are not accessible, studies rely on synthesis of wadsleyite using high-pressure and high-temperature apparatus, samples recovered from meteorites (Price, 1983) and computer modelling.
Figure 1.4: The structure of wadsleyite. Colours of ions are identical to those in figure 1.3

The second phase transition in the MgSiO$_4$ series is from wadsleyite to the spinel structured mineral ringwoodite, which is the major constituent of the mantle in the lower part of the transition zone (between 520 and 660 km deep). In the spinel structure the oxygen ions preserve the BCC structure of wadsleyite but the cations are rearranged to form a cubic structure with silicon forming isolated tetrahedra and magnesium in octahedral coordination. Above 660km ringwoodite decomposes into MgSiO$_3$ with the perovskite structure (distorted to an octahedral space group) and MgO.

The effect of iron on the phase diagram of olivine and its high-pressure polymorphs is shown in figure 1.5; adding iron to the structure decreases the stability field of wadsleyite until, at 20% iron, the wadsleyite phase is entirely absent. A second effect, which occurs whatever the impurity species is, is to widen the temperature – pressure range over which the phase transition takes place. This is the same effect observed when a multi-component system freezes – there is a temperature range between the liquidus and solidus – and can be used to provide an estimate for the maximum hydrogen content of the transition zone (Wood, 1995).
1.1.3 Mantle dynamics

While there is widespread agreement regarding the structure and composition of the Earth's interior, there is far less agreement regarding the dynamical behaviour of the mantle. The conclusion that the mantle must lose heat via convection rather than conduction or radiation is inescapable following an argument first proposed by Tozer (1965). This argument takes any reasonable estimate of the conductivity, optical opacity and temperature dependent viscosity the minerals making up the mantle and, when combined with the Earth's expected heat budget, leads to Rayleigh numbers indicating that the mantle will cool via convection (see Griffiths and Turner, 1998). The exact nature of this convection, and its relationship to plate tectonics is less clear, Davies (1999) has recently described the current state of knowledge regarding mantle convection and argues that two modes of mantle convection should be recognised. (1) A plate “mode” driven by internal radiogenic heating of the mantle and controlled by the descent of the brittle plates formed due to the strong temperature dependent viscosity of the mantle. In this mode of convection, up-welling at mid ocean ridges is passive and occurs over a large area while the down going parts of the convection cell is less extensive. (2) A plume mode, driven by heating at a lower thermal boundary layer taken to be the core mantle boundary and accounting for ~10% of the heat flux from the
Earth's interior and causing intraplate hot spots such as Hawaii. This mode of convection allows heat generated by the crystallization of the inner core to escape. There is evidence that both of these convection modes involve the whole mantle, but this is still a controversial assertion especially given the existence of geochemical evidence to the contrary. The geophysical and geochemical evidence for whole mantle and layered convection has recently been the subject of a Royal Society discussion meeting, at which a wide range of evidence was presented (see Davies et al., 2002). The conclusion, based on seismic tomography (e.g. Grand, 2002) and geodynamic modelling (e.g. Bunge et al., 2002) seems to be that large-scale whole mantle convection dominates and that the recognised geochemical reservoirs are not simply the upper mantle and lower mantle but are located elsewhere (Ballentine et al., 2002).
Chapter 2: Defects – the agents of crystal deformation

In the introduction, the dynamical behaviour of the Earth’s solid interior was described with the Earth’s mantle behaving as an extremely viscous convecting fluid. In order to understanding how the solid Earth behaves in this way, we must focus on defects in the minerals that make up the mantle. The importance of defects in controlling many aspects of mantle behaviour was described in the first chapter, in this chapter one aspect of defect behaviour is described in more detail – their action as agents of deformation in the solid state. The chapter begins by defining defects and considering the ways in which they move. This information is then brought together in order to understand how they allow the deformation of crystals and, ultimately, flow in the Earth’s interior.

Several books have covered this ground before and much of the material in this chapter is drawn from these. In particular Putnis (1992) provides a full survey of defects in minerals while Tilley (1987) considers point defects in detail and Hull and Bacon (1984) introduce dislocation theory. Poirier’s (1985) book on the creep of crystals provides an overview of models of deformation and introduction to rheology.

2.1 Point defects

Defects break the symmetry of the crystal containing them and are categorised by the dimensionality in which the symmetry breaking is extended. The simplest defect is a point defect where the local symmetry of the crystal is broken but the defect is localised to a single point. At its simplest a point defect is just an extra atom (an interstitial), a missing atom (a vacancy) or a “wrong” atom (an impurity) at a site in the crystal but it is also possible for multiple point defects to form together to make a defect cluster (figure 2.1).
Figure 2.1: Types of point defects. Atoms belonging to the crystal structure are shown in blue while those not in the chemical formula are pictured in red.

Point defects can be described in chemical reactions in several ways; Kröger-Vink notation (Kröger 1972) is particularly useful and is used throughout this thesis. In this notation, a defect is defined in terms of species, location and relative charge. The species and location uses standard chemical symbols with the species indicated in full sized text (with $V$ indicating a vacancy) and the location indicated with a subscript (with $I$ indicating an interstitial). Defects can have an effective charge, relative to the expected charge on the site of the defect, this is represented as a superscript with effective positive indicated by a dot, negative charge indicated by a prime symbol (') and neutral charge indicated by a cross. So, for example, $V_{Si}^{\infty}$ denotes a vacancy on the silicon site, which has an effective charge of $-4$ (four units of charge are absent), while $OH_{O}^{+}$ is a hydroxyl group at an oxygen site with a relative charge of $+1$ (oxygen ions have formal charge of $-2$ while OH groups have a formal charge of $-1$). Defect clusters are indicated by braces, so $\{V_{Si}^{-}+4OH_{O}\}^{+}$ is a cluster of four OH groups around a silicon vacancy with a relative charge unaltered compared to the bulk.
Having defined a nomenclature for describing point defects in chemical reactions, the likelihood of point defect formation can be rationalised by writing down the equations for the reactions and calculating (or estimating) the thermodynamics. In writing the equations, there are a few rules that must be followed (in addition to those governing “ordinary” chemical equations). The equation must conserve charge, i.e. the total effective charge on the right hand side of the equation must equal the total effective charge on the left otherwise the equation implies the creation of crystals with a net charge. The number of atomic species must be preserved (as in “ordinary” equations) and the atomic sites must be preserved. Two important ways of obeying these rules and making defects in ionic crystals are by forming Schottky and Frenkel defects. The Schottky defect involves forming a formula unit of vacancies and an extra formula unit of crystal, e.g. for MgO:

\[ \text{Mg}_{st}^n + O_{st}^n \rightarrow V_{st}^n + V_{st}^{n'} + \text{MgO} \, , \] (2.1)

and the Frenkel defect involves the creation of a vacancy and an interstitial, e.g. an Si interstitial in forsterite:

\[ Si_{st}^n \rightarrow V_{st}^{n'} + Si_{st}^{n''} \, . \] (2.2)

The important point about both of these reactions is that they obey the rules and preserve the stoichiometry of the crystals.

2.1.1 Formation of point defects

Ignoring so called “structural” point defects (which are just a convenient way to describe certain structures in terms of other structures e.g. the brownmillerite structure can be described in terms of the perovskite structure with structural oxygen vacancies) point defects can occur in three ways.

1) As the crystal grows it can incorporate impurity atoms into the structure. In terms of the point defect nomenclature, these can occur as impurities (on lattice sites) or interstitials (on interstitial sites). If the impurity replaces an ion of the same charge state (an isovalent substitution) then only one point defect need be created but, if the impurity ion is of a different charge state from the ion it replaces (a heterovalent substitution) then other defects must form to ensure charge neutrality.
For example if $\text{Al}^{3+}$ replaces $\text{Si}^{4+}$ in a magnesium silicate to form $\text{Al}^{\prime}_{\text{Si}}$, then another defect with net positive charge will form, for example we could have two $\text{Al}^{\prime}_{\text{Si}}$ defects charge balanced by one oxygen vacancy, or one $\text{Al}^{\prime}_{\text{Si}}$ defect balanced by a $\text{Al}^{\prime}_{\text{Mg}}$ defect. In any case, all of the defects are termed *extrinsic* defects.

(2) The second class of point defects occur when the crystal contains an element, such as iron, that can exist in more than one charge state within the crystal structure. The classic example is iron oxide where the iron can exist as iron III, iron II or as a combination of iron II and iron III in the same structure. The change in charge must be balanced by the formation of a compensating defect; for example in iron oxide the reduction of iron III to iron II increases the number of oxygen vacancies or reduces the number of oxygen interstitials. These types of defect are termed *non-stoichiometric* defects as the stoichiometry of the crystal changes on their formation.

(3) The first two classes of point defects require either the presence of an impurity or a change in the external chemical potential to form defects but there is a third type of defect that always exists, even in pure crystals. These *intrinsic* defects exist because of the increase in configurational entropy of a crystal on the formation of a point defect. Creating a point defect will involve an increase in the enthalpy of the crystal but also introduces disorder, increasing the entropy. In a defect free system at any finite temperature, the decrease in free energy caused by the increase in entropy will always tend to outweigh the increase in free energy caused by the increase in enthalpy and thus favour the formation of *some* defects. In fact, for a particular structure at a given temperature there will be an equilibrium number of defects found by expressing the configurational entropy as a function of point defect concentration. By writing down the free energy including this contribution, and finding the point defect concentration when the free energy is at a minimum, the equilibrium concentration can be found. For example Putnis (1992) derives an expression for the number, $n_s$, of
Schottky defects in a crystal with a formula MX (e.g. MgO or NaCl) at absolute temperature $T$:

$$n_s = N_0 e^{-\Delta H_s / 2kT},$$

(2.3)

where $N$ is the number of formula units in the crystal, $k$ is the Boltzmann constant and $\Delta H_s$ is the formation enthalpy of the Schottky defect (the energy change for equation 2.1). For ionic crystals with a more complex formula the derivation of expressions like equation 2.3 becomes very much more difficult because of the simultaneous requirement to maintain charge neutrality and minimise the free energy (which may involve a mixture of different types of defect). Ashcroft and Mermin (1976) provide details of the calculation required but the problem involves a constrained minimisation and is beyond the scope of this work. However, the number of intrinsic defects will always depend exponentially on the enthalpy of formation.

### 2.2 Movement of point defects - theories of diffusion

Diffusion is a ubiquitous process in many branches of science – in the Earth sciences alone diffusion is an important process for the movement of pollutants through aquifers, the transfer of heat and matter in the atmosphere and oceans and the transfer of heat, energy, momentum and matter in the Earth’s interior. Although the physical process governing diffusion may by different in all these situations, it is described by the same phenomenological relationships. A good review of diffusion with a geophysical bias can be found in Poirier (2000). The movement of matter through crystals is controlled by the diffusion of point defects through the crystal – a process that is described below.

#### 2.2.1 Phenomenology

From the phenomenological point of view, the diffusion of atoms through a solid can be described by the same relationships that are used to describe transport of matter or energy in the presence of a driving force in a wide range of situations. In the case of the diffusion of matter described here, the driving force can be thought of as a gradient in the concentration of the diffusion species, which is some kind of point defect. A simple example of solid-state
diffusion is pictured in figure 2.2. In this example, the flux of matter through an imaginary plane in the crystal, with a perpendicular concentration gradient, can be calculated by the appropriate application of Fick's Laws (only the one-dimensional case is described here). Fick's first law is appropriate in the situation where the concentration gradient does not vary with time. In one dimension it can be written as:

\[ J = -D \frac{dc}{dx} \]  

(2.4)

Here flux of matter (\( J \)) through a plane perpendicular to the \( x \) direction, in units of amount per unit time, per unit area of the plane, is proportional to the concentration gradient perpendicular to the plane (\( \frac{dc}{dx} \)). The constant of proportionality, \( D \), is known as the diffusion coefficient and varies as a function of, for example, temperature and pressure.

![Figure 2.2: Example of solid-state diffusion of one metal (e.g. gold) into another (e.g. silver), the system described by the upper diagram evolves with time following the lower graphs.](image)

In the situation where the concentration gradient changes with time Fick's second law can be used:

\[ \frac{dc}{dt} = \frac{d}{dx} \left[ D \frac{dc}{dx} \right] \]  

(2.5)
where \( t \) is time. Fick's laws can be generalised for a three dimensional system (Crank, 1975; Poirier, 2000), the diffusion coefficient is then a (second rank) tensor which must obey the crystal symmetry (Nye, 1957).

It is in the phenomenological context that most diffusion experiments are carried out; the aim is to find \( D \) for particular elements under controlled conditions. An example of a diffusion experiment is the determination of hydrogen diffusivity in forsterite (Demouchy and Mackwell, 2003). Here dry forsterite is heated in the presence of water for a measured time, the sample is quenched and the concentration of hydrogen across the sample measured by IR spectroscopy. The diffusion profile (concentration of hydrogen against depth in the sample) is fitted to an appropriate solution to Fick's law to extract the diffusion coefficient. Experiments at a range of temperatures give the temperature dependence of diffusion.

The diffusion of other elements requires slightly different analysis techniques and other sources of the diffusing species. In particular, chemical analysis is of no use if the diffusing species is one of the components of the mineral. In such cases self diffusion occurs, but is not measurable. The solution is usually to use an isotopic tracer, a rare isotope of the element of interest is used to enrich the source region and the concentration profile of this isotope is measured using, for example, secondary ion mass spectroscopy. Commonly \(^{17}\text{O}\) is used as a tracer for oxygen diffusion and \(^{30}\text{Si}\) is used as a tracer for silicon diffusion. The diffusivities measured by tracer diffusion are assumed to be equal to the self diffusion diffusivities apart from the isotope mass effect, which is difficult to study and only rarely considered (e.g. Freer et al., 1997). A second approach is available if two ions can exist on the same site, for example magnesium and iron in olivine. In this case inter-diffusion can be measured. For these studies two samples one enriched in e.g. iron and the second enriched in magnesium are placed together to form a "diffusion couple" and heated. Iron diffuses from the iron rich sample into the iron poor sample while magnesium diffuses in the opposite direction and the diffusion profile across the couple can be measured to extract diffusion co-efficients. However, diffusion coefficients vary with composition of the crystal so additional analysis is needed to extract the self-diffusion of the elements being studied. This relationship is discussed in the
case of olivine by Chakraborty (1997), who was able to reconcile magnesium tracer diffusion data with results from inter-diffusion experiments.

2.2.3 Thermodynamics of diffusion

When diffusion coefficients are measured experimentally at various temperatures they are found to show an Arrhenius behaviour, with a plot of the natural logarithm of the diffusion coefficient against the reciprocal temperature being a straight line (see figure 2.3). The details of diffusion are given below but to understand the empirical observation it is necessary to anticipate one result – that diffusion involves an addition of energy to the system in order to allow the ion to move. This can be thought of as the energy needed to squeeze an atom between two other atoms as shown in figure 2.4, the difference in energy between the ground state where the energy is at a minimum and high energy activated state where the energy is at a maximum is the activation energy for the hop. A simple analysis (see, for example Putnis, 1992) relies on the fact the probability, $p$, that an atom has energy $E$ above the average thermal energy $kT$ ($k$ is the Boltzmann constant) is given by:

$$
p \propto e^{-E/kT}.
$$

(2.6)

The diffusion rate will be proportional to this probability, so the relationship between the diffusion coefficient, $D$, and the temperature, $T$, can be written:

$$
D = D_0 \exp\left(-\frac{E}{RT}\right),
$$

(2.7)

where $R$ is the gas constant (so that the energy is expressed per mole rather than per atom), $E$ is the activation energy that describes the effect of temperature and $D_0$ is the pre-exponential factor which depends on other factors that control diffusion rate.
Figure 2.3: Example of an Arrhenius diagram for diffusion, the sharp kink in the line represents the change between two diffusion regimes.

Figure 2.4: An atom squeezing between two others as an elementary step in diffusion, the system passes through an activated state with high energy.

The activation energy introduced in equation 2.7 is no more than an empirical relationship between the diffusion coefficients and temperature. However, it is the basis of a thermodynamic description of diffusion. In particular, the activation energy can be considered a free energy and the variation in activation energy with temperature (a small effect resulting in a curve in plots like figure 2.3) and pressure can be expressed using the relationship:

\[ G^* = E^* + PV^* - TS^* \]  

(2.8)

where \( G^* \) is the activation free energy, \( V^* \) is the activation volume, \( S^* \) is an activation entropy term, \( E^* \) is the internal activation energy (i.e. the measured
value at zero Kelvin and zero pressure) \( T \) is the temperature and \( P \) is the pressure. \( S^* \) is generally small enough to be neglected leaving \( V^* \) to describe the effect of pressure.

The activation energy described above assumes that there is no energetic cost associated with forming the diffusing defect. This is equivalent to assuming that the diffusing defect is present in the crystal because it compensates for the presence of charged impurities. Diffusion under these conditions is described as extrinsic. At high temperature, the dominant contribution to the population of diffusing defects may be entropic. Diffusion is then termed intrinsic and the activation energy consists of two contributions, a fraction of the formation energy of the diffusing defect and the height of the migration energy barrier (see Vocadlo et al., 1995; Wright, 2003). In materials containing ions with variable valence, a third diffusion regime is possible: “transition-metal extrinsic diffusion” (Chakraborty 1997) or “constrained extrinsic diffusion” (Jaoul, et al. 1995). The diffusing defects arise in response to a change in the internal or external chemical potential. The activation energy for diffusion in this regime must include a contribution for the induced redox process.

### 2.2.4 Diffusion at the atomic scale

Fick’s Laws and the thermodynamic analysis above relate to the diffusion on the macroscopic scale observed in experiment, but in order to progress the understanding of diffusion further it is necessary to consider the discrete atomic nature of the crystal in which diffusion is occurring. On this scale diffusion can be considered to be the cumulative effect of a large number of distinct events on the atomic scale, and the diffusion mechanism\(^3\) described by the dominant event. Figure 2.5 outlines the most common atomic scale mechanisms; the simplest is the exchange mechanism (a) where two atoms swap positions in the crystal. A slightly more complex mechanism involves exchange between a number of ions forming a coupled exchange (b). The energy barriers for exchange mechanisms are often too high to allow significant diffusion, in these

---

\(^3\) Some authors use the term “diffusion mechanism” to distinguish between intrinsic and extrinsic diffusion. To save confusion I refer to this as the diffusion regime and reserve mechanism for the atomic scale mechanism described here.
cases the motion of point defects dominates. In close packed materials a common mechanism is vacancy diffusion (c), this mechanism involves an ion moving from its lattice site into an adjacent vacancy. The vacancy is preserved but moves to the once occupied lattice site enabling continued diffusion. Diffusion of interstitials are also possible, if an interstitial moves from one interstitial site to another (d), then such diffusion is termed interstitial diffusion but if the interstitial moves onto an occupied lattice site forcing the lattice ion into a second interstitial position (e) then the mechanism is interstitialcy diffusion.

![Diagram of atomic scale mechanisms for diffusion in crystalline solids](image)

**Figure 2.5:** Atomic scale mechanisms for diffusion in crystalline solids; exchange (a and b), vacancy (c) and interstitial (d and e) mechanism are shown, see text for details.

The atomic scale mechanism is very hard to determine by experiment. In chapter 5 various ways to study diffusion from atomic scale computer models are discussed and one of the major advantages of these methods is their ability to resolve the mechanism.
2.3 Line defects

The concept of the crystal dislocation, the only class of line defect considered in this thesis, arose from observations regarding plastic strain in the 1930s⁴. A simple picture of a deforming crystal involves breaking a plane of bonds and moving the atoms above the plane relative to the atoms below the plane (figure 2.6 a-c). However, estimates of the strength of crystals from this model are thousands of times higher than the observed strength. The solution is to invoke the presence of a dislocation, identified here as an extra half plane of atoms (although a better definition is given below) in figure 2.6 d. With the presence of the dislocation, movement can be achieved by breaking far fewer bonds (figure 2.6 d-f) and estimates of the strength of crystals become reasonable.

2.3.1 The geometry of dislocations

The dislocation was introduced above as the termination of an extra half plane of atoms, this definition is not particularly useful as the half plane for a dislocation can be defined in many ways (figure 2.7) and because dislocations exist which do not share the geometry of the dislocation in figure 2.6. Figure 2.8 shows the various terms used to describe a dislocation based on the concept of the Burgers circuit – a closed circuit in a perfect crystal which, when looped around a dislocation line is no longer closed. The vector between the start and end points of the circuit is known as the Burgers vector, which is the most fundamental property of a dislocation. If the Burgers vector is a whole multiple of a lattice vector of the crystal, then the dislocation only affects the crystal structure close to the dislocation line and the dislocation is termed a perfect dislocation. Imperfect, or partial, dislocations are possible and lead to the formation of planar defects such as stacking faults and coincidence boundaries. If the Burgers vector is perpendicular to the dislocation line, then the dislocation has an edge character but in cases where the Burgers vector is parallel to the dislocation line, the dislocation has a screw character. For edge dislocations, the plane containing the dislocation line and Burgers vector is known as the slip

---

⁴ The concept of dislocations being responsible for plastic strain was independently realised by Orowan, Polanyi and Taylor in 1934 (Hull and Bacon, 1984)
plane because this is the plane along which the crystal may be deformed by the dislocation (see figure 2.8). The Burgers vector can also be at some other orientation to the dislocation line giving dislocations with a mixed character.

The Burgers vector and slip plane can be used to concisely describe a dislocation. Through this thesis the convention is to describe an edge dislocation by the Burgers vector’s length, its direction and the slip plane, thus in an orthorhombic crystal a perfect edge dislocation with a dislocation line along [100] and a Burgers vector orientated along [010] would be described \( \mathbf{b}(010)(100) \), a screw dislocation with the same dislocation line would be described \( \mathbf{a}(100) \) (the slip plane is not defined).
Figure 2.6: plastic deformation of perfect (a – c) and dislocated (d – f) crystal. For the perfect crystal, in order to displace the top half of the crystal with respect to the bottom half a whole plane of bonds must be broken (red circles) but in the dislocated case the elemental step only involves the breaking of bonds close the core (extra half plane in blue).
Figure 2.7: Non-uniqueness of the extra half plane. The same edge dislocation (i.e. the same Burgers vector, dislocation line and slip plane) is shown in sections a and b but is described by the addition of two separate half planes of atoms (dark blue) with adjacent planes highlighted (light blue).

Figure 2.8: Definition of the Burgers vector. A closed circuit in a perfect crystal (a) marked by black arrows is not closed if it encircles a dislocation line (b) marked by X and perpendicular to the page. The Burgers vector is the vector to close the circuit (red arrow) and the plane containing this and the dislocation line is the slip plane (perpendicular to the page, trace marked as a grey dashed line).
2.3.2 Movement of dislocations

The dislocation line defines a boundary between slipped and unslipped portions of the crystal and must extend to the edge of the crystal or form a closed loop. For the crystal to deform, the slipped portion must grow so the dislocation must move. One way for a dislocation to move is by dislocation glide, movement of the dislocation in the glide plane as described in figure 2.6, but other modes of dislocation movement are possible. The glide motion of dislocations still requires some bond breaking or, alternatively, requires a change in the configuration of the dislocation core. This implies an energy barrier to glide termed the Peierls barrier, which is anticipated to be periodic (see figure 2.9). In addition, the whole dislocation is unlikely to move at a single time, instead kinks form which allow small portions of the dislocation to slip with a lower critical energy. As well as kinks – which displace the dislocation line in the glide plane – jogs exist (figure 2.10) which displace the dislocation line perpendicular to the glide plane and allow dislocation climb (dislocation movement out of its glide plane). A critical factor of climb is the need for an interaction with point defects, as the dislocation adsorbs vacancies the dislocation will climb upwards, if it adsorbs interstitials (or ejects vacancies) climb is downwards.

![Diagram of dislocation glide and Peierls barrier](image)

**Figure 2.9:** Glide of edge dislocation by the sideways propagation of kinks, view is of the slip plane with dislocation line in red and Burgers vector marked by the arrow on the left. Peierls valleys (low energy configuration of the core) are indicated on the slip plane by solid lines while the high-energy configuration corresponds to the dashed lines. A dislocation with kinks (a) moves forward by sideways movement of the kinks minimising the length of the dislocation line in an unfavourable geometry (transition from a to b) but eventually a new kink must be nucleated (c).
Figure 2.10: Movement of the dislocation line out of the slip plane is termed climb and achieved by interaction of the dislocation with point defects. Part (a) shows a half plane of atoms making up an edge dislocation with the dislocation line horizontal and the Burgers vector coming out of the page, the edge dislocation adsorbs a vacancy (b) to form a jog, locally moving the dislocation line out of the slip plane. Further vacancies result in the climb of the dislocation line (c) to a new glide plane parallel with the first.

2.4 Planar defects

Although planar defects can be extremely important in permitting deformation, this thesis does not describe any studies of them. Briefly, planar defects can be separated into several categories: interfaces or grain boundaries where two crystals are in contact, stacking faults left behind by the glide of a partial dislocation and twinning planes where one part of a crystal is separated from another by some defined symmetry operation. Putnis (1992) has a detailed discussion of planar defects and Poirier (1985) describes their importance for allowing deformation.

2.5 Rheology, deformation models and constitutive equations

Rheology is the study of the behaviour of materials undergoing deformation, which depends on the structure of the material at all scales. However, rheology is limited to the description of the deformation according to empirical "materials parameters" and the effect of external variables like temperature and not the
detailed structure of the material. Rheological behaviour can be divided into several categories including:

(i) Elastic behaviour, where the deformation is recovered on removal of the stress. In the limit of small deformation linear elasticity, with stress proportional to strain (with the constants of proportionality, or elastic constants, used to describe the relationship), dominates deformation.

(ii) Viscous behaviour, where a fixed stress results in a constant deformation rate, liquids tend to follow this behaviour.

(iii) Plastic behaviour, where deformation is not recovered on removal of stress, the stress – strain relationship may be complex but can often be expressed as a power law (see below).

When considering the plastic deformation of crystals at the atomic scale two classes of deformation mechanism are identifiable, \textit{diffusion creep}, where the principal mode of deformation is due to the formation and movement of point defects and \textit{dislocation creep}, where the movement of dislocations dominate. In general, materials will behave differently under different conditions of stress, temperature and pressure and one mathematical equation will be insufficient to describe their behaviour. However, when one particular atomic mechanism is active, it should be possible to express the strain rate $\dot{\varepsilon}$ as a function of stress ($\sigma$) and various other variables (including temperature, $T$, pressure, $P$, and possibly variables like chemical composition and external chemical potential):

$$\varepsilon = f(\sigma, T, P...) . \quad (2.9)$$

The form of equation 2.9 will depend on the atomic scale mechanism operating; often the process is thermally activated and so follows the relation:

$$\dot{\varepsilon} \propto \sigma e^{-\frac{\sigma}{RT}} , \quad (2.10)$$

where $R$ is the gas constant. Plotting log $\dot{\varepsilon}$ against log $\sigma$ tends to give a straight line over a limited range of data. Such creep, termed power law creep, can be described by an equation like:

$$\dot{\varepsilon} = \dot{\varepsilon}_0 \sigma^n e^{-\frac{\sigma}{RT}} , \quad (2.11)$$
with \( n \) between 1 (for viscous flow) and about 6 (for a strong stress dependence). Although \( Q, n \) and \( \dot{\varepsilon}_0 \) describe a particular deformation style, knowledge of these parameters is not enough to unambiguously assign a atomic scale mechanism, a point emphasised by Poirier (1985).

The atomic scale mechanisms can be further sub-divided. Dislocation creep is rate limited by the obstacles the dislocation must overcome to move, which could be the Peierls barrier or other objects such as point defects or even dislocations moving on other planes. If the deformation is controlled by movement along the glide plane then deformation is called dislocation glide but climb controlled slip is also possible. In either case deformation rate is insensitive to crystal grain size but depends on the presence of a source of new dislocations. Climb controlled creep is limited by the diffusion of point defects between dislocations climbing in opposite directions, and so should depend on the dislocation density, which may vary as a function of stress. In some materials (e.g. the BCC metals) dislocation glide is limited by the movement of screw dislocations, a process called cross slip recently demonstrated to operate by computer simulation (Marian et al., 2004). By contrast, diffusion creep is grain size dependent. The classic example of diffusion creep is termed Nabarro–Herring creep and involves the diffusion of point defects, generated at interfaces by non-hydrostatic pressure, through the crystal bulk. The strain rate is inversely proportional to the square of the grain diameter. Coble creep is essentially the same process but diffusion is along the grain boundaries and the strain rate dependence is inversely proportional to the cube of the grain size.

### 2.7 The deformation of olivine

Because it is assumed that the deformation of olivine determines the rheology of the upper mantle, there have been a large number of experiments to determine its behaviour. Early work, reviewed by Frost and Ashby (1982), Poirier (1985) and Kohlstedt and Goetze (1974), indicates that easy slip occurred along \([001](100)\) dislocations with \([100](010)\) dislocations also observed. The occurrence of \([010](100)\) dislocations, with a very long Burgers vector, was also reported. However, the results of the early studies diverged on extrapolation from the measured strain rates measured in the laboratory at high deviatoric
stress (1-10 kbars) to the much lower deviatoric stresses expected in the mantle (100 bars).

Work through the mid to late 1970’s and early 1980’s clarified the situation by undertaking single crystal deformation experiments under conditions of low deviatoric stress (Kohlstedt and Goetze, 1974; Durham and Goetze, 1977; Durham et al., 1977; Durham et al., 1979; Darot and Gueguen, 1981; Gueguen and Darot, 1982). These studies of dislocation microstructures were helped by the existence of decoration techniques, first developed for iron bearing olivine (Kohlstedt et al., 1975) then extended to cover iron free forsterite (Jaoul et al., 1979). These studies, which have been reviewed by Cordier (2002), revealed a complex dislocation microstructure in deformed olivine, with high temperature slip involving dislocations with Burgers vectors along [100] moving on several planes. Observations of cross slip between these planes were common. The studies showed that the [001] dislocations observed in the earlier work are characteristic of low temperature high stress deformation and that slip of dislocations with this Burgers vector occurs in other planes too. Cordier (2002) also notes that [100] screw dislocations have been observed to take on rather different characters depending on the glide plane of the associated edge dislocations, with the [100] screw dislocations associated with (001) glide planes being “straight” and those associated with (010) glide being bent. These studies also revealed a power law rheology for polycrystalline olivine with a flow law (equation 2.11) that depends on the orientation of the crystal, in dry olivine \( n \) is around 3, \( \dot{\varepsilon}_0 \) is between \( 10^6 \) and \( 10^7 \) and \( Q \) is of the order 550 kJmol\(^{-1}\) (Darot and Gueguen, 1981).

The interpretation of this data is complex – the dislocation microstructure seems to rule out simple glide while the activation energy does not seem to correspond to the activation energy for diffusion of any species in olivine. The first real attempt to understand the theoretical basis of plasticity in olivine was undertaken by Poirier (1975), this study considered the pseudo HCP packing of oxygen in the olivine lattice and compared it to simple metals. The comparison suggested the possibility of [100] screw dislocations dissociating onto the prism planes of the HCP lattice; indeed such dislocations have been observed in experiments. These dissociated screw dislocations may cross slip onto other
planes, a process accompanied by the formation of stacking faults and termed pencil glide. As one of the possible mechanisms for the onset of plasticity in metals is the beginning of movement of straight screw dislocations by pencil glide, this leads to a possible process constraining the deformation process. The model also predicts that pressure should enhance cross slip because the partials do not quite lie in the slip plane. It is interesting to compare this picture with recent studies of the motion of screw dislocations in BCC iron (Marian et al., 2004), the canonical exemplar of pencil glide. This work shows that the kink-pair mechanism, a competitor to the pencil glide theory, dominates in iron at low strain rates and a new and unexpected cross-kink mechanism takes over with increasing strain rate. What do these conclusions say about the proposed pencil glide mechanism in olivine? In any case, studies since 1975 suggest an alternative deformation mechanism in olivine that rather than being limited by the barrier to screw dislocation cross slip, olivine’s plasticity is limited by dislocation climb – which is in turn limited by the diffusion of point defects. A theoretical model of this process was presented by Jaoul (1990). Jaoul’s model seems to reconcile experimental measurements of point defect diffusivities (reviewed in chapter 5) with the activation energy for deformation, but the measured activation energy of silicon in olivine, and hence the need for the model, has recently been questioned (Dohmen et al., 2002).

Further work attempted to pin down the role of hydrogen in weakening olivine. Experiments show that many minerals without structural hydrogen can accommodate trace amounts of hydrogen in the structure. The properties, in particular the rheology, of these nominally anhydrous minerals (NAMs) can be dramatically changed by the presence of hydrogen, an effect first noted in quartz (Griggs and Blacic, 1965) and later in olivine (Mackwell et al., 1985). A TEM investigation of the dislocation microstructure of single crystal San Carlos olivine deformed under “wet” conditions by Mackwell et al. (1985) suggested that the hydroxyl enhanced dislocation climb and that both (010)[001] and (010)[100] slip systems were active in their experiments. Recent work on

---

5 Pencil glide is the movement of dislocations by the formation of double kinks in screw dislocations that have spread into several glide planes. The formation of the double kinks allows the movement of the dislocation (see Poirier, 1985).
polycrystalline samples (Mei and Kohlstedt, 2000a, 2000b) shows that water enhances diffusion rates under conditions where both diffusion creep and dislocation climb are presumed to be the dominant deformation mechanism.

Recent research in situ at high pressure involving the simultaneous deformation of two olivine samples with different grain sizes reveals that the deformation of olivine is not sensitive to grain size under mantle conditions (Li et al., 2003b). This observation rules out some simple diffusion controlled deformation mechanisms (such as Coble and Nabarro – Herring creep) but not deformation controlled by dislocation glide or climb (in turn controlled by the diffusion of point defects). In conclusion, the details of the deformation mechanism of olivine are not well understood – the application of computer modelling techniques described in the following chapters may allow useful new insights into the deformation process at the atomic scale.
Chapter 3: Methods in Computational Mineralogy

Over the past thirty years or so, simulation and modelling materials on the atomic scale has become a standard method used by the condensed matter physics, solid-state and materials chemistry and mineral physics communities. This approach to science has three major uses: (a) It allows the properties of materials to be predicted without recourse to experiment. (b) It allows an inherently atomic scale view of a process to be gained, which is often of utility in interpreting experiment. (c) It allows the testing of theories that have been derived to explain experiment. The success of modelling is not due to any major change in the basic underlying physical theories of matter on this scale (quantum theory is, after all, nearing its first centenary) but rather because of improvements in the implementation of the theories and a massive increase in the capacity of computers. In this chapter, some of the available methods for the simulation of minerals are described, with particular emphasis placed on methods used for the study of defects in later chapters. This chapter is necessarily an incomplete and biased survey of the subject, although I have attempted to weight the description of each method to its prominence in the later chapters.

Two basic physical models are commonly used for atomic scale simulation. The ideal method is to base the description of the simulated system in quantum mechanics. The adiabatic Born-Oppenheimer approximation is assumed and only the electrons are considered to behave as quantum particles; the wave function for a system consisting of electrons in the potential of the atomic nuclei is found and used to calculate the energy of the system. This approximation relies on the fact that the electrons are very much lighter and more mobile than the nuclei, so that for any change in the arrangement of the nuclei the electrons rapidly move to their new configuration. This rapid movement is approximated as an instantaneous rearrangement of the electrons, which allows an energy hypersurface to be defined that only depends on the nuclear degrees of freedom.
The second method replaces the complexities of describing the electrons with relatively simple parameterised functions describing the energy of interactions between atoms. This second method requires far less computational resources per simulated atom than the quantum based methods – a distinct advantage if the simulation of systems containing many atoms (and electrons) is the aim. Of course, there are disadvantages associated with simply averaging out the effects of the electrons as some simple set of parameters. Note that the Born-Oppenhimer approximation is implicit in parameterised models. Whichever method is used, a simulation can be separated into three distinct parts. First, the total energy of the system is calculated and this energy defines a point on the energy hypersurface corresponding to the starting atomic or nuclear positions. Then the atoms or nuclei must be moved across the hypersurface so as to minimise the energy, this requires repeated calculations of the energy because the shape of the hypersurface is not known in advance. The first two steps are repeated iteratively and, finally, properties of the simulated system can be found.

By applying periodic boundary conditions to a simulation cell commensurate with the crystal’s unit cell a model of an infinite perfect crystal can be designed. This is also the starting point for models of defective crystals, where the translational symmetry is lost. For this reason much of this chapter considers modelling techniques for the perfect crystal (the next four sections of the chapter deal with each of three stages of a simulation in turn, with the calculation of energy described separately for the quantum mechanical and atomistic methods) before turning to an outline of the methods available for dealing with the effect of temperature and modelling defects in solids.

3.1 Quantum mechanical models

The application of quantum mechanical methods to study condensed matter at the atomic scale is described in a number of textbooks (Forsman and Frisch, 1996; Jensen, 1999; Leach, 2001), review articles (Segall et al., 2002; Abarenkov, 2003; Harrison, 2003; Postnikov, 2003) and other publications (Oganov, 2002; Kohn, 2003; Pople, 2003) from which much of the material in this section is taken. The aim of this section is to provide a general overview of
the use of quantum mechanical methods, not to provide a rigorous description of
the underlying physics, in order to place the later chapters in context. The total
energy of a system of electrons and nuclei can be found by solving the (time
independent, non relativistic) Schrödinger equation:
\[ \hat{H}\Psi = E\Psi, \tag{3.1} \]
where \( \hat{H} \) is the Hamiltonian operator, \( E \) is the total energy and \( \Psi \) is the
wavefunction. The Hamiltonian operator:
\[ \hat{H} = \left( \sum_{i} -\frac{1}{2} \nabla_{i}^{2} + \sum_{i} v_{e-n}(\mathbf{r}_{i}) + \sum_{i} \sum_{j \neq i} \frac{1}{|\mathbf{r}_{i} - \mathbf{r}_{j}|} \right), \tag{3.2} \]
is composed of three parts, the first term on the right hand side is the kinetic
ergy operator for the electrons, the second describes the potential due to the
nuclei felt by the electrons and the third describes the Coulombic electron-
electron interactions. Atomic units are used, \( N \) is the number of electrons, \( \mathbf{r} \) is a
vector describing their spatial positions (for simplicity the spin co-ordinate is
neglected throughout) and the Coulombic nuclei – nuclei interactions must be
added. Despite the apparent simplicity of the Hamiltonian (equation 3.2), the
solution of equation 3.1 for systems with large numbers of electrons is difficult
because there are \( 3N \) electronic coordinates in the extremely complex
wavefunction. Although numerical solutions to the many body problem are
becoming possible using, for example, the Quantum Monte Carlo approach (see
Foulker et al., 2001 for a review) most practical calculations carried out today
replace the many electron Schrödinger equation with \( N \) single electron
Schrödinger like equations with each electron experiencing an effective field
\( (v_{\text{effective}}) \) of the other electrons:
\[ \hat{H}_{i}\psi_{i} = \left( -\frac{1}{2} \nabla_{i}^{2} + v_{e-n}(\mathbf{r}_{i}) + v_{\text{effective}} \right) \psi_{i} = E_{i}\psi_{i}. \tag{3.3} \]
The problem becomes one of finding suitable ways of describing the effective
electron interactions, and of describing the total wavefunction from the \( N \) single

---

6 Atomic units simplify the equation by setting the electronic mass \( (9.110 \times 10^{-31} \text{ kg}) \), charge
\( (1.602 \times 10^{-19} \text{ C}) \) and action \( (1.055 \times 10^{-34} \text{ Js}) \) to unity and deriving the related units, such as
energy (the Hartree = \( 4.360 \times 10^{-18} \text{ J} \)) and length (the Bohr radius = \( 5.292 \times 10^{-11} \text{ m} \)). See
appendix D of Jensen (1999).
electron wavefunctions $\psi_i$ with energy $\epsilon_i$. Another issue is that each one-electron equation relies on the solutions to all the others. The trick is to solve the equations together using the self-consistent field (SCF) approach. An initial guess is used as a starting point and used to calculate the effective interactions, this allows the parameters to be changed and a new, better, approximation to the effective interaction to be calculated. Iterations through this cycle result in a solution to all of the equations together. At the end of each cycle, the total wavefunction can be built and diagonalised in order to extract the energy.

3.1.1 Hartree-Fock theory

Hartree (1928) proposed using a simple approximation for the effective electron interactions, that of the classical Coulomb interaction between the electron of interest and an inhomogeneous electron gas (the potential due to this interaction is known as the Hartree potential). The Schrödinger like equations then have the form:

$$\hat{H}_i\psi_i = \left( -\frac{1}{2} \nabla_i^2 + v_{e-n}(r_i) + \int \frac{\rho(r) dr}{|r - r'|} \right) \psi_i = \epsilon_i \psi_i,$$

where the third term in equation 3.4 is the Hartree potential and the other terms are as defined in equation 3.3. This approximation neglects exchange (spin - spin interactions) and correlation (that the interaction depends on the position of the other electrons, not just an average of the electron density) of electrons. In this approximation, the total wavefunction is just the product of the $N$ one-electron functions.

Fock (1930) added exchange to the Hartree approach, creating Hartree-Fock theory, which is used today and is the basis for many high accuracy methods in quantum chemistry. The key addition is the concept of exchange, which forces the wavefunction to obey Fermi statistics. The Schrödinger like equations then have the form:

$$\hat{H}_i\psi_i = \left( -\frac{1}{2} \nabla_i^2 + v_{e-n}(r_i) + \int \frac{\rho(r) dr}{|r - r'|} \right) \psi_i$$

$$-\sum_j \delta_{i,j} \int dr \frac{\epsilon_j(r')}{|r - r'|} \epsilon_j(r) = \epsilon_i \psi_i,$$

\(48\)
which is just the equation from Hartree’s description with the addition of the exact exchange potential ($\delta_{ij}$ is the Kronecker delta function and $\epsilon_i^e$ is the complex conjugate of the energy of the $i^{th}$ orbital). The exchange term depends on all the other electrons in the system and so is non-local, leading to a computationally expensive method. The total wavefunction is a single determinant (called the Slater determinant):

$$\psi_{HF} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(r_1, \sigma_1) & \psi_1(r_2, \sigma_2) & \cdots & \psi_1(r_N, \sigma_N) \\ \psi_2(r_1, \sigma_1) & \psi_2(r_2, \sigma_2) & \cdots & \psi_2(r_N, \sigma_N) \\ \cdots & \cdots & \cdots & \cdots \\ \psi_N(r_1, \sigma_1) & \psi_N(r_2, \sigma_2) & \cdots & \psi_N(r_N, \sigma_N) \end{vmatrix},$$

(3.6)

with spatial ($r$) and spin ($\sigma$) co-ordinates explicitly represented. This means exchanging any two electrons, which corresponds to swapping columns in equation 3.6, results in a change of sign of the wavefunction. This is the correct behaviour for fermions (spin half particles like electrons); the wavefunction is antisymmetric and exchange is exact. This means that in the Hartree-Fock approximation, electrons of the same spin are prevented from being in the same place but there is no such effect for electrons of opposite spin, so there is an exchange hole but no correlation hole around each electron.

The lack of a treatment for correlation can be overcome by various refinements to Hartree-Fock theory collectively known as post-Hartree-Fock methods (Foresman and Frisch, 1996; Jensen, 1999; Pople, 2003). These methods, which include Möller-Plesset perturbation theory, the configuration interaction methodology and the coupled cluster approach are extremely demanding in terms of computer time, with scaling as bad as the number of electrons raised to the seventh power (Hartree-Fock theory scales as the number of electrons to the fourth power). Additional problems are that the configuration interaction method is not size consistent and the coupled cluster approach is not variational. These problems have prevented the use of post-Hartree-Fock methods for studies of condensed matter. However, an approach that is mathematically (but not practically) exact, computationally fast, popular and applicable to the periodic solid is density functional theory (DFT).
3.1.2 Density functional theory

DFT is a method that, in principle, solves the many body Schrödinger equation exactly, including both exchange and correlation, without recourse to the expensive post-Hartree-Fock wavefunction methods described above. This is possible because of the existence of two theorems, proposed and proved by Hohenburg and Kohn (1964), (1) that the external potential of a system of electrons is a unique functional\(^7\) of the electron density, and (2) that the ground state and energy of the system is variational with respect to the electron density. The idea of using the electron density to determine the properties of a system was not new, it dates back to the earlier Thomas-Fermi method but, in proving these two theorems, Hohenburg and Kohn showed that such an approach leads to a formally exact method to extract the total energy (and any other ground state property) without the need to construct the many body wave function. Kohn and Sham (1965) then recast the problem in terms of Schrödinger like equations for single non-interacting electrons moving in the mean field of the other electrons (like equation 3.3) which leads to improved accuracy in determining the electronic kinetic energy and is the foundation of modern DFT.

Within the framework of DFT as defined by Hohenburg and Kohn the exact energy is written as:

\[
E = E[\rho(r)] = \int \text{d}r V_{\text{ext}}(r)\rho(r) + F[\rho(r)],
\]

(3.7)

where both the external potential, \(V_{\text{ext}}\), and the "universal" functional, \(F[\rho(r)]\), are functionals of the electron density. The universal functional \(F[\rho(r)]\) can be conveniently separated into three parts:

\[
F[\rho(r)] = E_K[\rho(r)] + E_H[\rho(r)] + E_{xc}[\rho(r)],
\]

(3.8)

a kinetic energy contribution, a electron-electron interaction contribution (equal to the Hartree energy described in section 3.1.1) and the exchange-correlation contribution respectively. The exchange-correlation functional is defined to make up all missing contributions, but is roughly equal to missing contributions

---

\(^7\) The functional is unique up to an additive constant. A functional is a function of a function, denoted by \(F[f(x)]\), the functional \(F\) is a function of \(f\), which is a function of the variable \(x\).
from Hartree's method. All that is needed is to vary the electron density
distribution in order to minimise the energy found by the application of equation
3.7 and 3.8; this will give both the ground state energy and the ground state
electronic structure. Although this formulation of DFT has succeeded in
defining the energy of an \( N \)-electron system (which requires solution in \( N \)-
dimensional parameter space) in terms of a three dimensional minimisation
problem, the results are not accurate. The reason is that no accurate functional
for calculating the kinetic energy term in equation 3.8 from the electron density
has yet been found.

The Kohn-Sham approach solves the kinetic energy problem. The solution is to
express the total electron density as a sum over one-electron wavefunctions, we
then need to solve \( N \) Schrödinger like equations (using the same SCF and
diagonalisation methodology described in section 3.1). This means that the
electron density at a given point is the sum of the probability of finding an
electron form any of the one-electron orbitals at that point:

\[
\rho(r) = \sum_{n=1}^{N} \psi_n^*(r) \psi_n(r),
\]

but, unlike in Hartree-Fock theory, the orbitals have no physical meaning\(^8\). The
kinetic energy term in equation 3.8 can be replaced with the sum of one-electron
kinetic energies from the Kohn-Sham one electron Schrödinger like equations.
This is a very good approximation and any energy deficit can simply be
subsumed into the exchange-correlation functional.

Up to now the form of the exchange-correlation functional has not been
described. While it has been shown (by Hohenburg and Kohn) that an exact
functional must exist its form remains unknown. Approximating the exchange-
correlation functional is the only inaccuracy in DFT that does not arise from its
numerical implementation. Kohn and Sham introduced a very simple, but at the
same time astonishingly\(^9\) accurate approximation, that of the exchange-

\(^8\) Except that they are connected to the true electron density \textit{via} equation 3.9 and the magnitude
of the energy of the highest occupied orbital relative to the vacuum is equal to the ionisation
energy. However, the Kohn-Sham orbitals are often, usefully, used as a proxy for physically
meaningful orbitals.

\(^9\) According to Kohn (1999)!
correlation energy of a uniform electron gas (called the local density approximation, or LDA). The energy arising from exchange and correlation at a point of electron density \( \rho \) is equal to the exchange and correlation energy in a system of electrons of uniform density \( \rho \):

\[
E_{xc}[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) e_{xc}(\rho(\mathbf{r})).
\]  

(3.10)

This is known very accurately from a combination of exact calculation and Monte Carlo simulation (Ceperley and Alder, 1980; Oritz and Ballone, 1994). The LDA functional has been parameterised by Perdew and co-workers (Perdew and Zunger, 1981; Perdew and Wang, 1992). The use of the LDA is usually found to lead to over binding in condensed systems (the total binding energy is too high); it predicts bond lengths that are too short. It is, however, sufficiently accurate for many purposes, often because there is a cancellation of errors between a small underestimate of the exchange energy balanced by a large overestimate of the smaller contribution due to correlation (Kohn, 2003). A common refinement to the LDA is to include a dependence on the local gradient of the electron density in the exchange correlation functional. This approach, known as the Generalised Gradient Approximation (GGA), often leads to slight under binding (bonds that are too long and energies that are too low). Popular GGA parameterisations include “PW91” (Wang and Perdew, 1991) and “PBE” (Perdew et al., 1996). Hybrid functionals, such as “B3LYP” (Lee et al., 1988; Becke, 1993), combine the exact exchange of Hartree-Fock theory with an approximate DFT exchange-correlation functional. Functional development is an active field with further improvements in accuracy expected.

3.1.3 Implementing electronic structure calculations

Both the Hartree-Fock and Kohn-Sham approaches described above do not allow, by themselves, the simulation of solids; some additional ‘machinery’ is needed. If perfect solids are to be studied, an ability to apply periodic boundary conditions is essential in order to avoid the need to represent an infinite number of atoms in the calculation. Functional forms used for the expansion of the orbitals need to be chosen and a method to reduce the number of electrons can be useful. Two common approaches are used as exemplars and described in some detail. The first, which dominates in condensed matter physics, is the
"plane wave and pseudopotential" approach, the second uses Gaussian functions for the basis set and is, perhaps, the most common approach in quantum chemistry. Use is made of both of these methods later in the thesis.

3.1.3.1 Plane waves and pseudopotentials

DFT became a common method following the implementation of the first principles simulation approach described in the seminal paper of Car and Parrinello (1985). The five major features which characterise the approach are (Segall et al., 2002):

1. Use of a plane-wave basis set to represent the Kohn-Sham orbitals.
2. The replacement of the nuclei and core electrons with pseudopotentials.
3. Using fast Fourier transforms to take advantage of optimisations available by carrying out some of the calculation in real space and some in reciprocal space.
4. Minimisation of the total energy to find the ground state.
5. Fictitious dynamics for the electrons to allow propagation of the wavefunction in time to avoid repeated expensive energy minimisation.

Not all of these techniques are always used. A plane wave basis set makes it essential to describe the simulated system using 3-dimensional periodic boundary conditions, which means that the system is infinite. This apparent problem is solved by using Bloch’s theorem, where a representation of the wavefunction in reciprocal space allows a single unit cell to be studied, the cost is a need to sample the one electron wavefunctions over the first Brillouin zone. The \( n \) one-electron Kohn-Sham wavefunctions that are expressed as plane waves using Bloch’s theorem have the form:

\[
\psi_{n,k}(\mathbf{r}) = \sum_{G} u_{n,k}(G) \exp\left(i(k + G) \cdot \mathbf{r}\right),
\]

where \( G \) are wave vectors of the plane waves and \( k \) is the sampling point in reciprocal space and \( u \) are the coefficients of the plane wave expansion. Both the set of wave vectors and the sampling of the first Brillouin zone should be, in principle, infinite. In practice, this is not possible. Sampling a restricted number of ‘k-points’ is found to be a reasonable approximation; more k-points are needed as the band structure becomes more rapidly varying or with
decreasing unit cell size. Choosing a number of wavevectors to include involves choosing when to stop adding extra ones, each one must be commensurate with the unit cell so the set of plane waves starts with a wave with wavelength equal to a cell length, then a half, the a quarter and so on. In practice the set is described by a cut off energy, equal to the energy of a free electron whose wavevector has the largest wavevector in the plane-wave basis.

Computationally unfeasible cut off energies are required to describe the wavefunction close to the cores of the atom where the curvature of the wavefunction is largest. To avoid this pseudopotentials are introduced to replace the atomic cores (nuclei and inner electrons). These are constructed by fitting to all electron calculations for single, isolated atoms or ions using the same methodology that will be used for the extended calculation. The pseudopotential replaces the atomic core with a smoothed potential with lower curvature – which reduces the number of plane-waves included in the basis. Outside the core the pseudopotential wavefunction and potential matches the all electron one and is usually norm-conserving, that is the pseudowavefunction gives the correct number of electrons within the core. The fact that this restriction can be lifted was realised by Vanderbilt (Vanderbilt, 1990) who suggested the use of ultra soft pseudopotentials, which require the use of still smaller numbers of plane-waves.

3.1.3.2 Atomic centred basis sets

A common alternative to using plane waves for the basis set is to use a linear combination of atomic orbitals (LCAO) basis set. In this approach, the basis is built up from the electron orbitals present in isolated atoms making up the system being studied. A brief outline of the differences between LCAO and plane wave basis sets is outlined in table 3.1.
Table 3.1: Major differences between atom-centred and plane wave basis functions. Modified from Postnikov (2003). *Used without permission.*

<table>
<thead>
<tr>
<th>Plane Waves</th>
<th>Atom-centred functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete basis set possible by increasing a single parameter (the cut-off energy).</td>
<td>Size of basis set increases lineally with number of atoms - including vacuum does not increase basis.</td>
</tr>
<tr>
<td>Easy analytical manipulation, very useful for implementing new ideas (see Segall <em>et al.</em>, 2002).</td>
<td>Periodic or non-periodic boundary conditions possible.</td>
</tr>
<tr>
<td>Boundary conditions limited to periodic in three dimensions. Although ideal for studies of perfect crystalline solids, leads to problems for systems which are not inherently periodic.</td>
<td>No systematic method to improve quality of basis set. Convergence on complete basis set not guaranteed.</td>
</tr>
<tr>
<td>Rapidly changing fluctuations in charge near to atomic nuclei prevents use for all electron calculations.</td>
<td>Complex mathematical manipulation needed.</td>
</tr>
<tr>
<td>Cost of the calculation increases with increasing simulation box size even if the number of atoms is kept the same.</td>
<td></td>
</tr>
</tbody>
</table>

LCAO are defined in terms of atom-centred orbitals that are the product of an angular and radial part. The radial part can accurately be described using Slater-type functions, which match the exact analytical solution for the orbital in the hydrogen ion. However, taking integrals of Slater functions is a computationally expensive procedure, so Gaussian functions are commonly used instead. In this case the basis functions are:

\[
\psi_r = \sum_p d_p \left( cx^n y^m z^l e^{-ar^2} \right). \tag{3.12}
\]

where \(a\) is a parameter controlling the radius of the basis function \(n, m\) and \(l\) control the shape of the function, \(c\) is a normalisation constant and \(d\) is a fixed constant for the basis set. This scheme leads to the possibility of fast analytic integrals but more functions are required to describe the kato cusp close to \(r=0\).
There is a de-facto standard nomenclature for Gaussian type basis sets (Foresman and Frisch, 1996), ranging from the minimal STO-3G basis with just three Gaussian functions approximating a Slater function for each orbital through split valence basis sets such as 6-31G to basis sets with added functions to account for polarization (6-31G*) and with added diffuse functions with large $a$ parameters (6-31+G). Ultimately, very large basis sets are available, but there is no general method to systematically improve the basis (unlike the case of a plane-wave basis).

3.1.3.3 Computer codes and other approaches

The plane wave and pseudopotential method has been implemented using DFT in a wide range of computer codes including VASP and various codes called CASTEP (Segall et al., 2002). There are of course many other approaches to electronic structure calculations, here I mention a couple of them. The Projector Augmented Plane Wave method (implemented in VASP) is a close cousin of the plane wave and pseudopotential approach, but instead of replacing the core with a pseudopotential, projector functions are used. This can be more accurate than the pseudopotential method because problems with transferability are not encountered. To treat all electrons in a plane wave type code the basis set can be separated into two parts, valence electrons are described using plane waves and an LCAO basis set is used to describe the core. This approach, the Linearly Augmented Plane Wave method, offers very high accuracy but at the cost of extreme computational complexity, this scheme is implemented in the Wien2k code.

Codes that make use of an LCAO basis set can be divided between those that use Gaussian functions (see section 3.1.2.2) and those that make use of a numerical basis set. Gaussian type orbitals are used in the Gaussian code (Frisch et al., 1998) for studies of isolated molecules (the newest version, Gaussian03, includes 1, 2 and 3 dimensional periodicity too) and the CRYSTAL code, used for studies of periodic systems. The use of numerical basis sets tends to require fewer functions per orbital but the computational cost per function is increased. The numerical approach has an advantage that the functions can be made to go strictly to zero beyond some critical radius, this can be used to create a method which scales linearly with system size allowing very large systems to
be studied. One example of this approach is the SIESTA code (Sánchez-Portal et al., 1997; Artacho et al., 1999; Soler et al., 2002; Craig et al., 2004).

3.2 Parameterised potential models

The size and speed of modern computers limits the size of electronic structure calculations using DFT or Hartree-Fock theory to less than about 500 atoms. One approach to study larger systems is to parameterise the Hamiltonian; this leads to tight binding theory (Goringe et al., 1997). Although this method increases the number of atoms that can be studied and keeps a useful description of the electronic structure it is difficult to apply to ionic systems because it is based on a description of the electronic structure consisting of neutral atoms. An alternative to the complexities of attempting to solve the Schrödinger equation is to describe the total energy of the simulation cell by parameterised functions that depend on the relative atomic positions. This is an enormous simplification when it comes to calculating the potential energy hypersurface, which leads to drastic savings in computer time; many more atoms can be included in the calculation. The approach has a long history of successful application to minerals and other ionic materials (e.g. Catlow and Price, 1990; Patel et al., 1990; Price and Vocadlo, 1996; Wright, 2003) but suffers from some drawbacks, principal amongst these are that processes involving bond breaking are difficult to describe, potentials are tied to particular crystal structures and the model may break down on extrapolation away from the conditions under which it was derived.

The basic idea can be illustrated by considering a pair of interacting atoms or ions (Walton, 1983). At very large separations, the two atoms do not exert forces on one another but as they are brought together, they begin to interact. The forces they apply on another varies with their separation, at very short separations interactions between the two electron clouds will tend to force the atoms apart, at larger separations the electron clouds will tend to be polarised

---

10 The ability of a potential derived to describe a particular crystal to describe a different one is termed transferability. When fitting potential parameters there is often a trade off between the quality of the reproduction of the properties of the crystal being fitted to and the transferability of the resulting potential.
and van der Waals forces will tend to bring the atoms back together. Coulombic forces will be superimposed on top of this picture; if the two atoms have like charge there will be an additional repulsive force but if they are of unlike charge there will be an attractive force. At some separation, the forces will be balanced and the atoms will be separated by their equilibrium bond length. A schematic representation of the force as a function of separation is shown in figure 3.1. If we consider the potential energy of the pair of infinitely separated atoms to be zero then the potential energy of the atoms \( U \) at some smaller separation can be found by calculating the work done against the interatomic forces \( F \) in bringing the atoms together from infinite separation:

\[
U(r) = \int F(r) dr
\]  

(3.13)

and the equilibrium separation found by searching for a minimum on the energy versus separation graph (figure 3.1). Alternatively, the force can be found as the derivative of the potential energy with separation:

\[
F(r) = \frac{dU}{dr}
\]  

(3.14)

Figure 3.1: Schematic potential energy between two atoms as a function of separation. The harmonic approximation represented by the thin line gives a good approximation of the potential energy curve (the thick line) in the region of the energy minimum. (Based on figure 3.13 of Poirier, 2000).
Of course, a solid is made up of more than two atoms. Bringing a third atom into the system means that we cannot just account for the separations of each pair of atoms, but also the relative position of all three atoms. In a larger system the relative positions of all atoms are important. We can expand the total lattice energy (the energy required to bring all the atoms together from infinite separation) as a Taylor expansion:

$$U_{\text{latice}} = \sum_{i}^{N} U_{i} + \frac{1}{2} \sum_{i}^{N} \sum_{j}^{N} U_{ij} + \frac{1}{6} \sum_{i}^{N} \sum_{j}^{N} \sum_{k}^{N} U_{ijk} + \ldots$$ \hspace{1cm} (3.15)

The first term is the sum of the self-energies of all the atoms, the second the pairwise interaction between all pairs of atoms and so on, the fractions prevent over counting of interactions. For ionic systems this expression can be written more conveniently as:

$$U_{\text{latice}} = \frac{1}{2} \sum_{ij} \varphi_{ij} + \frac{1}{6} \sum_{ijk} \phi_{ijk} + \ldots$$ \hspace{1cm} (3.16)

where the self energy of an isolated ion has been defined as zero, the summations expressed in a truncated form and the Greek letters refer explicitly to parameterised functions of the energy. The pairwise term has been split into two parts, the first describing the Coulombic interaction between pairs of charged ions, accounting for \(\sim 90\%\) of the cohesion in ionic solids (Ashcroft and Mermin, 1976), and a ‘short ranged’ part which describes the other interactions.

As well as representing interactions derived from distinct phenomena, the separation is made for computational reasons: the short range part can be summed to convergence only accounting for interactions between atoms separated by less than some finite distance, while special techniques must be used for the conditionally convergent Coulomb sum.

### 3.2.1 Coulomb summation

The first term in equation 3.16 describes the energy of interaction between point charges. The difficulty in handling this term is due to the slow rate of decay in the interaction energy with the separation of the point charges, in finite systems the problem is easily overcome by including all pairs of point charges in the sum but when periodic boundary conditions are used, this becomes impossible.

In an infinite (or periodic) solid the evaluation of this term is difficult because
the strength of the interaction decays as $1/r$ while the number of particles to be included in the summation increases as $r^3$. Physically this means that atoms on the surface of any truncation sphere will play an important role in determining the final value of the sum and it can yield different values depending upon the order in which the terms are taken; the sum is conditionally convergent (see, Ashcroft and Mermin, 1976). To overcome this problem special Coulomb summation techniques have been developed which allow the sum to be evaluated. The most commonly used method is the Ewald summation (1921); this involves a mathematical trick that exploits the periodicity of the simulation cell. The sum is separated into two parts and summing one in real space and the second in reciprocal space results in two convergent sums. Physically this approach is akin to adding a shielding potential to distant ions summing interactions of the close ions and then correcting for the shielding (figure 3.2a).

Figure 3.2: Graphical representation of the Ewald summation (part a, from Leach, 2001) and the Wool summation (part b, from Wolf et al., 1999). The Ewald sum consists of the summation of point charges and a shielding potential (upper two segments) in real space combined with a sum to remove the shielding potential undertaken in reciprocal space (lower segment). In the Wolf sum each atom, $i$, interacting with a second atom, $j$, within a truncation sphere of radius $R_c$ has a compensating charge added at the edge of the truncation sphere. Figures reproduced without permission.
The Ewald method is difficult to parallelise to large multi-processor computers because of the requirement that the location of all the atoms in the simulation cell must be known to perform the long ranged part of the sum. In order to overcome this problem a number of Coulomb summation methods have been developed that scale more efficiently with system size and can be more easily parallelised. These methods include the fast multipole method (Greengard and Rokhlin, 1989) and smoothed particle-mesh Ewald approach (Essmann et al., 1995). Recently Wolf et al. (1999) devised a method which allows the summation to be carried out in real space using a finite radius truncation sphere. The key to this approach is to ensure that the truncation sphere carries no net charge, a condition enforced by the addition of phantom charges around the perimeter of the sphere (figure 3.2b). Convergence is further enhanced by the application of a shielding potential, chosen to be identical to the real space part of the Ewald sum. The shielding potential negates the need for additional corrections for multipoles across the truncation sphere. The Wolf sum is therefore controlled by two parameters: one controlling the shielding and a second giving the radius of the truncation sphere.

A similar problem is encountered in systems that are periodic in two dimensions and finite in the third. Now, once $r$ approaches the height of the system, the number of ions increases as $r^2$ again resulting in a non-convergent summation. This problem has been overcome by Parry (1975; 1976) who developed an approach similar to the Ewald summation for systems periodic in two dimensions. The one dimensional case (periodic in one dimension and finite in the other two) is absolutely convergent so, in principle, no special methods are needed, but the method developed by Saunders et al. (1994) speeds the summation. One-dimensional simulation cells are used in chapters 6 and 7 to model dislocations, both the Saunders approach and the Wolf sum (which functions identically in zero, one, two and three dimensional cases) have been used.

### 3.2.2 Pair-potentials

In the simple ionic model outlined above, the Coulomb forces are complemented by short-range repulsion from the overlap of electrons and attractive van der Waals forces. As the energy contribution from these terms
decays more rapidly than $r^2$, their summation is convergent in real space and it is usual to set the contribution to zero at a separation beyond a predetermined radius. This procedure may introduce discontinuities in the potential energy surface so sometimes the potential is tapered to zero (Gale and Rohl, 2003) or shifted in order to remove any discontinuity. Short-ranged two body interactions are also of use to describe covalent chemical bonds; in this case, the interactions usually exclude any Coulombic terms between bonded atoms.

The simplest form of short-ranged two-body interaction is a harmonic potential:

$$\varphi(r_{ij}) = \frac{1}{2} k_{ij} (r_{ij} - r_0)^2, \quad (3.17)$$

where the force on the atom, controlled by the parameter $k$, is proportional to its displacement from an equilibrium separation ($r_0$). This form is mostly used to describe chemical bonds in force fields derived for organic chemistry and in the shell model used to describe the polarisability of ions described in section 3.2.4. A more accurate picture of a chemical bond is provided by the Morse potential form:

$$\varphi(r_{ij}) = D_{ij} \left[ \left( 1 - e^{-a_{ij}(r_{ij} - r_0)} \right)^2 - 1 \right], \quad (3.18)$$

which has the advantage of describing the anharmonic nature of the bond (extension is easier than compression). The parameters of the Morse potential have clear chemical meanings: $D$ is the dissociation energy of the bond (the depth of the potential energy well), $r_0$ is the equilibrium bond length and $a$ describes the shape of the minimum.

Interactions between atoms that are not chemically bonded include terms to describe Pauli repulsion caused by the overlap of electron clouds and the correlated motion of electrons that generate Van de Waals attraction. The Lennard-Jones potential form:

$$\varphi(r_{ij}) = -\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^9}, \quad (3.19)$$

achieves this with four terms, the $A$ parameter controls the strength of the attraction which decays with the exponent $n$ while repulsion is controlled by $B$ and the exponent $m$; in order to have an energy minimum $m$ must be larger than
n. In many cases, the dominant contribution to the attractive interaction comes from the mutual interaction of fluctuating dipoles, quadrupoles and so on (van der Waals forces, see Jensen, 1999), which result in an attractive interaction:

\[ \varphi_{\text{vdW}}(r_{ij}) = \frac{c_{ij}}{r_{ij}^6} + \frac{b_{ij}}{r_{ij}^4} + \frac{e_{ij}}{r_{ij}^2} + \ldots, \]

and only taking the first terms gives the common choice of \( n = 6 \) in the Lennard-Jones potential. Setting \( m = 12 \) then leads to great computational efficiency using reduced units (Allen and Tildesley, 1986).

A more realistic description of the short-range repulsion term is as an exponentially decaying function provided by the Born-Mayer potential:

\[ \varphi(r_{ij}) = A_{ij} e^{-\rho_{ij}}, \]

with the repulsion controlled by two parameters, \( A \) and \( \rho \). The exponential decay is justified by the exponential long-range asymptotic behaviour of the electron density of atoms (see, Oganov, 2002). Adding van de Waals attraction gives the Buckingham potential form:

\[ \varphi(r_{ij}) = A_{ij} e^{-\rho_{ij}} - \frac{C_{ij}}{r_{ij}^6}, \]

which is by far the most commonly used function for the simulation of ionic solids and adds a single parameter, \( C \), to the Born-Mayer form. Many more two body interactions have been proposed, a wide range are tabulated by Gale and Rohl (2003).

### 3.2.3 Three and four body interactions

For some systems, useful results can be extracted from truncating the expansion of equation 3.17 at two-body terms, but in other systems higher order terms are needed. The typical use of a three body potential is to describe the electrostatic repulsion between bonds in a covalent system to give, for example, carbon in methane its correct tetrahedral coordination. The three-body interaction has also been successfully used in silicates to describe the silicate tetrahedron. The simplest form is to impose a restoring force proportional to the deviation of the bond angle from a specified ideal angle. This gives a harmonic potential form:
where $\theta$ is the angle made between three atoms, $\theta_0$ is the ideal angle and $k$ is a force constant. Other potential forms are described by, for example, Gale and Rohl (2003) and include the exponential three body interaction and the Axilrod-Teller interaction (Axilrod and Teller, 1943), designed to represent three centre dispersion effects (a three atom analogue to van der Waals forces). Four body interactions have two main uses, to impose the correct torsional angles in organic molecules such as ethane and to keep molecular ions, such as CO$_3^{2-}$, planar.

### 3.2.4 Polarisation

An additional important contribution to the energy of crystalline systems arises from the polarisation of their constituent ions. The inclusion this effect by one of several routes leads to a much better description of the behaviour of the solid. In the simplest approximation, the point charge representing the ion is supplemented by a point dipole, and possibly higher order terms, and many “covalent” effects can be captured (Madden and Wilson, 1996). In this model the dipole moment is given by:

$$\mu = \alpha V_f,$$  \hspace{1cm} (3.24)

and the ion contributes an additional self-energy term to the lattice energy given by

$$U_{\text{Polarization}} = -\frac{1}{2} \alpha V_f^2,$$ \hspace{1cm} (3.25)

where $\alpha$ is the polarisability and $V_f$ is the electric field. Although this is a simple model it adds a level of complexity to the calculation of the system’s energy because the interactions between induced dipoles must be handled in a self consistent manner – simply adding each contribution will not lead to the correct result as one dipole will change the value of the dipole on the surrounding atoms.

An alternative to this approach is the shell model first proposed by Dick and Overhauser (1958). This model keeps the simplicity of point monopoles for all of the electrostatic interactions but replaces the point charge representing the ion.
with two charges, one to represent the nuclei and a second to represent the surrounding cloud of electrons. As the two point charges move apart, a dipole is generated. Coulomb interactions between the core and shell charge representing a single ion are removed and the two particles are coupled by a short ranged potential function, usually a harmonic function (equation 3.17). Short range interactions between the polarisable ion and other ions in the system generally operate through the shells rather than the cores however some complex models (e.g. the apatite model of Mkhonto and de Leeuw, 2002) include short range interactions between cores of polarisable ions. Coulomb terms between all cores and shells (i.e. core-core, shell-shell and core-shell) representing different ions are included in the evaluation of the lattice energy.

\[ \alpha = \frac{q_{\text{shell}}}{k_{\text{core-shell}}} \]  

(3.26)

where \( q_{\text{shell}} \) is the charge on the shell and \( k_{\text{core-shell}} \) is the force constant is the force constant between the core and the shell.

Potential models based only on the shell model and pair wise interactions between ions have been used with considerable success for modelling ionic and
semi-ionic materials. However, they suffer from one major problem when they are used to model structures with the rock salt structure (e.g. NaCl, MgO and AgBr); they are forced to obey the Cauchy relations despite experimental knowledge that the materials violate these relationships. This problem is overcome by a further refinement – the breathing shell model (Schröder, 1966; Sangster, 1973). In this model (figure 3.3b), the point charge representing the shell is replaced by a sphere of variable radius. The shell charge is still located at the centre of the sphere but the short-range potentials act on its surface so the effective radius for the pair potentials is reduced by the radius of the shell. The radius of the shell becomes a variable in the calculation and changes in the radius are coupled to the energy by a restraining potential, which is usually harmonic but can also be an exponential function (Matsui et al., 2000), or, in principle, any other potential form. The inclusion of such off centred forces has the important consequence of allowing for the Cauchy violation in rock salt structured materials, discussed further in section 6.3.

3.3 Energy minimisation and parameter fitting

At absolute zero, the stable configuration of the atoms is that with the minimum energy. Having outlined the methods that allow the energy of a particular configuration of atoms to be calculated, the problem becomes one of arranging the atoms such as to minimise the energy. The difficulty is that the relationship between the energy and the atomic configuration is complex, and there are far too many possible configurations to calculate the energy of them all to find the minimum. This type of mathematical problem occurs in many places (e.g. in modelling financial markets and in computer games) and there are a wide range of techniques to solve it. There are two approaches to the problem, a search for the global minimum of the system or for local minima. The global minimum is the lowest possible energy that can be found; techniques that attempt to find this include simulated annealing (Kirkpatrick et al., 1983) and the use of genetic algorithms (recently reviewed by Woodley, 2004). Usually the global minimum is not the desired structure, for example performing a global minimisation for any of the SiO₂ polymorphs should result in the structure of alpha-quartz and the diverse crystal chemistry of the dense polymorphs and the siliceous zeolites
would be missed. Instead, if the starting configuration of the system is close to the local minimum corresponding to the system of interest then one of the local optimisation techniques is more appropriate.

### 3.3.1 Local optimisation algorithms

In one way or another local optimisation of multi-dimensional functions is a major part of most modelling codes; as well as using optimisation algorithms to move the atoms and find stable structures, optimisation procedures are used for the convergence of electron density in DFT codes and to obtain the parameters of the potential functions used in parameterised potential models (see section 3.3.2). In the case of finding the stable geometry of a collection of atoms the optimisation amounts to the problem of finding values for \( r \) (which, in the case of fixed cell parameters with \( N \) ions is a \( 3N-3 \) dimensional vector) that give a minimum total energy, \( E(r) \), but we do not have a global picture of the shape of \( E(r) \). If care has been taken in the derivation of the potential model then \( E(r) \) is continuous. At a minimum the gradient of the potential energy surface, \( \nabla E(r) \), will be zero and the matrix of second derivatives, \( A \), will be positive. Close to the minimum the potential function is well described by a harmonic energy well (c.f. figure 3.2).

There are many ways to search for local minima on the potential energy surface (for some common examples see Press et al., 1992), the most commonly used approach is based on the Newton-Raphson method. To second order, the energy at some point \( r \) a short distance from the current point, \( r_i \), which is close to the minimum, is given by:

\[
E(r) = E(r_i) + (r - r_i) \cdot \nabla E(r_i) + \frac{1}{2}(r - r_i) \cdot A \cdot (r - r_i),
\]

so the gradient at \( r \) is given by:

\[
\nabla E(r) = \nabla E(r_i) + A \cdot (r - r_i).
\]

If \( r \) is to be a minimum then we set \( \nabla E(r) \) to zero and the step to take to the minimum is just \( r - r_i \), in other words the step is:

\[
r - r_i = -A^{-1} \cdot \nabla E(r_i).
\]
If the function is not strictly harmonic at \( r \), then using the Newton-Raphson method iteratively should lead to a minimum, but this approach is hardly ever used and instead quasi-Newtonian algorithms are implemented.

In the quasi-Newtonian methods the second derivative matrix is not explicitly calculated, instead an approximation to the inverse of the second derivative matrix, the Hessian, \( H \), is constructed and updated in each iteration through the Newton-Raphson algorithm so that it becomes equal to the exact inverse second derivative matrix when the function is close to a minimum. This approach is taken for a couple of reasons: (1) Calculating the matrix of second derivatives can be extremely time consuming as it requires at least \( 3N^2 \) energy evaluations for an N dimensional system. (2) For the Newton-Raphson method to lead to a decrease in energy \( A \) must be positive definite, away from the minimum there is no guarantee that \( A \) is positive definite but it is always possible to choose the starting approximation for the quasi-Newton method to be so (the unit matrix is a common choice). A second issue is that the Newton-Raphson step (found in equation 3.29) may be too long when evaluated away from the quadratic region of the potential energy surface, in the quasi-Newtonian method a fraction of the step is taken determined by a line search in the direction of the step. There are two common approaches to updating the Hessian during each iteration, the most widely used the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method (Shanno, 1970) where the updated Hessian is given by\(^{11}\):

\[
H_{i+1} = H_i + \frac{(r_{i+1} - r_i) \otimes (r_{i+1} - r_i)}{(r_{i+1} - r_i) \cdot (\nabla E(r_{i+1}) - \nabla E(r_i))} - \frac{\left[H \cdot (\nabla E(r_{i+1}) - \nabla E(r_i))\right] \otimes \left[H \cdot (\nabla E(r_{i+1}) - \nabla E(r_i))\right]}{(\nabla E(r_{i+1}) - \nabla E(r_i)) \cdot H \cdot (\nabla E(r_{i+1}) - \nabla E(r_i))}
+ \left[(\nabla E(r_{i+1}) - \nabla E(r_i)) \cdot H \cdot (\nabla E(r_{i+1}) - \nabla E(r_i))\right] u \otimes u
\]

(3.30)

where the vector \( u \) is defined:

\[
u = \frac{(r_{i+1} - r_i)}{(r_{i+1} - r_i) \cdot (\nabla E(r_{i+1}) - \nabla E(r_i))} - \frac{H_i \cdot (\nabla E(r_{i+1}) - \nabla E(r_i))}{(\nabla E(r_{i+1}) - \nabla E(r_i)) \cdot H_i \cdot (\nabla E(r_{i+1}) - \nabla E(r_i))}.
\]

\(^{11}\otimes \) is the direct or outer product of two vectors, the \( ij \) component of \( u \otimes v \) is \( u_i v_j \).
The alternative method of Davidson, Fletcher and Powell (Fletcher and Powell, 1964) is identical to the BFGS approach but the final term of \( u \) is neglected.

Storing the Hessian matrix can be impossible for large systems (more than a few thousand atoms) as the storage requirement increases with the square of the number of atoms. In such situations, the conjugate gradients method, with storage requirements of order \( N \), can be used at the expense of many more steps in the optimisation (Press et al., 1992). This approach utilises a series of line searches over the energy surface in order to find the minimum. The first step is down the steepest gradient from the starting point and subsequent steps are downhill but perpendicular to the previous step.

An elegant half way house between the conjugate gradients and BFGS algorithm is the “limited memory BFGS” recently implemented in the GULP code (J. D. Gale, Personal Communication). In this approach only a limited set of diagonal elements of the Hessian are evaluated and stored, hopefully reducing the number of cycles compared to the conjugate gradients approach without running out of memory! On the other hand if evaluation of the energy is computationally expensive it is worth minimising the number of steps, this can be achieved by having a good guess for the starting Hessian. The new CASTEP code achieves this by building an initial Hessian from the approximate average vibrational frequencies and bulk modulus of the system (Segall et al., 2002). Finally the rational function optimisation technique (RFO, Banerjee et al., 1985) is interesting. In this method the exact Hessian is calculated and diagonalised, the search direction is then along one of the eigenvectors. This approach is particularly useful if a stationary point on the potential energy surface, which is not a minimum (or maximum), is sought or if the quasi-Newton methods fail to converge to a minimum.

### 3.3.2 Derivation of parameters

Optimisation algorithms are also used in the derivation of the parameters of particular potential models. Rather than moving the atoms in order to minimise the energy the atoms are held fixed and the parameters varied so that the model reproduces some pre-selected energies or properties (see section 2.4) for the structure called the observables. This selection of properties can be drawn from
experiment or other calculations using electronic structure methods and more than one structure can be used in the fitting procedure. The quantity to be minimised in fitting is the sum of the squares, $F$, of the disagreement between the calculated and supplied observables defined by:

$$F = \sum_{i=1}^{N_{\text{obs}}} w_i (f_{i}^{\text{obs}} - f_{i}^{\text{calc}}),$$

(3.31)

where $N_{\text{obs}}$ is the number of observables, $f_{i}^{\text{obs}}$ and $f_{i}^{\text{calc}}$ are the values of each observed and calculated observable and $w_i$ is a weighting factor for each observable. In principle, the parameters to be fitted to minimise $F$ include all of the parameters in the potential model, including the charges on each of the ions and the partitioning of the charge between the cores and shells, if the shell model is used. Often the set of observables is not sufficient to adequately constrain all of these parameters and formal charges are assumed and some short range parameters can be transferred from existing models of related systems. The only complications that arise are handling the location of any shells in the model and determining the exact way the calculated observables are calculated. An easy approach is to calculate the observed properties of the force field at some predefined (usually experimental) structure, varying the parameters of the force field which vary the calculated observables and a good fit is sought. The structure is included in the fit by including the gradients of the potential energy surface in the fit (and setting the observed gradients to zero). In this scheme, the position of the shells cannot be easily determined – no experimental data helps in locating these fictional particles. The solution is so called simultaneous fitting where the location of the shells are determined by an energy minimisation at each fitting step, or equivalently zero force on the shells is required in the fit and the positions of the shells become unknown parameters to be found by the fit. The problem with this approach is that the properties are calculated away from an energy minimum (which is not strictly correct). This is overcome by “relax fitting” (Gale, 1996), where, at each fitting step, the system is subject to energy minimisation prior to calculating the observables. The structure is then directly included as a set of observables and there is no ambiguity in determining the position of the shells. The procedure is obviously more computationally time consuming but it usually results in a better fit; of

70
course a good approximate model is needed to start relax fitting otherwise the initial energy minimisation will fail.

### 3.4 Calculation of properties

Once the equilibrium geometry has been found various properties of the system can be extracted from the curvature of the potential energy surface at the minimum; these properties include the elastic constants (and compliance) tensor, the dielectric and piezoelectric constants, refractive indices, electrostatic potentials and the vibrational properties. This section concentrates on the calculation of two properties of particular geophysical significance, the elastic behaviour of the solid that can be used to calculate seismic velocities and calculation of vibrational properties, which provide a way to probe the effect of temperature on a solid – obviously essential to understand mantle mineralogy – and to make contact with experiments that probe the parts of the vibrational spectra such as IR spectroscopy. The calculation of the dielectric tensor is also briefly considered because it is important when it comes to studying defects. The calculation of other properties have been considered by, for example, Gale and Rohl (2003) and Catlow and Mackrodt (1982). The calculation of other properties, such as NMR shielding tensors, require an explicit knowledge of the electronic structure of the system so can only be calculated from the results of such calculations.

#### 3.4.1 Elastic properties

The elastic constants define the relationship between stress and strain. Only Hooke’s Law (the linear approximation where stress is proportional to strain) will be considered here. In three dimensions the second order stress tensor ($\sigma_{ij}$) is related to the second order strain tensor ($\varepsilon_{ij}$) by the fourth order elastic constants tensor ($c_{ijkl}$):

$$\sigma_{ij} = \sum_{kl} c_{ijkl} \varepsilon_{kl}$$  \hspace{1cm} (3.32)

Because of translational symmetry there are a maximum of 21 independent elastic constants, so the tensor can be represented in a reduced form as a 6×6 diagonal matrix. The point group symmetry may further reduce the number of
independent elastic constants. For example cubic crystals only have three independent constants and orthorhombic crystals have nine (see Nye, 1957). The elastic constants tensor can be re-expressed in terms of the potential energy change associated with a change in the strain:

$$C_{ij} = \frac{1}{V} \left( \frac{\partial^2 U}{\partial \varepsilon_i \partial \varepsilon_j} \right), \quad (3.33)$$

(Nye, 1957), Voigt notation is used to reduce the number of subscripts (see Poirier, 2000). It is possible to calculate the elastic constants from this expression by calculating the energy of the cell at different strains and fitting this data to equation 3.33. This method is generally used in electronic structure calculations, where the second derivative matrix is not readily available. In this process it is essential to allow the internal degrees of freedom to relax at each applied strain, so each strain involves a geometry optimisation with fixed cell parameters. For parameterised potential methods the second derivatives matrix is available, so an alternative approach is taken. Defining three matrices for the second derivatives of energy with the degrees of freedom of atomic co-ordinates and cell lattice vectors:

$$D_{ee} = \left( \frac{\partial^2 U}{\partial \varepsilon_i \partial \varepsilon_j} \right)_{\text{internal}},$$

$$D_{ei} = \left( \frac{\partial^2 U}{\partial \varepsilon_i \partial \alpha_j} \right)_{\varepsilon}, \quad (3.34)$$

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

Equation 3.33 can be expressed as:

$$c_{ij} = \frac{1}{V} \left( D_{ee} - D_{e} D_{ij} D_{e} \right). \quad (3.35)$$

In equations 3.33 – 3.35 $V$ is the volume of the unit cell, $U$ is the internal energy of the unit cell, $\alpha$ and $\beta$ are the positions of atoms in the cell $\varepsilon$ is the strain and $\sigma$ is the stress.

From the elastic constants matrix, the compliance matrix (the constant of proportionality between strain and stress) can be readily calculated (one is just the inverse of the other). More interestingly, this allows the calculation of the
elastic properties of rocks by assuming that they are isotropic polycrystalline samples made up of a random arrangement of anisotropic crystals and therefore the calculation of seismic wave speeds. Such data is invaluable to compare mineralogical models of the mantle with data from seismic observations. There is, however, no unique method of averaging the elastic constants to achieve this. The Reuss average assumes uniform stress through the polycrystalline material with the strain accommodated by all the grains in series while the Voigt average assumes the strain is uniform and is supported by individual grains in parallel. These give lower and upper bounds of the possible value the shear and bulk moduli could take. The arithmetical mean (Hill average) of these two values are often used (e.g. Poirier, 2000).

3.4.2 Dielectric properties

The second order dielectric constants tensor describes the polarisation of the crystal in an applied electrical field. The dielectric behaviour depends on the frequency of the applied field but here only the low frequency static limit is described (Gale and Rohl, 2003, provide details of the calculation of the high frequency and frequency dependent tensors). High or low values of the dielectric constants can be a useful property in technological applications but, in the context of this thesis, their calculation is outlined due to the importance of the dielectric response of a crystal for the calculation of defect properties. From the second derivatives matrix \( (D) \), the cell volume \( (V) \) and a vector of ionic charges \( (q) \) the static dielectric constant \( (\varepsilon^0) \) is given by:

\[
\varepsilon^0_{\alpha\beta} = \delta_{\alpha\beta} + \frac{4\pi}{V} q D_{\alpha\beta}^{-1} q. \tag{3.36}
\]

3.4.3 Lattice vibrations

The atoms in a solid are far from stationary, at any temperature they vibrate about their equilibrium positions. It is possible, in a periodic solid, to describe these movements as wave-like oscillations of the atoms but quantum mechanics tells us that these vibrations must be quantised, so we can also describe them as phonons, quantum particles that describe the atomic motions.

The phonon frequencies are the square root of the eigenvalues of the dynamical matrix (see Cochran, 1973), which is constructed from the second derivative
matrix multiplied by the reduced masses of the atoms. For periodic solids a phase factor must be added and the frequencies become wave-vector dependent leading to phonon dispersion relations. The vibrational frequencies can be used for a number of tasks; firstly, imaginary numbers as eigenvalues of the dynamical matrix results in unphysical negative frequencies. These negative frequencies indicate that the model is not in an energy minimum and is instead at a turning point on the potential energy surface. Secondly, the frequencies can be compared to experiment. The allowed frequencies in experiment depend on the methodology used to measure them – the selection rules. Inelastic neutron scattering can probe the entire vibrational spectrum while more common Infra-Red or Raman spectroscopy only probe a limited number of frequencies. In any case, the observed frequency can be compared to the calculated one and the relative intensities of the response estimated (Gale, 2001). For Infra-Red spectroscopy the intensity can be calculated from:

$$I_{\text{IR}} \propto (\sum q d)^2,$$

(3.37)

with $q$ being the charge on the site and $d$ the displacement vector associated with the vibration frequency; the summation is over all sites in the crystal. Raman intensities are given by:

$$I_{\text{Raman}} = \left(1 + \frac{1}{\exp\left(\frac{hv}{k_BT}\right)}\right)^2 \chi^2,$$

(3.38a)

with $\chi$ defined:

$$\chi = \sum_i \sum_j (r_i d_i) (r_j r_{ij}).$$

(3.38b)

This method gives, for any structure, the set of frequencies associated with any wave vector. It is now possible to calculate the phonon frequencies over all allowed wave vectors to determine the phonon density of states. Using the density of states it is possible, using statistical mechanics, to find all thermodynamic properties of the structure under investigation. For example the free energy, $F$, is given by:
\[ F = kT \sum_{\nu} \left( \frac{\hbar \omega \nu}{kT} + \ln \left( 1 - e^{\hbar \omega \nu/(kT)} \right) \right), \]  

(3.39)

where \( k \) is Boltzmann's constant, \( h \) is the Planck constant, \( T \) is temperature, \( M \) is the total number of phonon frequencies and \( \omega \nu \) is each phonon frequency. The calculation of phonon frequencies is, however, computationally time-consuming but fortunately thermodynamic quantities converge rapidly as the number of wave vectors sampled increases.

### 3.5 Accounting for the effect of pressure and temperature.

At any temperature above zero Kelvin, the atoms of a real crystal vibrate vigorously about their mean positions. In modelling solids, this motion should be accounted for in order to account for phenomena such as thermal expansion and vibrational entropy but, so far, the discussion has only involved a model static lattice. Ideally, one would solve a time dependent Schrödinger type equation for the atomic positions; in general this is not possible so classical equations of motions are solved instead. The equations can be solved using two techniques that are complementary, lattice dynamics and molecular dynamics. Lattice dynamics treats the motions as a series of wave like displacements and provides a method of solving the equations analytically. Molecular dynamics solves the equations of motion by numeral integration over a large number of time steps.

Before discussing temperature, the effect of pressure in static simulations will be addressed. This is a relatively simple problem, first the cell parameters are allowed to vary in the energy minimisation and secondly rather than minimising the internal energy the enthalpy, \( H \):

\[ H = U + PV, \]  

(3.40)

is minimised (\( U \) is the internal energy, \( P \) is the pressure and \( V \) is the cell volume).

Free energy minimisation follows from a calculation of the phonon density of states during each step of the energy minimisation. This is used to calculate the free energy using equation 3.39 and this is minimised during the optimisation. For maximum efficiency, this is done using analytical derivatives (Gale, 1998)
although the use of numerical derivatives is also possible (Parker and Price, 1989), especially if additional approximations are made.

An alternative to finding analytic solutions to the Newtonian equations of motion, is to solve them numerically. This is the basis of molecular dynamics simulations. The atoms are assigned masses and initial velocities that are propagated through time, after a discrete time step, the total force on each atom is calculated and this is used to calculate the acceleration of the atom. The velocities of the atoms are updated and they again move for a time step. Details of integration algorithms and coupling to thermostats and barostats to perform constant pressure and temperature dynamics is beyond the scope of this thesis, the reader is referred to Allen and Tildesley (1986) for more information.

3.6 Modelling point defects

Turning away from a model of a perfect crystal to one containing defects introduces interesting problems. A key feature of any defect is that it breaks the long-ranged translational symmetry of the crystal, making any direct application of the methodologies described so far in this chapter difficult. The point defect population is usually quite small. This leads to the "dilute limit", a situation where the defects are considered to be isolated in an otherwise perfect infinite crystal. The defect energy is then defined as the difference in free energy between this defective case and a non-defective reference system. For the purposes of this section, only the defect internal energy will be considered. The defect energy (enthalpy at zero Kelvin, neglecting zero point energy contributions) is therefore defined by:

\[ E_{\text{defect}} = E_{\text{reference}} - E_{\text{defective}} \]  

(3.41)

where \( E_{\text{defect}} \) is the defect energy in the dilute limit and \( E_{\text{reference}} \) and \( E_{\text{defective}} \) are the internal energies of the infinite reference and defective systems respectively.

Defect – defect interactions can be considered as a modification to the situation in the dilute limit and taking the energy difference between a system with interacting defects and the sum of the isolated defect energies gives a binding enthalpy for the defects. It is worth noting at the outset that although binding enthalpies are almost always negative for defects of opposite charge (thus
favouring binding), there is also a large decrease in the entropy of the system when defects are forced together, which acts against binding. The defective system can be described by a large cluster of atoms surrounding a single defect or by the use of periodic boundary conditions applied to a “super-cell” consisting of a number of unit cells (figure 3.4). Cluster based methods clearly succeed in isolating each defect – there is only one defect represented in the simulation. However, some effects of the defect are quite long ranged, especially if the defect is charged, which can mean that too many (i.e. upwards of a few thousand) atoms must be included in the simulation. This problem can be solved using approximate methods to describe the response of the crystal away from the defect.

Figure 3.4: (a) Schematic representation of the Mott-Littleton approach to the modelling of defects. The inner region (I) is described using interatomic potentials and is made large enough to allow the displacement of atoms in the rest of the crystal (region II) to be small enough to be described by the harmonic approximation. This allows the energy stored in region II (which extends to infinity) to be evaluated by only considering a finite number of atoms, region IIa. (b) Representation of the super-cell method. The interactions between the infinite array of defects must be removed (see section 3.6.1).

3.6.1 The super cell approach

The most straightforward method to study defects is to impose artificial translational symmetry on the system. This allows any computer code for the simulation of bulk systems to be used to study defects; the problem then is that introducing a defect into the simulation cell causes images of the defect to be introduced into surrounding cells. Around each defect, the atoms are displaced.
from their equilibrium positions. The displacement fields around adjacent
defects may overlap causing the structure of the defect in the simulation cell to
be incorrect, additionally if the defect is charged, or carries some dipole, there is
a long ranged defect – defect interaction as well as problems calculating the
energy of a infinitely charged solid. These problems can generally be solved;
the simulation cell can consist of many unit cells in order to increase the defect
– defect separation and partially isolate the defect. Additionally corrections
exist which can be used to remove residual defect – defect interactions.

3.6.1.1 Charged defects in super cells

The most difficult case for the super cell method of studying point defects is
when the defect is charged. In this case, there are two problems. Firstly,
because of the long ranged nature of charge – charge interactions, it is
exceptionally inefficient to make the separation large enough to make the
defects non-interacting by simply enlarging the super-cell. The second
complication arises because the simulation cell is not charge neutral and
therefore the periodic system has infinite charge and the energy is difficult to
define.

For atomistic models, the charge neutrality problem can be solved by applying a
simple correction to the Ewald summation that corresponds to introducing a
uniform background charge to neutralise each simulation cell (Leslie and Gillan,
1985). The background correction is:

\[ E_{\text{Background}} = -\frac{1}{2} \pi Q^2 / V_0 \gamma' \]  

(3.42)

where \( Q \) is the charge on the simulation cell, \( V_0 \) is the volume of the simulation
cell and \( \gamma \) is the Ewald parameter. The background correction is simply added to
the Coulomb energy calculated in the usual way (section 3.2.1).
The charge – charge interaction between defects remains but can be removed by
the application of a second correction described by Leslie and Gillan (1985).
The charge – charge interactions in the model crystal are approximated as
charge – charge interactions in a dielectric continuum, the dielectric constant of
which is that calculated for the perfect crystal (section 2.4.2). The interaction
energy for a cubic lattice is then given by:

\[ E_{\text{interaction}} = -\frac{1}{2} \alpha Q^2 / \varepsilon_0 L , \]  

(3.43)

where \( \alpha \) is the Madelung constant, \( Q \) is the charge on the defect, \( \varepsilon_0 \) is the static
dielectric constant and \( L \) is the cell parameter. This interaction energy can also
easily be found by calculating the energy of a periodic array of ions in a vacuum
and dividing by the dielectric constant. For non-cubic systems, there is an
additional problem here. The dielectric behaviour is properly described by a
second order tensor (Nye, 1957), Brodholt (1997) suggests circumventing this
problem by using an average experimental value for the dielectric constant
whilst Woodley et al. (2003) used the trace of the calculated tensor. These
approximations seem sensible given the small magnitude of the correction term
but this could lead to errors if the simulation cell is small or the dielectric
anisotropy of the crystal is very large. It is also questionable that the bulk
dielectric behaviour, a macroscopic property, is appropriate for use on the small
length scale achieved by super-cell calculations.

Despite these corrections, for charged defects, a large super-cell is required for
defect energies to converge; this often makes the super-cell method more
computationally expensive then the Mott-Littleton method (section 3.6.2.1).
However, some useful techniques (e.g. lattice dynamics or density functional
theory with a plane wave basis) require a periodic model such as that provided
by the super cell method.

3.6.2 Cluster based methods

The alternative to imposing artificial periodicity on the defective crystal is to
find a way to model an isolated defect in an infinite crystal. This is not easy
because an infinite crystal has an infinite number of atoms. The solution is to
use an approximate continuum model to describe the crystal away from the
defect and only use an atomistic model where it is needed – close to the defect.
At its simplest, this approach involves making a finite cluster of the crystal surrounding the defect. Atoms close to the edge of the cluster are held fixed in order to prevent relaxation caused by the presence of vacuum around the cluster, while the atoms closer to the defect are relaxed to the energy minimum. The fixed atoms ensure that, from the point of view of the defect and its surrounding atoms, the crystal appears to be infinite. This approach involves a substantial approximation, especially in the case of charged defects, because the structure of the crystal away from the defect is altered by the presence of the defect, the more sophisticated methods described below account for this.

### 3.6.2.1 The generalised Mott–Littleton method

For charged defects in ionic materials, atoms distant from the defect are displaced away from their equilibrium lattice sites by long ranged electrostatic forces acting between the defect and the ions. This leads to problems for a simple fixed boundary cluster based approach because it neglects the effect of the polarization of the lattice away from the defect, which can feed back into the structure of the material close to the defect. An approximate method for treating this long ranged effect in ionic crystals was first proposed by Mott and Littleton (1938); they derived a formula for the energy cost of the polarization for a cubic lattice. With the development of computational algorithms, this idea was used as the basis of a very widely used methodology to model point defects in ionic materials (see, for example, Catlow and Mackrodt, 1982; Lidiard, 1989; Harding, 1990; Islam, 2000). Modern generalised implementations of the Mott-Littleton method are due to work carried out at Harwell during the 1970’s (Norgett, 1974) leading to the HADES, and later, CASCADE codes. With further refinements, this approach is implemented in the GULP code, which was used for the research presented in this thesis using the Mott-Littleton method. The major problem is to include the effect of an infinitely large crystal without having an infinite number of atoms, as this would be an intractable problem.

The approach taken is to divide the crystal into two regions, a spherical inner region (region I) surrounding the defect and an outer region (region II), which extends to infinity. Initially, assuming all interactions are short ranged, the total energy of the crystal is decomposed into three parts (equation 3.44 and figure 3.5), interactions between atoms in region I, $E_1$, interactions between atoms in
region II, $E_3$, and interactions between atoms in region I and region II, $E_2$. The energy $E_1$ is only a function of the positions of the atoms in region I, $r$, the energy of region II is a function of the positions of the atoms in region II, $\xi$, and the interaction energy between region I and II depends on both sets of coordinates. i.e.

$$E = E_1(r) + E_2(r,\xi) + E_3(\xi),$$

(3.44)

where $E$ is the energy of the defective or perfect system of equation 3.43 depending on whether a defect has been introduced or not.

By making region I large the displacements of the atoms in region II become small, this allows the energy associated with the displacements to be approximated as an harmonic function of the displacements:

$$E_3 = \frac{1}{2} \xi \cdot \mathbf{A} \cdot \xi,$$

(3.45)

where $\mathbf{A}$ is the second derivative matrix. Applying the condition that at equilibrium there is no energy gradient with respect to the co-ordinates of region II for any set of co-ordinates for the inner set of atoms:

$$\frac{\partial E(r,\xi)}{\partial \xi} = \frac{\partial E_2(r,\xi)}{\partial \xi} + \mathbf{A} \cdot \xi = 0,$$

(3.46)

and substituting this into equation 3.44 leads to an expression for the total energy that does not require the calculation of $E_3$ and is therefore tractable:

$$E = E_1(r) + E_2(r,\xi) - \frac{1}{2} \xi_0 \cdot \mathbf{A} \cdot \xi \bigg|_{\xi = \xi_0}.$$

(3.47)

By minimising this expression with respect to $r$, we arrive at a structure of the defective system. This procedure removes the need to calculate the self-energy of region II by approximating the energy cost as a harmonic function.

A further refinement is that the defect energy is calculated directly, rather than calculating the energy of the defective and non-defective systems separately and finding the difference. This leads to cancellations of terms; in the simple case where all interactions are two-body in origin and no shells are present the defect energy is given by:
Here $r$ describes the locations of the atoms in the defective system and $R$ their co-ordinates in the non-defective system. The summations are performed using a useful convention: it is assumed that the energy of the ion at infinity that is to be added to the perfect system (to form an interstitial) or that has been removed from the perfect system (to form a vacancy) is zero and the sum excludes terms that would lead to double counting (see Norgett, 1974). Impurities can be treated as a combined vacancy and interstitial on the same site. In order to solve equation 3.48, the co-ordinates of the atoms in the reference system must be stored and used in the calculation – adding substantially to the computational overhead.

The potential model actually consists of long ranged electrostatic contributions as well as the short-range contributions. To handle this region II is split into two sub-regions, region IIa extends from the edge of region I to a point where the short range interactions have decayed to zero (i.e. by a radius equal to the short range cut-off) and region IIb, which includes the rest of the crystal. The forces on the atoms in region IIa are calculated explicitly (although an additional approximation – that the forces only arise from the defect itself – is often used) given the position of the atoms in region I. The forces on atoms in region IIb must be calculated implicitly using a knowledge of the dielectric behaviour of the system. In region IIb only electrostatic interactions with region I are important, early work utilised an isotropic approximation to the dielectric tensor but in the more modern codes the correct anisotropic description (Catlow et al., 1982) is used. In this scheme the energy of region IIb is given by:

$$E_{\text{Defect}} = \sum_{i,j \in I, \ i \neq j} \left\{ \varphi(r_i, r_j) - \varphi(R_i, R_j) \right\} + \sum_{j \in \Pi} \left\{ \varphi(r_i, r_j) - \varphi(R_i, r_j) \right\}$$

$$- \frac{1}{2} \sum_{i \in I, \ j \in \Pi} \left[ \frac{\partial \varphi(r_i, r_j)}{\partial r_j} - \frac{\partial \varphi(r_i, r_j)}{\partial r_j} \right] \cdot (r_j - R_j).$$

(3.48)
with the matrix $M$ defined by:

\[ M_{ij}^{\alpha\beta} = \sum_{\gamma} \left( (D^{-1})^{\alpha\gamma} q \right) \left( \varepsilon^{-1} \right)^{\alpha\beta}. \]  

(3.49b)

3.6.2.2 Embedded cluster methodologies

Although the generalised Mott-Littleton method provides a good model of the isolated defect in an infinite crystal, the methodology is limited to the use of parameterised potentials; it is not possible to implement the model using electronic structure methods. This is because too many atoms must be included in region $I$, the correct method of describing region $II$ is also far from clear. Recent work on embedded methodologies provides a solution though. Many chemical reactions of interest occur by local cleavage and formation of bonds in an extended system. Binding of molecules to active sites in enzymes, heterogeneous catalysis and solvation of ions are all examples of local problems. The embedding approach describes the local reaction using a high quality electronic structure description and surrounds this by a lower quality representation of the surroundings. This scheme is clearly of relevance to studies of point defects in solids where an electronic structure description of the
defect (where bonds may be broken and the description should be good) can be embedded in a parameterised potential model of the surrounding crystal.

The use of embedding methods for the study of point defects in ionic materials has gone through a number of developmental stages. Early approaches had an electronic structure description of the defect (using Hartree-Fock theory) with the surrounding crystal either represented by fixed point charges (Grimes et al., 1989) or by a local confining potential calculated for a perfect crystal (Kunz and Vail, 1988). In either case, the outer embedding region is held fixed at either the bulk structure or a structure supplied before the calculation. A major advance is the coupling of the structure of the surrounding outer region to the quantum mechanical inner region, this was achieved by Shluger and co-workers (1986; 1993) with the *simultaneous* optimisation of the electronic structure and embedding regions implemented by Shluger and Gale (1996). This methodology has been implemented into the GUESS code (Sushko et al., 2000a; 2000b) and included as part of Chemshell (Sherwood et al., 2003), a general scripting environment for chemical simulations. Recent applications of this methodology to defects in forsterite using the GUESS code are described by Brathwaite et al. (2002; 2003); as this the approach used for the research presented in this thesis it is this methodology that will be presented in detail. Further discussion of the background to other embedding methodologies has been given by Sherwood et al. (2003).

The model used in this thesis consists of an inner region surrounding the defect consisting of up to about 100 atoms described using either density functional or Hartree-Fock theory (in principle, of course other methods could be used). This is surrounded by a region of ions described by a parameterised potential model that is allowed to relax during the geometry optimisation, surrounded in turn by fixed point charges needed to converge the electrostatic potential in the inner regions. Figure 3.6 shows this schematically.
The description of the quantum mechanical region is handled by the GAUSSIAN code (Frisch et al., 1998) with the outer regions represented by point charges, separated from the QM region by shell of cations represented by pseudopotentials. The GAUSSIAN code is used to calculate the energy of the QM cluster, and the forces on each QM ion, in the electrostatic potential of the rest of the system. The energy and forces are then fed back into the main GUESS code, which also calculates the energy of interactions between species in the environment, and short ranged potential functions that cross the QM/MM boundary. From the energy and forces, the atoms can be moved in order to minimise the total energy (section 3.3). The total energy of the system can therefore be expressed as:

\[
E = \langle \Phi | H_0 + V_{\text{env}} | \Phi \rangle + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{q_i q_j}{r_{ij}} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} q_i r_{ij}^2 + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} q_i r_{ij}^2,
\]

(Sushko et al., 2000b). The first term of 3.50 is the energy of the QM region in the Coulombic field of the rest of the model, given by:

\[
V_{\text{env}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{q_i q_j}{|r_i - R_j|} + \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{Z q_i q_j}{|R_i - R_j|},
\]

(q is the charge on an MM ion and Z the effective charge on a QM ion according to the potential model), the second term is the Coulombic contribution between ions in the outer regions (q is an MM charge and r is the
position of an MM ion), the third term is the short range pair-wise potential, the fourth term accounts for the polarization of ions in the outer regions represented by a harmonic shell model (spring constant $k$ and core – shell separation $r_{oc}$) and the final term is the short range contribution between ions in the environment and ions in the QM region. Three body and higher order terms have been excluded from equation 3.50 for simplicity.

An important issue is that the electrostatic potential experienced by both the QM region and the flexible MM region must be converged and the whole system should be charge neutral and have no net dipole moment. This is the region for the outer shell of fixed ions (note that these ions also have short ranged potentials associated with them to prevent boundary effects at the edge of the flexible region). The electrostatic potential experienced by the inner regions can be converged in two ways, in the approach adopted by the QUASI developers a large number of point ions are positioned around the outer, fixed MM ions and the magnitude of the charges at these sites varied until the distribution of electrostatic potentials over a grid of points in the centre of the cluster converges to the correct values. The alternative, used in GUESS, is to converge the potential by construction as described in figure 3.7.

![Figure 3.7](image.png)

**Figure 3.7:** Termination scheme for the edge of a simulation cluster of SiO$_2$. By terminating the cluster with oxygen and assigning terminating oxygen ions a charge of $-1$ and bridging oxygen ions a charge of $-2$ the cluster is guaranteed to be charge natural. A similar scheme for forsterite involves assigning each terminating oxygen a different charge depending on the cations it co-ordinates (see Braithwaite *et al.*, 2002).
A final point is that fixed boundaries used in embedding methods are not strictly equivalent to the Mott-Littleton method. Additional corrections can be made to account for the polarization of the lattice away from the point defect (i.e. beyond the start of the fixed region) but this is just an additive correction and it cannot feed back into the description.

3.7 Extended defects

Compared to point defects much less attention has been paid to the simulation of extended defects in minerals. Methodologies for the study of surfaces are well advanced and include the use of three-dimensional slab models and two-dimensional periodic boundary conditions (see Gay and Rohl, 1994; de Leeuw et al., 2003; Finocchi et al., 2003 for more information). It is worth noting that this methodology is also suitable for the simulation of interfaces, stacking faults, domain walls and most other planar defects. Dislocations have received even less attention. Again, there is a choice between the use of periodic boundary conditions (Cai et al., 2001) and a non-periodic approach (Hoagland et al., 1976). However, as dislocation – dislocation interactions only decay as $1/r$, the non-periodic approach offers clear advantages. A methodology has been developed for the simulation of line defects in minerals (described fully in chapter 6). This involves embedding a simulation cell containing the dislocation core in an elastic representation of the infinite crystal. It is assumed that the dislocation is straight, this allows the cell to be periodic in one dimension, thus considerably simplifying the calculation.

The remaining chapters describe the application of the methodology outlined above to models of point defects and dislocations in forsterite. Chapter 4 and 5 describe models of point defects while chapter 7 outlines models of dislocations. In order to model dislocations in forsterite (and other structurally complex materials) the modelling methodology was modified – these modifications are described in chapter 6.
Chapter 4: Models of point defects in olivine

This chapter contains a collection of results in the form of determinations of the structure and energies of a wide range of point defects in forsterite. These results are an essential prerequisite for the studies of diffusion considered in chapter 5 and in themselves, provide useful data on the likely point defects in olivine. Before describing the point defects, some of the results of the many previous computational studies of forsterite, wadsleyite and ringwoodite are outlined, both from the point of view of the perfect, defect free crystal and the previous work of defective structures. This literature review is followed by a brief overview of studies of defect free forsterite, details of the methodologies applied in this work for the study of defects, with differences to previous studies highlighted, and a presentation and discussion of the major results.

4.1 Review of previous work

There has been a wide array of computational studies of olivine and its high pressure polymorphs, with most research groups with an interest in computational studies of minerals contributing to the literature at one time or another. A discussion of this work is essential in placing the results presented in this chapter and in the remainder of the thesis in context. The literature review is rather arbitrarily separated into those studies that concentrate on the periodic, defect free crystal and those which address problems related to defects. As studies of the perfect crystal are needed prior to undertaking studies of defects, this literature is described first.

4.1.1 Bulk properties

Forsterite has been used as a model silicate material from the earliest days of computational mineralogy. Early work included studies of the structure and energetics of the Mg_2SiO_4 polymorphs (Price and Parker, 1984; Parker and Price, 1985) using static energy minimisation methods based on the use of interatomic potentials. Further work concentrated on the lattice dynamics of forsterite (Price et al., 1987a) and the thermodynamics of the phases and the
pressure and temperature of the phase transitions, all calculated in the quasiharmonic approximation (Price et al., 1987b; Parker and Price, 1989; Price et al., 1989). These early studies showed the importance of including a description of the rigidity of the SiO$_4$ tetrahedra through three body interactions to describe the repulsion between the Si–O bonds.

The earliest electronic structure studies utilised the Hartree-Fock (HF) approximations using a basis set of Gaussian type orbitals implemented in the CRYSTAL code (Silvi et al., 1993). However, cell optimisation for non-cubic cells at the time could only be performed by hand – a demanding task – so only the ringwoodite cell was fully relaxed; studies of forsterite and wadsleyite were limited to calculations based on the experimentally determined cell. The key result of this study comes from an analysis of the electron density that shows that the magnesium ions are isolated and the silicate units form covalently bonded units with net negative charge.

Early studies using density functional theory attempted to elucidate the compressional behaviour of forsterite using the local density approximation to the exchange and correlation functional and the plane wave and pseudopotential methodology to solve the Kohn-Sham equations (Brodholt et al., 1996; Wentzcovitch and Stixrude, 1997). da Silva et al. (1997) calculated the individual elastic constants using a similar methodology and applying small strains to the unit cell. Haiber et al. (1997) carried out similar calculations concentrating on calculating the equation of state and hydrogen solubility for each phase (discussed below), again using the LDA. This work is interesting as it is the only published DFT study to examine all three Mg$_2$SiO$_4$ polymorphs using the same methodology. Because the equations of state of all three phases were calculated, the common tangent construction can be used to extract the pressure of the phase transitions. Comparison of calculated transition pressures with experiment is a rather extreme test for any computational method; not only does the location and curvature of individual minima of the nuclear potential energy surface need to be described (these are also needed for calculation of the structure and elastic constants), but also the relative location of different minima are needed. Bearing in mind the difficulty of this calculation, and the use of the simple LDA functional, the success of the calculation of the transition pressure
is reassuring (see also section 4.2 and table 4.5). Further calculations using a GGA functional and ultra soft (rather than norm-conserving) pseudopotentials was undertaken by de Leeuw and Parker (2000). The computational details and results of these electronic structure studies are summarised later, in table 4.3.

The properties of fayalite have recently been examined using DFT (Cococcioni et al., 2003). Although this study was successful in reproducing the atomic structure of fayalite the magnetic structure was incorrectly evaluated. The high pressure polymorphs have also been studied using the plane wave and pseudopotential approach, apart from the work early of Haiber et al. (1997). Kiefer and co-workers (2001) calculated the elastic constants and compression mechanism of wadsleyite while studies of ringwoodite include calculation of its elastic constants (Kiefer et al., 1997) and lattice dynamics using numerical evaluation of the matrix of second derivatives (Lazewski et al., 2001). Kiefer’s calculations yield useful information for the interpretation of seismic data, and are in reasonable agreement with existing experimental data. A problem with the calculation of Lazewski et al. (2001) is the computational cost, the result of this is that the individual calculations may not have been completely converged with respect to plane wave cut off and k-point sampling possibly resulting in inaccuracies in the calculated frequencies (certainly, the calculated lattice parameter was significantly different from the experimental one).

4.1.2 Defect properties

The earliest work to include the calculation of defect energetics in forsterite was carried out by Wright and Catlow (1994) who used parameterised potentials and the Mott-Littleton method to study the incorporation of water into the forsterite lattice. A similar approach was used by Jaoul et al. (1995) in order to study the formation and diffusion of defects on the magnesium sub lattice. Richmond and Brodholt (2000) used parameterised potentials and the super cell methodology to study the incorporation of iron(III) into forsterite. The incorporation of other impurity species into forsterite has also been studied (Purton et al., 1996; 1997), with results used to analyse trace element partitioning between mantle phases and melts. A key difference in the calculations carried out by Richmond and Brodholt (2000) and Purton et al. (1996; 1997) is that in the former case the charge balancing defects were assumed to be sufficiently separated to be non-
interacting while in the latter they were explicitly bound together. Bringing the
defects of opposite charge together obviously tends to lower the internal energy
contribution to the defect formation energy, it is not so clear that the free energy
will decrease because the association of the defects will tend to decrees the
entropy of the defect formation reaction.

Methods based on parameterised potential functions have also been used to
study the surface structure of forsterite crystals. The first surface calculations of
any silicate mineral were undertaken for dry crystals of pure forsterite (Watson
et al., 1997), these calculations yield calculated equilibrium crystal
morphologies in approximate agreement with the experimental morphology but
showed that in some respects kinetic effects are important. Further work (de
Leeuw et al., 2000a) included the effect of water on the surface structure (both
adsorption of H₂O molecules on the surface and dissociated hydrogen from
water molecules hydrating under-bound oxygen ions on the forsterite surface).
These calculations yielded an equilibrium morphology for forsterite in hydrous
conditions similar to that observed experimentally but with the (010) and (001)
faces significantly too large indicating calculated relative growth rates that were
too slow for these surfaces. Further calculations (de Leeuw et al., 2000b)
concentrated on the (010) grain boundary and surface. The surface calculations
showed the importance of (001) and (100) steps on the (010) surface, this
allowed tilt boundaries to be constructed. In addition, the replacement of
magnesium ions by two hydrogen ions was studied and found to be more
favourable in the grain boundary than in the bulk.

\textit{Ab initio} electronic structure methods based on Density Functional Theory have
also been used. Haiber \textit{et al.} (1997) considered the mechanism of hydrogen
incorporation using the local density approximation and, in the case of
forsterite, found that OH groups form preferentially on the O3 site and
suggested that the binding of OH groups to metal vacancies would be
unfavourable. The generalized gradient approximation was used with a plane
wave basis set to describe a charged cell containing ionic (Brodholt, 1997) and
protonated defects (Brodholt and Refson, 2000). Key observations from these
calculations include the inequivalence in energy required to form vacancies on
the three oxygen sites with vacancies on the O3 site favoured and that hydrogen
substitution decreases the energetic penalty of forming silicon vacancies, thus providing a possible explanation of water weakening in olivine. The results of Brodholt and Refson (2000) and later calculations by de Leeuw and Parker (2000) show that protons will tend to bind to metal vacancies (while still forming OH groups with lattice oxygen ions), this is in agreement with the results of Wright and Catlow (1994) but contrary to the suggestion of Haiber et al. (1997). It is possible that Haiber's erroneous conclusion was due to the use of an under damped, damped molecular dynamics optimisation strategy within a relatively small super cell.

Recently a combined atomistic – electronic structure method was used to examine metal vacancies and hydrogen defects in forsterite (Braithwaite et al., 2002; Braithwaite et al., 2003). This method exploits the strengths of the electronic structure and classical methods in unison. It bypasses possible artefacts caused by introducing a periodically repeating net charge within a super cell that can present difficulties in electronic structure methods and the limitations of the pair-potential technique that may inadequately describe the detailed structure of the defect. This is achieved by embedding an electronic structure description of the defect and its immediate surroundings in a deformable and polarisable atomistic model of the perfect solid. These calculations are in good agreement with the earlier electronic structure calculations and are used as the basis for further calculations described below.

4.2 Calculation of bulk properties

In order to construct a model of defective olivine, it is first necessary to possess a good model of a perfect infinite olivine crystal. It is the purpose of this section to describe and validate these models of the perfect crystal used throughout the thesis. Two basic types of models, electronic structure and parameterised potentials, are used.

For the parameterised potential model well-tested "THB1" potential parameters (Catlow, 1977b; and Sanders et al., 1984; Lewis and Catlow, 1985) were used to model forsterite, fayalite and other phases. Additionally, a simple model discussed below was used to describe the presence of OH defects. The model is therefore based on an ionic description of the inter-atomic interactions with
pair potentials described by formal ionic charges combined with a short range Buckingham potential of the form given in equation 3.22 used to describe oxygen – oxygen and cation – oxygen interactions. This is supplemented by a harmonic three-body bond bending interaction to provide the rigidity required to accurately model semi-covalent SiO$_4$ tetrahedra. The polarization of oxygen is described using the shell model of Dick and Overhauser (1958) described in section 3.2.4. In this model, the charge of the ion is distributed between a negatively charged shell (notionally with no mass) and a positively charged core that interact only via a harmonic potential. Interactions between ions include core-core, core-shell and shell-shell Coulombic interactions and short range interactions that only operate between shells. The formal charges in the model probably over emphasise the ionic nature of olivine, an effect that becomes important in chapter 5. These parameters, derived empirically from experimental data for simple binary oxides (supplemented by quantum mechanical data for the oxygen – oxygen interactions), have been successfully used for the modelling of the bulk (e.g. Price et al., 1987a; Catlow and Price, 1990) and defect (e.g. Wright and Catlow, 1994; Jaoul et al., 1995; Richmond and Brodholt, 2000) properties of forsterite and of a wide range of other silicates.

Validation of this potential model was undertaken by calculating the structure and properties of forsterite, wadsleyite and ringwoodite. Table 4.2 gives the cell parameters, elastic constants and dielectric properties calculated for structures found by minimising the lattice energy using the BFGS local optimisation algorithm starting from the experimental structure employing the GULP code (Gale, 1997; Gale and Rohl, 2003).
Table 4.1: Parameters of the potential model used in this study. Sources and equations are given in the text. The pair potentials are truncated beyond 10Å and the three body interactions only operate between central silicon ions and the first shell of oxygen ions.

<table>
<thead>
<tr>
<th>Ions</th>
<th>Charges (units of electronic charge)</th>
<th>Core - shell spring constant eV·Å²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mg</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Fe²⁺</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Fe³⁺</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>Si</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>O²⁻</td>
<td>0.848190</td>
<td>-2.848190</td>
</tr>
<tr>
<td>O¹.⁴⁻</td>
<td>-1.4260</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>0.4260</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Buckingham potential</th>
<th>A (eV)</th>
<th>ρ (Å)</th>
<th>C (eV·Å⁶)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si – O²⁻</td>
<td>1283.90734</td>
<td>0.32052</td>
<td>10.66158</td>
</tr>
<tr>
<td>Si – O¹.⁴⁻</td>
<td>983.556</td>
<td>0.32052</td>
<td>10.66158</td>
</tr>
<tr>
<td>O²⁻ – O²⁻</td>
<td>22764.0</td>
<td>0.149</td>
<td>27.88</td>
</tr>
<tr>
<td>O²⁻ – O¹.⁴⁻</td>
<td>22764.0</td>
<td>0.149</td>
<td>27.88</td>
</tr>
<tr>
<td>O¹.⁴⁻ – O¹.⁴⁻</td>
<td>22764.0</td>
<td>0.149</td>
<td>27.88</td>
</tr>
<tr>
<td>H – O²⁻</td>
<td>311.97</td>
<td>0.25</td>
<td>0.0</td>
</tr>
<tr>
<td>H – O¹.⁴⁻</td>
<td>311.97</td>
<td>0.25</td>
<td>0.0</td>
</tr>
<tr>
<td>Mg – O²⁻</td>
<td>1428.5</td>
<td>0.29453</td>
<td>0.0</td>
</tr>
<tr>
<td>Mg – O¹.⁴⁻</td>
<td>1060.5</td>
<td>0.29453</td>
<td>0.0</td>
</tr>
<tr>
<td>Fe²⁺ – O²⁻</td>
<td>599.4</td>
<td>0.33990</td>
<td>0.0</td>
</tr>
<tr>
<td>Fe³⁺ – O²⁻</td>
<td>1102.4</td>
<td>0.32990</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Morse (coulomb subtracted)

<table>
<thead>
<tr>
<th>Morse</th>
<th>D (eV)</th>
<th>α (Å⁻¹)</th>
<th>r₀ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H – O¹.⁴⁻</td>
<td>7.0525</td>
<td>2.030</td>
<td>0.94850</td>
</tr>
</tbody>
</table>

Harmonic three-body

| O²⁻ – Si – O²⁻ | 2.09724 | 109.47 |
| O¹.⁴⁻ – Si – O²⁻ | 2.09724 | 109.47 |
| O¹.⁴⁻ – Si – O¹.⁴⁻ | 2.09724 | 109.47 |
Table 4.2: Comparison of properties of forsterite, wadsleyite and ringwoodite calculated using the potential model with experimental data (in italics).

<table>
<thead>
<tr>
<th>Property</th>
<th>Forsterite</th>
<th>Wadsleyite</th>
<th>Ringwoodite</th>
</tr>
</thead>
<tbody>
<tr>
<td>a (Å)</td>
<td>4.782</td>
<td>5.652</td>
<td>8.016</td>
</tr>
<tr>
<td></td>
<td>4.734(6)</td>
<td>5.696(1)</td>
<td>8.0649(1)</td>
</tr>
<tr>
<td>b (Å)</td>
<td>10.246</td>
<td>11.391</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>10.1902(15)</td>
<td>11.453(1)</td>
<td>-</td>
</tr>
<tr>
<td>c (Å)</td>
<td>5.986</td>
<td>8.275</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>5.9783(7)</td>
<td>8.256(1)</td>
<td>-</td>
</tr>
<tr>
<td>C_{11} (GPa)</td>
<td>358.62</td>
<td>434.60</td>
<td>412.96</td>
</tr>
<tr>
<td></td>
<td>328(6)</td>
<td>360(6)</td>
<td>370.5(78)</td>
</tr>
<tr>
<td>C_{22} (GPa)</td>
<td>206.62</td>
<td>425.65</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>383(4)</td>
<td>367.7(65)</td>
<td>-</td>
</tr>
<tr>
<td>C_{33} (GPa)</td>
<td>281.14</td>
<td>331.68</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>235</td>
<td>272.4(58)</td>
<td>-</td>
</tr>
<tr>
<td>C_{44} (GPa)</td>
<td>44.23</td>
<td>101.61</td>
<td>136.74</td>
</tr>
<tr>
<td></td>
<td>67</td>
<td>112(2)</td>
<td>131(2)</td>
</tr>
<tr>
<td>C_{55} (GPa)</td>
<td>74.54</td>
<td>113.47</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>118(4)</td>
<td>122.5(40)</td>
</tr>
<tr>
<td>C_{66} (GPa)</td>
<td>84.29</td>
<td>100.57</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>98(4)</td>
<td>103.1(39)</td>
</tr>
<tr>
<td>C_{12} (GPa)</td>
<td>93.83</td>
<td>118.61</td>
<td>169.40</td>
</tr>
<tr>
<td></td>
<td>69</td>
<td>75(9)</td>
<td>65.6(45)</td>
</tr>
<tr>
<td>C_{13} (GPa)</td>
<td>96.18</td>
<td>136.19</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>69</td>
<td>98(4)</td>
<td>95.2(52)</td>
</tr>
<tr>
<td>C_{23} (GPa)</td>
<td>87.70</td>
<td>144.10</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>73</td>
<td>105(9)</td>
<td>105.1(44)</td>
</tr>
<tr>
<td>ε_{11}₀</td>
<td>6.09</td>
<td>7.13</td>
<td>7.23</td>
</tr>
<tr>
<td>ε_{22}₀</td>
<td>6.97(21)</td>
<td>7.34</td>
<td>7.85</td>
</tr>
<tr>
<td>ε_{33}₀</td>
<td>5.96</td>
<td>7.27</td>
<td></td>
</tr>
<tr>
<td>ε_{11}₁₀⁺</td>
<td>2.11</td>
<td>2.24</td>
<td>2.31</td>
</tr>
<tr>
<td>ε_{22}₁₀⁺</td>
<td>2.06</td>
<td>2.28</td>
<td></td>
</tr>
<tr>
<td>ε_{33}₁₀⁺</td>
<td>2.09</td>
<td>2.23</td>
<td></td>
</tr>
</tbody>
</table>

Cell parameters for forsterite (a, Fujino et al., 1981) wadsleyite (b, Sawamoto et al., 1984) and ringwoodite (c, Sasaki et al., 1982). Elastic constants are determined by single crystal Brillouin spectroscopy for forsterite (d, Ahrens, 1995), wadsleyite (b, Sawamoto et al., 1984) and (e, Zha et al., 1997a) and ringwoodite (f, Jackson et al., 2000). The dielectric constants for 1MHz at 25 °C (g, Cygan and Lasaga, 1986).

As can be seen from table 4.2 the cell parameters of all three phases are well described by the model, with all parameters agreeing with experiment to within 0.8% and the elastic constants generally agree to within 10%. The forsterite cell is predicted to be too large while the ringwoodite cell is predicted to be too small. Some of the elastic constants are in greater disagreement with experiment, notably for ringwoodite where an improved model utilising the
breathing shell model has recently been constructed (Blanchard et al., In prep) with much better reproduction of the elastic constants. The principal elastic constants of wadsleyite, and to a lesser extent forsterite, are over-estimated. This results in a calculated bulk modulus which is too high. Interestingly the calculated phase transition is in fair agreement with experiment (Parker and Price, 1989) because the difference in energy of the phases is reduced, cancelling the error. It is also worth noting that the ordering of some of the elastic constants is wrong, notably the calculated shear constants ($C_{44}$, $C_{55}$ and $C_{66}$) for forsterite are higher than the off diagonal elements ($C_{12}$, $C_{13}$ and $C_{23}$) while the experimental shear constants are stiffer than the off diagonal elements. In wadsleyite the situation is reversed with calculated off diagonal elements too stiff and shear elements not stiff enough, again inverting the order of some of these elements. These problems are only expected to lead to minor errors in studies of point defects (where the dielectric response is more important than the elastic deformation) while the errors are much more significant for studies of dislocations. This issue is discussed in more detail in chapter 7.

As well as the use of parameterised inter-atomic potentials, use is also made of electronic structure methods in this thesis. The bulk structure and compressional behaviour of the three polymorphs of Mg$_2$SiO$_4$, forsterite, wadsleyite and ringwoodite, have been calculated using the new version of the CASTEP code (Segall et al., 2002) and the linear scaling SIESTA code (Soler et al., 2002). Recently Civalleri and Harrison (Civalleri and Harrison, 2002) emphasised the importance of ensuring that the exchange - correlation functional used to generate all electron results to fit the pseudopotentials is compatible with the functional used in the calculation using the pseudopotential. In the light of this, and the availability of a new version of the CASTEP code, it was decided to calculate the equations of state for all three phases. Two sets of ultra soft pseudopotentials and one set of norm conserving pseudopotentials were used. The norm conserving pseudopotentials are from the library supplied with the old version of the CASTEP code, they are believed to be identical to those used by Brodholt and co-workers (Brodholt et al., 1996; Brodholt, 1997; Brodholt and Refson, 2000). The first set of ultra-soft pseudopotentials is that supplied with the Accelrys version of the new CASTEP code, fitted to LDA all
electron results and similar to those used by de Lceuw and Parker (2000). The second set of ultra-soft pseudopotentials are those derived by Civalleri and Harrison (2002) fitted to GGA all electron results, it should be noted that the Mg pseudopotential for this study is from the Accelrys library as no GGA pseudopotential was available. Although this is certainly far from ideal it was decided that it would be worthwhile to pursue this study with a “correct” description of the SiO$_4$ units.

The results for forsterite, along with details of the previous studies described above, are presented in table 4.3. Several points can be made about these results, firstly all the LDA results in the literature (Brodholt et al., 1996; Wentzcovitch and Stixrude, 1997) and calculated as part of this study, underestimate the cell volume by 2-7%. This is to be expected given the tendency for LDA to over-bind. In contrast, the GGA calculations are more confused, with some studies (e.g. de Lceuw and Parker, 2000) giving cell volumes a few percent too large but the results of Brodholt and co-workers giving volumes a few percent too small (Brodholt, 1997; Brodholt and Refson, 2000). Close examination shows that the improvement of cell volume is at the expense of an accurate description of cell shape with the errors on the $c/a$ and $b/c$ ratios often being in the opposite sense. When using GGA pseudopotentials with a GGA functional, the errors on the cell edges tend to be equal, indicating that the whole cell has expanded rather than changed shape. This is closer to the expected behaviour for a GGA functional and suggests that all electron calculations would be of value. Results of using the linear scaling SIESTA code are in fair agreement with the new plane-wave calculations, but the size of the problem was too small for efficient use of the order-$N$ parts of the code so use of SIESTA was not pursued.
Table 4.3: Summery of electronic structure studies of forsterite, including available computational details, cell parameters and errors.

<table>
<thead>
<tr>
<th></th>
<th>a (Å)</th>
<th>b (Å)</th>
<th>c (Å)</th>
<th>V (Å³)</th>
<th>c/a</th>
<th>b/c</th>
<th>RMS error on cell edges</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% error</td>
<td>% error</td>
<td>% error</td>
<td>% error</td>
<td>% error</td>
<td>% error</td>
<td></td>
</tr>
<tr>
<td>Brodholt et al. &amp;</td>
<td>4.64</td>
<td>9.99</td>
<td>6.07</td>
<td>281.68</td>
<td>1.31</td>
<td>1.64</td>
<td></td>
</tr>
<tr>
<td>Brodholt (1996; 1997):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCP, 600eV cut off,</td>
<td>LDA, 1kt.</td>
<td>4.69</td>
<td>10.08</td>
<td>5.99</td>
<td>283.18</td>
<td>1.28</td>
<td>1.68</td>
</tr>
<tr>
<td>CETEP.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-2.32%</td>
<td>-1.98%</td>
<td>1.60%</td>
<td>-2.73%</td>
<td>4.02%</td>
<td>-3.53%</td>
<td>1.97%</td>
</tr>
<tr>
<td></td>
<td>-1.33%</td>
<td>-1.08%</td>
<td>0.20%</td>
<td>-2.21%</td>
<td>1.55%</td>
<td>-1.27%</td>
<td>0.87%</td>
</tr>
<tr>
<td></td>
<td>-2.18%</td>
<td>-1.38%</td>
<td>1.53%</td>
<td>-2.04%</td>
<td>3.79%</td>
<td>-2.87%</td>
<td>1.69%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGA, 1 kpt.</td>
<td>4.65</td>
<td>10.05</td>
<td>6.07</td>
<td>283.67</td>
<td>1.31</td>
<td>1.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-2.18%</td>
<td>-1.38%</td>
<td>1.53%</td>
<td>-2.04%</td>
<td>3.79%</td>
<td>-2.87%</td>
<td>1.69%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGA, 4 kpts.</td>
<td>4.71</td>
<td>10.15</td>
<td>5.96</td>
<td>284.93</td>
<td>1.27</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.91%</td>
<td>-0.39%</td>
<td>-0.31%</td>
<td>-1.61%</td>
<td>0.61%</td>
<td>-0.09%</td>
<td>0.54%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haiber et al. (1997)a.</td>
<td>4.73</td>
<td>10.09</td>
<td>5.96</td>
<td>284.47</td>
<td>1.26</td>
<td>1.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.50%</td>
<td>-0.97%</td>
<td>-0.31%</td>
<td>-1.76%</td>
<td>0.19%</td>
<td>-0.67%</td>
<td>0.59%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wentzcovitch, and</td>
<td>4.68</td>
<td>9.95</td>
<td>5.84</td>
<td>272.00</td>
<td>1.25</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td>Stixrude (1997): LDA,</td>
<td>70 Ry cut off,</td>
<td>4.65</td>
<td>9.97</td>
<td>5.85</td>
<td>270.99</td>
<td>1.26</td>
<td>1.71</td>
</tr>
<tr>
<td>NCP.</td>
<td>2<em>2</em>2 MP grid.</td>
<td>-1.50%</td>
<td>-2.33%</td>
<td>-2.36%</td>
<td>-6.07%</td>
<td>-0.87%</td>
<td>0.04%</td>
</tr>
<tr>
<td></td>
<td>64 Ry cut off,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2<em>2</em>2 MP grid.</td>
<td>-2.20%</td>
<td>-2.15%</td>
<td>-2.21%</td>
<td>-6.42%</td>
<td>-0.02%</td>
<td>0.06%</td>
</tr>
<tr>
<td></td>
<td>8 kpts. 70 Ry cut off.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.59%</td>
<td>-2.34%</td>
<td>-2.31%</td>
<td>-6.11%</td>
<td>-0.74%</td>
<td>-0.02%</td>
<td>2.08%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Leeuw and Parker</td>
<td>4.78</td>
<td>10.30</td>
<td>6.01</td>
<td>295.92</td>
<td>1.26</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td>(2000).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.58%</td>
<td>1.05%</td>
<td>0.55%</td>
<td>2.19%</td>
<td>-0.03%</td>
<td>0.50%</td>
<td>0.73%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This work. K-BP 600eV</td>
<td>4.65</td>
<td>9.99</td>
<td>6.08</td>
<td>282.03</td>
<td>1.31</td>
<td>1.64</td>
<td></td>
</tr>
<tr>
<td>cut off, NewCASTEP.</td>
<td>LDA 600eV 1kt.</td>
<td>4.66</td>
<td>10.03</td>
<td>6.07</td>
<td>283.51</td>
<td>1.30</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>-2.28%</td>
<td>-1.98%</td>
<td>1.68%</td>
<td>-2.61%</td>
<td>4.06%</td>
<td>-3.61%</td>
<td>1.98%</td>
</tr>
<tr>
<td></td>
<td>-2.05%</td>
<td>-1.62%</td>
<td>1.60%</td>
<td>-2.10%</td>
<td>3.73%</td>
<td>-3.17%</td>
<td>1.76%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This work. USP (fitted</td>
<td>4.72</td>
<td>10.12</td>
<td>5.95</td>
<td>284.61</td>
<td>1.26</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td>to LDA data) New</td>
<td>LDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASTEP, 700eV cut off,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3<em>2</em>3 MP grid,</td>
<td>GGA</td>
<td>4.77</td>
<td>10.30</td>
<td>6.04</td>
<td>296.47</td>
<td>1.27</td>
<td>1.71</td>
</tr>
<tr>
<td>NewCASTEP.</td>
<td>PBE</td>
<td>4.77</td>
<td>10.31</td>
<td>6.05</td>
<td>297.10</td>
<td>1.27</td>
<td>1.71</td>
</tr>
<tr>
<td></td>
<td>0.31%</td>
<td>1.16%</td>
<td>1.12%</td>
<td>2.60%</td>
<td>0.81%</td>
<td>0.04%</td>
<td>0.86%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This work. USP (fitted</td>
<td>4.83</td>
<td>10.39</td>
<td>6.07</td>
<td>304.25</td>
<td>1.26</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td>to GGA data) New</td>
<td>PW91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASTEP, 700eV cut off,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3<em>2</em>3 MP grid,</td>
<td>PBE</td>
<td>4.83</td>
<td>10.38</td>
<td>6.08</td>
<td>304.78</td>
<td>1.26</td>
<td>1.71</td>
</tr>
<tr>
<td>NewCASTEP.</td>
<td>1.59%</td>
<td>1.90%</td>
<td>1.67%</td>
<td>5.25%</td>
<td>0.08%</td>
<td>0.23%</td>
<td>1.72%</td>
</tr>
<tr>
<td>Atom-like basis (SIESTA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.84</td>
<td>10.45</td>
<td>6.12</td>
<td>309.48</td>
<td>1.27</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.76%</td>
<td>2.56%</td>
<td>2.40%</td>
<td>6.87%</td>
<td>0.63%</td>
<td>0.15%</td>
<td>2.24%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment: Fujino et al. (1981), Acta Cryst,</td>
<td>4.7534</td>
<td>10.1902</td>
<td>5.9783</td>
<td>289.578</td>
<td>1.258</td>
<td>1.705</td>
<td>-</td>
</tr>
<tr>
<td>B37513-518</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Further investigation of the accuracy of the DFT calculations involved calculating the athermal equations of state (EOS) for each phase (see figure 4.1). This was done by performing a full geometry and cell optimisation under applied hydrostatic stress between 0 and 60 GPa (i.e. through the whole transition zone) – no assumptions were made regarding the locations of ions or ratios of lattice parameters remaining equivalent on compression. The resulting enthalpies and cell volumes were fitted to the Murnaghan EOS in order to be consistent with the results of Haiber et al. (1997):

\[
E(V) = \frac{K_0 V}{K'_0} \left[ \left( \frac{V_0}{V} \right)^{K'_0} - 1 \right] + E_0
\]

(4.1)

where the enthalpy, \( E(V) \), at a given volume, \( V \), is a function of the zero pressure energy, \( E_0 \), and volume, \( V_0 \), along with the bulk modulus, \( K_0 \), and its pressure derivative \( K'_0 \). Although this is perhaps not the best EOS for the extrapolation of experimental data (see Cohen et al., 2000; Poirier, 2000) it is certainly sufficient to compare with the results of Haiber et al. (1997) and for interpolation to examine the effect of changing the pseudopotential.

Table 4.4: Parameters of fitted Murnaghan athermal equation of state (equation 4.1) using two different sets of pseudopotentials compared to the results of Haiber et al. (1997). Volume - energy data is presented in figure 4.1, volume in Å\(^3\) per 56 atom cell, \( K_0 \) in Mbar, \( K'_0 \) is dimensionless and \( E_0 \) in eV per 56 atom cell.

<table>
<thead>
<tr>
<th></th>
<th>LDA study (Haiber et al., 1997)</th>
<th>This study, GGA with LDA pseudopotentials</th>
<th>This study, GGA with GGA pseudopotentials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha ) ( \beta ) ( \gamma )</td>
<td>( \alpha ) ( \beta ) ( \gamma )</td>
<td>( \alpha ) ( \beta ) ( \gamma )</td>
</tr>
<tr>
<td>( V_0 )</td>
<td>586.6 545.2 533.5</td>
<td>591.7 547.2 531.4</td>
<td>607.2 563.4 548.3</td>
</tr>
<tr>
<td>( K_0 )</td>
<td>1.37 1.79 1.89</td>
<td>0.79 1.04 1.15</td>
<td>0.76 0.99 1.11</td>
</tr>
<tr>
<td>( K'_0 )</td>
<td>3.75 4.26 4.08</td>
<td>3.49 3.52 3.59</td>
<td>3.53 3.63 3.52</td>
</tr>
<tr>
<td>( E_0 )</td>
<td>0.00 2.69 3.95</td>
<td>0.00 2.43 3.30</td>
<td>0.00 2.05 3.94</td>
</tr>
</tbody>
</table>
Figure 4.1: Data points from DFT calculations of equation of state of forsterite (red) wadsleyite (green) and ringwoodite (blue) using LDA (a) and GGA (b) ultra-soft pseudopotentials. Lines are the result of the fitted equation of state.
The fitted values of the EOS parameters are given in table 4.4. Comparing the GGA results with the LDA results shows the general tendency for GGA to yield larger cell volumes and lower values for the bulk modulus than LDA results. Table 4.5 gives the estimated transition pressure for the various studies derived using a common tangents construction. The GGA results predict transition pressure within the experimental error.

Table 4.5: Estimated transition pressures (in GPa) for the two phase transitions in the Mg$_2$SiO$_4$ system. Experimental results are taken from Akimoto and Kawada and extrapolated to zero Kelvin by Haiber (1997). Haiber’s results are for DFT calculations within the LDA. Results from this study are discussed in the text.

<table>
<thead>
<tr>
<th></th>
<th>Experimental results</th>
<th>LDA study (Haiber et al., 1997)</th>
<th>This study, GGA with LDA pseudopotentials</th>
<th>This study, GGA with GGA pseudopotentials</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$ to $\beta$</td>
<td>10 ±2</td>
<td>12.4</td>
<td>15.2</td>
<td>11.2</td>
</tr>
<tr>
<td>$\beta$ to $\gamma$</td>
<td>15 ±2</td>
<td>20.0</td>
<td>16.5</td>
<td>13.1</td>
</tr>
</tbody>
</table>

4.3 Modelling point defects - methodology

Two methods have been used to calculate point defect properties in olivine. The generalised Mott-Littleton method is used to study all defect species while the QM/MM embedded cluster approach, previously used for studies of defects in forsterite by Braithwaite and co-workers (2002; 2003), is used here to study selected defects. The embedded cluster methodology is not used for studies of all defects due to its large computational cost. The General Utility Lattice Program ‘GULP’ (Gale, 1997; Gale and Rohl, 2003) was used to perform the Mott-Littleton calculations and GUESS (Gaussian Used for Embedded System Studies, Sushko et al., 2000a; Sushko et al., 2000b) was used for the QM/MM studies.

4.3.1 Mott-Littleton calculations

The energy of point defects and defect clusters are calculated using the generalised Mott-Littleton method described in section 3.6.2.1 and as implemented within GULP. In this method, a region around the defect (region I) is relaxed to an energy minimum using a model based on the interatomic
potentials given in table 4.1. Setting the size of region I is one of the major sources of numerical error in Mott-Littleton calculations. If it is too small then the assumption of harmonic displacements in region II is invalid and the calculated defect energy is too high, but the maximum size of region I is limited by the speed and memory size of the available computational resources. Following convergence tests for a number of defects region I was truncated at a radius of 10Å from the defect. Further increasing the size of region I lowers the energy by less than 0.1 eV. This gives an estimate of the error in the determination of defect energies due to the Mott-Littleton approach which is much smaller than the expected errors due to the use of parameterised potentials. The size of region IIa for both the convergence testing and production calculations was set to the radius of region I plus the longest short ranged cut off plus 0.5 angstroms. By construction, this should be a large enough size for region IIa to avoid boundary effects, a fact further confirmed by convergence tests.

In all cases the default BFGS optimiser was used to find low energy defect structures and the vibrational frequencies\textsuperscript{12} were calculated to confirm that a minimum (rather than some other stationary point) had been reached. The centre of region I was always displaced from the defect itself to break the symmetry of the initial defect structure. In a few cases unphysical negative frequencies were calculated, these indicate the presence of imaginary eigenvalues in the diagonalised second derivatives matrix showing that the optimiser had converged to a stationary point that was not a minimum. In these cases the structure was re-optimised with a slightly different initial configuration or using a different optimisation strategy until a true minimum

\textsuperscript{12} Within GULP, vibrational frequencies calculated for defect structures determined using the using the Mott-Littleton method are calculated in the same way as calculations for an isolated cluster. The cluster used comprises of all atoms in region I and the second derivatives of the energy with respect to all these atoms' degrees of freedom are calculated. This leads to possible problems because atoms in the correct position for the Mott-Littleton calculation in region I need not be at energy minimum in the cluster case (a factor exacerbated by the need to optimise the Mott-Littleton calculation using a force balance rather than energy minimisation scheme) because of forces applied from atoms in region IIa. This is a particular problem for the accurate calculation of vibrational frequencies coupled to long wavelength lattice vibrations (Gale and Rohl, 2003) but does not appear to cause major problems here – in all cases it was possible to successfully find an energy minimum within the Mott-Littleton method that yielded only positive frequencies for the cluster based calculation.
was located (a combination of the conjugate gradients optimizer followed by use of the rational function optimizer once the gradient of the energy began to decrease often proved to be a successful choice).

4.3.3 Embedded cluster calculations

The QM/MM embedded cluster method described in section 3.6.2.2 has been used to study a small number of the defect species in forsterite. This work is best considered as an extension to the previous studies by Braithwaite et al. (2002; 2003) and an identical setup is used (in fact, an identical basic bulk like input file was used as a starting point). Two simulation systems are therefore used, both consist of a 30 Å spherical "nanocluster" of ions. For the smaller calculations the outer 18 Å of which is held fixed at the lattice sites, and the inner 12 Å is allowed to relax to an energy minimum during the calculation, this nanocluster is centred on an Mg1 ion. The larger nanocluster is centred on a silicon ion and the inner 14 Å is allowed to relax. Close to the defect, atoms are treated using an electronic structure method in the "QM cluster". For the larger setup used to study defects around the Si site 97 ions are included in the QM cluster and for the smaller setup, used for defects close to the M1 site, 43 ions are included\textsuperscript{13}. The nanocluster itself is constructed to give a good representation of the long-range electrostatic properties over the inner part. This is achieved in two ways, firstly the cluster is centred on an inversion centre or mirror plane and secondly the edge of the nanocluster is terminated as described in figure 3.7. The electrostatic potential across region I is shifted and a correction term described below must be applied.

The potentials used to describe the molecular mechanics part of the model were fitted by Braithwaite et al. (2002) using charges for the ions calculated using natural population analysis of the Hartree-Fock electron density. Use of these potentials, which are reported in table 4.6, gives consistent forces in the QM and MM parts of the model.

\textsuperscript{13} This included the sites with Mg pseudopotentials used to terminate the QM cluster.
Table 4.6: Parameters of the potential model used for the molecular mechanics part of the embedded cluster study, fitted by Braithwaite et al. (2002). Sources and equations are given in the text. The pair potentials are truncated beyond 15 Å and the three body interactions only operate between central silicon ions and the first shell of oxygen ions with O – O distances less than 3.0 Å and Si – O distances less than 2.0 Å.

<table>
<thead>
<tr>
<th>Charges</th>
<th>Core</th>
<th>Shell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mg</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Si</td>
<td>2.8</td>
<td>-2.4</td>
</tr>
<tr>
<td>O</td>
<td>0.7</td>
<td>53.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Buckingham potential</th>
<th>A (eV)</th>
<th>ρ (Å)</th>
<th>C (eV·Å⁻⁶)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si – O</td>
<td>802.41</td>
<td>0.32052</td>
<td>10.66</td>
</tr>
<tr>
<td>O – O</td>
<td>22764.00</td>
<td>0.14900</td>
<td>27.88</td>
</tr>
<tr>
<td>Mg – O</td>
<td>1316.88</td>
<td>0.29453</td>
<td>0.00</td>
</tr>
<tr>
<td>H – O</td>
<td>311.97</td>
<td>0.25000</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Harmonic three-body</th>
<th>K (eV·rad⁻²)</th>
<th>θ₀(degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O – Si – O</td>
<td>2.09</td>
<td>109.47</td>
</tr>
</tbody>
</table>

A series of additional contributions are needed to allow the calculation of defect energies by the QM/MM scheme. The first of these is the self energy of an isolated ion, which is not defined to zero in electronic structure schemes. The second contribution arises because the electrostatic potential in the nanocluster is not equal to the electrostatic potential on the same site in the infinite lattice (a problem not encountered in Mott-Littleton calculations because of the use of lattice summation techniques to evaluate the electrostatic contributions from an infinite region II). This correction term is given by:

\[ E_{ESP} = \Delta V \times q_i \]  \hspace{1cm} (4.2)

where \( \Delta V \) is the change in the electrostatic potential and \( q_i \) is the charge on the defect. In this case \( \Delta V \) is 10.60 eV. The third contribution is due to the neglect of polarisation in the outer region of the nanocluster, where ions are held fixed at their lattice sites, and the unrepresented infinite crystal. This contribution can be estimated by performing a Mott-Littleton calculation with region I equal to the radius of the relaxed portion of the nanocluster. The polarization
contribution to the total defect energy should be the same for the Mott-Littleton
and QM/MM methods\textsuperscript{14}. The defect energy can then be written as:

$$E_{\text{defect}} = \left( E_{\text{defect}} + E_{\text{ML}} + \sum E_{\text{self}} + \sum E_{\text{ESP}} \right) - \left( E_{\text{bulk}} + \sum E_{\text{self}} + \sum E_{\text{ESP}} \right)$$ (4.3)

$E_{\text{defect}}$ is the total energy of a nanocluster containing a defect, $E_{\text{bulk}}$ is the total
defect of the same nanocluster without the defect, $E_{\text{ML}}$ is the Mott-Littleton
correction. $E_{\text{ESP}}$ is defined in equation 4.2 and $E_{\text{self}}$ is the energy of an isolated
ion, for vacancies, this appears in the first term, for interstitials it appears in the
second term and for impurities self energies appear in both parts of the equation.

\textbf{4.4 Modelling point defects - Results}

Some of the defect energies calculated using both the Mott-Littleton and
embedded cluster methods are presented in table 4.7 and the total energies of the
component embedded cluster calculations are reported in appendix 2. The
defect energies are the amount of energy required to form an isolated defect in
an otherwise perfect crystal ignoring thermal vibrations and at zero pressure. In
general, there is good agreement between energies calculated using each
method; the following subsections address the structure and defect energies of
each defect in turn.

\textbf{4.4.1 Vacancies}

The structures of the vacancies are determined by the simple approach of
removing the appropriate ion from the simulation and minimising the energy,
there is no ambiguity in setting up the simulation. The smaller QM cluster was
used for embedded cluster calculations of defects on the Mg\textsubscript{1} site while the
larger cell was used for calculations on the Si and O sites. The smaller
simulation cell for the magnesium vacancy allowed a detailed study of the effect
of basis set truncation, which informs the choice of methodology for the larger
cell where detailed study becomes impossible because of the large size of the
cell and the unfavourable scaling of some of the methods used.

\textsuperscript{14} But in the QM/MM method the polarization of the lattice at long range is not permitted to
have any effect on the defect structure, it seems that this contribution is only a minor
cortribution to the total defect energy.
### Table 4.7: Defect energies for a range of possible intrinsic defects in forsterite.

<table>
<thead>
<tr>
<th>Defect (a)</th>
<th>Mott-Littleton method (eV)</th>
<th>Embedded cluster method (b) (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{O_1}^\ast$</td>
<td>27.97</td>
<td>-</td>
</tr>
<tr>
<td>$V_{O_2}^\ast$</td>
<td>25.20</td>
<td>- (c)</td>
</tr>
<tr>
<td>$V_{O_3}^\ast$</td>
<td>24.54</td>
<td>- (c)</td>
</tr>
<tr>
<td>$O_{i(1)}^-$</td>
<td>-14.37</td>
<td>-</td>
</tr>
<tr>
<td>$O_{i(2)}^-$</td>
<td>-16.11</td>
<td>-</td>
</tr>
<tr>
<td>$V_{Mg1}^\ast$</td>
<td>24.48</td>
<td>25.63</td>
</tr>
<tr>
<td>$V_{Mg2}^\ast$</td>
<td>26.40</td>
<td>-</td>
</tr>
<tr>
<td>$M_{Si(1ko)}^\ast$</td>
<td>-17.75</td>
<td>-18.04</td>
</tr>
<tr>
<td>$V_{Si}^\ast$</td>
<td>100.81</td>
<td>96.08</td>
</tr>
<tr>
<td>$S_{li}^\ast$</td>
<td>-76.60</td>
<td>-</td>
</tr>
<tr>
<td>$Fe_{Mg1}^\ast$</td>
<td>1.08</td>
<td>-</td>
</tr>
<tr>
<td>$Fe_{Mg2}^\ast$</td>
<td>0.97</td>
<td>-</td>
</tr>
<tr>
<td>$Fe_{Mg1}^\ast$</td>
<td>-22.55</td>
<td>-</td>
</tr>
<tr>
<td>$Fe_{Mg2}^\ast$</td>
<td>-23.24</td>
<td>-</td>
</tr>
</tbody>
</table>

(a) Defects are described using Kröger-Vink defect notation and include vacancies in all three oxygen positions, oxygen interstitial ions occupying both free octahedral sites on both magnesium sites, a split interstitial magnesium defect, a silicon vacancy and a silicon interstitial. Defect energies are quoted with respect to the perfect forsterite lattice and the ion at infinity.

(b) Embedded cluster results are for the B3LYP hybrid functional with the structure optimised using the 6-31G basis set and energies calculated using the 6-31+G* basis set. Details of the effects of basis set truncation and the use of different methods are given in the text and in tables 4.8, 4.9 and 4.10.

(c) Although the energies of these defect species have been calculated the defect energy is not available – see text for details.

### 4.4.1.1 Magnesium vacancies

The first point to note is that vacancies on the Mg1 site are favoured over vacancies on the Mg2 site by approximately 2 eV. This means that essentially all magnesium vacancies should form on the Mg1 site. The embedded cluster calculations show little basis set truncation effects (table 4.8, figure 4.2), explicit relaxation with a 6-31+G* or 6-311+G* basis set alters the calculated energy by less than 0.1 eV. Calculations using the HF approximation give the lowest defect energy while DFT with the PW91 functional gives the largest defect energy (almost 1 eV higher in energy than the HF result) suggesting that
correlation effects tend to increase the binding of Mg to the forsterite lattice. As expected the B3LYP hybrid functional yields intermediate energies.

The embedded cluster calculations, which are limited to calculations of the Mg1 vacancy, are in good agreement with the Mott-Littleton results, with calculated defect energy approximately 0.5 eV lower than the HF result. This agreement is hardly surprising given that magnesium is an ionic species in forsterite, with Mulliken population analysis of the electron density giving charges in the region of +2. The fact that the ion is so ionic is the reason for the remarkable agreement between ML and embedded cluster calculations – the parameterised potential is a very good description of a spherical closed shell ion.

\[ \begin{align*}
\text{Comparison of defect energy convergence with basis set size.}
\end{align*} \]

\textbf{Figure 4.2:} Convergence in defect energy of a Mg1 vacancy in forsterite with basis set size. The red points are the embedded cluster DFT results with the PW91 GGA functional, blue is the hybrid B3LYP methodology and black uses the HF approximation. The Mott-Littleton results are plotted in green for comparison. The basis set is given on the abscissa and refers to the basis used for geometry optimisation and final energy calculation respectively (so 6-31+G*//6-31G means optimisation using the 6-31G basis with final energy calculation using the 6-31+G* basis, these results are quoted in table 4.7). Rapid convergence is observed for each method with basis set size.
Table 4.8: Defect energies for Mg1 vacancies in forsterite calculated using the embedded cluster method. See caption to figure 4.2 for the meaning of the basis set symbols.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>Defect energy (eV)</th>
<th>Hartree-Fock</th>
<th>DFT (PW91)</th>
<th>DFT (B3LYP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-31G//6-31G</td>
<td>25.832</td>
<td>27.127</td>
<td>26.875</td>
<td></td>
</tr>
<tr>
<td>6-31+G*/6-31G</td>
<td>24.942</td>
<td>25.725</td>
<td>25.564</td>
<td></td>
</tr>
<tr>
<td>6-31+G*/6-31+G*</td>
<td>24.979</td>
<td>25.762</td>
<td>25.610</td>
<td></td>
</tr>
<tr>
<td>6-311+G*/6-31+G*</td>
<td>25.002</td>
<td>25.780</td>
<td>25.632</td>
<td></td>
</tr>
<tr>
<td>6-311+G*/6-311+G*</td>
<td>25.017</td>
<td>25.773</td>
<td>25.625</td>
<td></td>
</tr>
</tbody>
</table>

4.4.1.2 Silicon vacancies

The worst agreement between Mott-Littleton and embedded cluster methods, where both sets of data have been calculated, is for the silicon vacancy. This discrepancy has been noted by Braithwaite et al. (2003) and is due to at least two possible causes. Firstly the 6-31+G* basis set is likely to be insufficient to describe the electronic polarization of oxygen ions around the silicon vacancy (with an effective charge of +4) and secondly two 5-co-ordinate silicon atoms are formed in the embedded cluster calculations (shown in figure 4.3). As it is based on parameterised potentials, it may be difficult for the Mott-Littleton method to reproduce the energetics of this structure.

As with the magnesium vacancy the calculated defect energies of the silicon vacancy is lowest using the HF approximation and highest for DFT within the GGA, indicating that the effect of the inclusion of correlation is to destabilise the defect (see table 4.9). Further discussion of this defect is presented in section 4.5.
Figure 4.3: Structure of silicon vacancy predicted by embedded cluster method. [100] is orientated up the page and the view is approximately along [010] (rotated to show structure more clearly). Note the two 5-coordinate silicon atoms (at 5 and 7 o'clock) that form bonds with the O3 oxygen ions, the two isolated oxygen ions are O1 and O2. Magnesium ions are blue, silicon ions are pink and oxygen ions are red.

Table 4.9: Defect energies for vacancies on the silicon site using various basis sets and methodologies within the embedded cluster methodology.

<table>
<thead>
<tr>
<th>Method</th>
<th>Defect energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF/6-31G//HF/6-31G</td>
<td>92.039</td>
</tr>
<tr>
<td>HF/6-31+G*/HF/6-31G</td>
<td>95.333</td>
</tr>
<tr>
<td>PW91/6-31G//PW91/6-31G</td>
<td>94.118</td>
</tr>
<tr>
<td>PW91/6-31+G*/PW91/6-31G</td>
<td>95.774</td>
</tr>
<tr>
<td>B3LYP/6-31G/B3LYP/6-31G</td>
<td>93.875</td>
</tr>
<tr>
<td>B3LYP/6-31+G*/B3LYP/6-31G</td>
<td>96.083</td>
</tr>
</tbody>
</table>
4.4.1.3 Oxygen vacancies

The defect energies for oxygen vacancies within the electronic structure methods cannot be directly calculated. This is because the energy of the isolated $O^{2-}$ ion cannot be evaluated, as the second electron is not bound to the nucleus (it is only stabilised in the solid state by the confining potential of the surrounding cations). It is, however, possible to calculate the difference between the energy needed to produce a vacancy on each of the oxygen sites, presented in table 4.10. There is excellent agreement between the methods for this value. The inequivalence between the three sites has important implications for diffusion as discussed in chapter 5.

The structure of a vacancy on the O3 site calculated using the embedded cluster method is shown in figure 4.4. The major observation is that there is relatively little relaxation around the defect and that the central silicate unit maintains its tetrahedral shape, despite the loss of one of the oxygen atoms. This is also observed for the less favourable O2 vacancy. However, in all cases where the Mott-Littleton method is used, the Si ion moves towards the three remaining oxygen ions to form a near planar SiO$_3$ unit. Although this highlights an error with the potential energy functions the fact that the defect energies are similar is reassuring because it holds open the possibility of using the potential model for extended calculations.
Table 4.10: Energy differences between point defects on the O2 and O3 sites in forsterite calculated using the embedded cluster and Mott-Littleton method and the results of periodic DFT calculations using the GGA PW91 functional.

<table>
<thead>
<tr>
<th>Method</th>
<th>Defect energy difference between O2 and O3 (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mott-Littleton method</td>
<td>0.66</td>
</tr>
<tr>
<td>Periodic DFT (Brodholt, 1997)</td>
<td>1.07</td>
</tr>
<tr>
<td>HF/6-31G//HF/6-31G</td>
<td>0.53</td>
</tr>
<tr>
<td>PW91/6-31G//PW91/6-31G</td>
<td>0.46</td>
</tr>
<tr>
<td>B3LYP/6-31G//B3LYP/6-31G</td>
<td>0.50</td>
</tr>
</tbody>
</table>

(a) See section 4.5.1.1 for a discussion of the discrepancy between the current data and the data from Brodholt (1997).

Figure 4.4: structure of O3 vacancy calculated using embedded cluster methodology looking along [010] with [100] oriented up the page. Note that the central SiO$_3$ unit preserves a tetrahedral shape with the Si ion moving towards the oxygen vacancy, this is in contrast to the more planar arrangement observed in the Mott-Littleton calculations with the Si moving away from the vacancy.
4.4.2 Interstitial defects

In practical terms, vacancies are created by simply removing the relevant ion from the simulation cell while more effort is required to establish the structure of interstitials. The presence of interstitial oxygen ions is indicated from a range of experimental data (e.g. Gérard and Jaoul, 1989; Ryerson et al., 1989) described in chapter 5. It seemed likely that the relatively large oxygen ions could only occupy vacant octahedral sites in the olivine structure so interstitial defects were placed in a number of locations in and around the I1 and I2 site. Upon geometry optimisation of the resulting Mott-Littleton models, two clear energy minima were located, one close to the centre of each site.

In the case of interstitial magnesium, location of energy minima was less straightforward. Initial calculations with interstitial magnesium ions in either of the two vacant octahedral sites resulted in very large ionic displacements on relaxation and final defect energies that were very sensitive to the initial geometry. This is an indication of a failure of the geometry optimisation procedure, probably due to a starting configuration away from an energy minimum. In order to locate the minimum energy configuration for a magnesium interstitial defect a large number of starting geometries were created by placing interstitial magnesium ions on a regular 0.5 angstrom grid across the symmetry irreducible portion of the unit cell and performing a defect optimisation using the large condor pool at University College London. Following removal of unphysical structures where the coulombic attraction between oxygen ions and inappropriately located magnesium ions had overcome the short range repulsion leading to very large negative energies, the lowest energy structures were examined. The low energy configurations were all various split interstitial configurations where the interstitial ion and a displaced lattice magnesium ion were located close to opposite faces of an M1 or M2 octahedron. Embedded cluster calculations (see below and appendix 2)

---

15 Condor is a system to utilise large numbers of commodity computers as a distributed high throughput computational resource when the computer is not used for its primary purpose. Developed at The University of Wisconsin (http://www.cs.wisc.edu/condor/) the UCL pool consists of approximately 1000 teaching PCs and is an ideal resource allowing large number of small calculations to be quickly performed.
show a difference in energy between an octahedrally co-ordinated magnesium interstitial and the split interstitial indicates that the split interstitial is 4.4 eV more stable than the octahedral interstitial on the 11 site.

The lowest energy site was chosen for further investigation and Mott-Littleton and embedded cluster calculations were set up with the structure (with two interstitial ions and a vacancy) as input. Details of this structure are shown in figure 4.5, the two magnesium ions form a split interstitial across the M1 site orientated in the [010] direction with each magnesium ion in distorted tetrahedral co-ordination. The Mg – O bond distances in this structure are similar to those found in crystals with structural magnesium tetrahedrally co-ordinated by oxygen. For example in a recently synthesised bismuth magnesium vanadate (Uma and Sleight, 2002) tetrahedral Mg – O bonds are ~1.95 angstroms long, in this structure the bonds are 1.89, 2.02, 1.96 and 1.86 angstroms long for the Mg – O3a, Mg – O1, Mg – O3b and Mg – O2 bonds respectively.

**Figure 4.5:** Structure of magnesium split interstitial defect from embedded cluster calculations. (a) Looking along [010] with [100] oriented up the page. (b) Looking along [100] with [001] oriented up the page. The two tetrahedral magnesium ions are shown in black, otherwise the key is identical to that for figure 4.2.

---

16 O3a, O1 and O2 refer to oxygen sites co-ordinating the M1 forming one of the triangular faces of the M1 octahedron while the O3b refers to a oxygen site in the second O shell from the M1 site.
The energetics of the magnesium split interstitial defect calculated using the QM/MM method and shown in table 4.11, follow the opposite pattern to the magnesium vacancy and are in excellent agreement with the Mott-Littleton methodology. The HF approximation predicts defect energies almost 1 eV higher than DFT with B3LYP being in excellent agreement with DFT in this case. The Mott-Littleton results fall between the HF and DFT calculated energies. Figure 4.6 shows the convergence of the defect energy with basis set. Convergence is not as good as in the case of the magnesium vacancy (perhaps due to the partial occupation of d-orbital not represented in the smaller basis sets), but in any case the change in energy from the 6-31+G*/6-31-G to 6-311+G*/6-311+G* is only about 0.1 eV.

![Figure 4.6](image)

**Figure 4.6:** Convergence in defect energy of the magnesium split interstitial in forsterite with basis set size. See caption of figure 4.2 for more information.

The minimum energy for the incorporation of a silicon interstitial was found in exactly the same way as the magnesium interstitial. The lowest energy configuration was found to be with the silicon interstitial on the mirror plane in tetrahedral co-ordination close to the I2 site, the interstitial forms 4 bonds of about 1.7 Å to oxygen atoms already bonded to three silicon atoms and all Si–O–Si bridging is through O3 oxygens. The Si–O bond lengths are rather
long, perhaps suggesting a breakdown of the potential model, future embedded cluster calculations on this structure are required.

**Table 4.11:** Defect energies for the magnesium split interstitial across the Mg1 site, see the caption to figure 4.2 for the meaning of the basis set symbols.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>Defect energy (eV)</th>
<th>Hartree-Fock</th>
<th>DFT (PW91)</th>
<th>DFT (B3LYP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-31G//6-31G</td>
<td>-18.029</td>
<td>-18.470</td>
<td>-18.478</td>
<td></td>
</tr>
<tr>
<td>6-31+G*//6-31G</td>
<td>-17.366</td>
<td>-17.952</td>
<td>-17.947</td>
<td></td>
</tr>
<tr>
<td>6-31+G*//6-31+G*</td>
<td>-17.403</td>
<td>-18.001</td>
<td>-17.993</td>
<td></td>
</tr>
<tr>
<td>6-311+G*//6-31+G*</td>
<td>-17.503</td>
<td>-18.045</td>
<td>-18.038</td>
<td></td>
</tr>
<tr>
<td>6-311+G*//6-311+G*</td>
<td>-17.508</td>
<td>-18.049</td>
<td>-18.043</td>
<td></td>
</tr>
</tbody>
</table>

**4.4.3 Hydrous defects**

Defects involving hydrogen in normally anhydrous minerals (NAMs), such as forsterite, are of great interest to the geophysicist and provide some interesting problems for modelling. If a model based on interatomic potential functions is used, it is difficult to see where the potentials can come from (there are rather few experimental determinations of the structure of hydrated point defects). On the other hand, if electronic structure methods are used then there is the challenge of describing hydrogen bonding.

One of the major questions is the mechanism of hydrogen incorporation, it is known that hydrogen is not adsorbed as molecular water because of the absence of the characteristic H-O-H bending mode in spectroscopic measurements. Infrared active O-H stretching modes between 3300 – 3700 cm\(^{-1}\) are however observed in all NAMs and can be used to estimate hydrogen content. In detail the IR-spectra of olivine is rather complex (e.g. Demouchy and Mackwell, 2003) with many distinct O-H stretching modes, this suggests that hydrogen is incorporated in a number of distinct ways, usually thought of as hydroxyl groups closely associated with point defects with negative effective charge – metal vacancies are usually postulated. This suggests a mechanism for the experimentally observed “water weakening” described in chapter 2, if the
association with hydroxyl groups lowers the energy needed to introduce metal vacancies then this should increase the number of metal vacancies in wet, compared to dry, olivine, which will decrease its viscosity. To determine if hydrogen – metal vacancy association is energetically feasible a large number of calculations have been carried out with hydrogen bound to metal vacancies and hydrogen isolated in an otherwise perfect crystal.

4.4.3.1 Mott-Littleton calculations

In the potential model used for the Mott-Littleton method, O-H bonds are described using a Coulomb subtracted Morse potential of the form given in equation 3.18. Initially a model previously used by Wright et al. (1994) was used. These potentials use fractional charges for the O and the H in the OH⁻ ion, but the hydroxyl has a formal charge of -1. In this description, hydrogen bonding is not directly represented (although the Buckingham potentials contain an interionic attraction between non bonded hydrogen and oxygen ions). The parameters were fitted to quantum-chemical data by Saul et al. (1985) and modified by Schröder et al. (1992) accounting for the decrease in charge on the oxygen ion. These potentials were found to give unreasonable calculated frequencies for the IR active O-H stretching mode. To avoid this problem Gatzemeier et al. (In preparation) recently re-parameterised the model in order to reproduce the O-H stretching modes of the \( \left\{ V_{\text{Mg}}^{2}OH_{\text{O2}} \right\}^{*} \) defect as calculated by Braithwaite et al. (2003). This model is used here; the parameters of the potential model are given in table 4.1.

Initial calculations of the energy of hydrous defects in forsterite were undertaken using the Mott-Littleton method. Five types of environment have been considered: (i) hydrogen bonded to lattice oxygen in an otherwise defect free crystal of forsterite, hydrogen bonded to (ii) interstitial oxygen ions and hydrogen bonded to oxygen coordinating a vacancy on the (iii) Mg1 site, the (iv) Mg2 site and the (v) Si site. The point symmetry of the three metal sites has been used to reduce the number of calculations needed. For the octahedral Mg1 site, which lies on an inversion centre, there are only three distinct ways of arranging a single O-H group and six distinct ways of arranging two O-H groups. The Si site has only three ways of arranging one O-H group (the two
O3 sites are identical), four ways of arranging two O-H groups, three ways of arranging three O-H groups and only one way of fully protonating the vacancy to form a “hydrogarnet” substitution or “hydroxyl nest”. The Mg2 site is more complex and is co-ordinated by one O1 oxygen, one O2 oxygen and four O3 oxygens, the four O3 oxygens can be separated into two pairs, the first pair (O3a) are bonded to a common silicon ion while the second pair (O3b) share a silicon atom with another O3 group that does not form part of the Mg2 octahedra. This means there are a total of four distinct ways of arranging the first O-H group and eleven distinct ways of arranging the second. Because of this it is particularly important to consider the four ways of arranging two O-H groups on O3 sites, there are two distinct ways to arrange one O-H on a O3a site and a second on an O3b site, the first arrangement has the two O-H groups on corners of one of the triangular faces of the octahedra (the “face” configuration) while the second has the two O-H groups on opposite corners of the octahedra (the “cross” configuration). The 38 ways of locating OH groups in forsterite, along with the energies and IR active OH stretching frequencies calculated using the Mott-Littleton method are reported in table 4.12. A small number of these configurations fail to converge to a stable minimum despite various optimisation procedures being used. The reason for this convergence failure was investigated and discovered to be due to the unphysical dissociation of OH groups during the relaxation. So far these structures have not been considered further.
Table 4.12: Calculated OH defect energies using the Mott-Littleton method.

<table>
<thead>
<tr>
<th>Defect</th>
<th>Defect energy (eV)</th>
<th>Calculated OH stretching modes (cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( OH'_{O1} )</td>
<td>21.63</td>
<td>4580</td>
</tr>
<tr>
<td>( OH'_{O2} )</td>
<td>16.83</td>
<td>3435</td>
</tr>
<tr>
<td>( OH'_{O3} )</td>
<td>16.54</td>
<td>3569</td>
</tr>
<tr>
<td>( OH'_{(1)} )</td>
<td>-0.39</td>
<td>3563</td>
</tr>
<tr>
<td>( OH'_{(2)} )</td>
<td>-2.64</td>
<td>3744</td>
</tr>
<tr>
<td>{( V_{Mg}OH_{O0} )}</td>
<td>40.94</td>
<td>3525</td>
</tr>
<tr>
<td>{( V_{Mg}OH_{O2} )}</td>
<td>38.74</td>
<td>3380</td>
</tr>
<tr>
<td>{( V_{Mg}OH_{O3} )}</td>
<td>38.84</td>
<td>3419</td>
</tr>
<tr>
<td>{( V_{Mg}OH_{O0}_OH_{O0} )}</td>
<td>55.94</td>
<td>3387, 3386</td>
</tr>
<tr>
<td>{( V_{Mg}OH_{O0}_OH_{O2} )}</td>
<td>54.35</td>
<td>3389, 3323</td>
</tr>
<tr>
<td>{( V_{Mg}OH_{O0}_OH_{O3} )}</td>
<td>54.92</td>
<td>3586, 3313</td>
</tr>
<tr>
<td>{( V_{Mg}OH_{O2}_OH_{O1} )}</td>
<td>54.13</td>
<td>3415</td>
</tr>
<tr>
<td>{( V_{Mg}OH_{O3}_OH_{O1} )}</td>
<td>56.49</td>
<td>3094, 3523</td>
</tr>
<tr>
<td>{( V_{Mg}_OH_{O0}_OH_{O1} )}</td>
<td>42.51</td>
<td>3456</td>
</tr>
<tr>
<td>{( V_{Mg}_OH_{O0}_OH_{O2} )}</td>
<td>40.49</td>
<td>3338</td>
</tr>
<tr>
<td>{( V_{Mg}_OH_{O0}_OH_{O3} )}</td>
<td>38.94</td>
<td>3400, 3470</td>
</tr>
<tr>
<td>{( V_{Mg}_OH_{O2}_OH_{O0} )}</td>
<td>58.64</td>
<td>3263, 3443</td>
</tr>
<tr>
<td>{( V_{Mg}_OH_{O2}_OH_{O2} )}</td>
<td>56.35</td>
<td>3367, 3252</td>
</tr>
<tr>
<td>{( V_{Mg}_OH_{O2}_OH_{O3} )}</td>
<td>56.12</td>
<td>3460, 3270</td>
</tr>
<tr>
<td>{( V_{Mg}_OH_{O3}_OH_{O0} )}</td>
<td>55.09</td>
<td>3445, 3519</td>
</tr>
<tr>
<td>{( V_{Mg}_OH_{O3}_OH_{O2} )}</td>
<td>55.08</td>
<td>3483, 3420</td>
</tr>
<tr>
<td>{( V_{Mg}_OH_{O3}_OH_{O3} )}</td>
<td>56.17</td>
<td>3308, 3297</td>
</tr>
<tr>
<td>{( V_{Mg_OH_{O0}_OH_{O3} )}</td>
<td>59.95</td>
<td>3212, 3701</td>
</tr>
<tr>
<td>{( V_{Mg_OH_{O2}_OH_{O3} )}</td>
<td>54.97</td>
<td>3476, 3372</td>
</tr>
<tr>
<td>{( V_{Mg_OH_{O3}_OH_{O3} )}</td>
<td>61.97</td>
<td>3877, 3874</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O0} )}</td>
<td>112.19</td>
<td>2909</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O1} )}</td>
<td>112.73</td>
<td>3558</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O2} )}</td>
<td>112.92</td>
<td>3434</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O0}_OH_{O2} )}</td>
<td>124.30</td>
<td>3513, 3583</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O0}_OH_{O3} )}</td>
<td>124.50</td>
<td>3460, 3683</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O2}_OH_{O3} )}</td>
<td>124.36</td>
<td>3459, 3398</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O3}_OH_{O3} )}</td>
<td>123.92</td>
<td>3405, 3454</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O0}_OH_{O0}_OH_{O1} )}</td>
<td>136.49</td>
<td>3572, 3645, 3483</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O0}_OH_{O2}_OH_{O1} )}</td>
<td>136.33</td>
<td>3669, 3615, 3517</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O2}_OH_{O0}_OH_{O1} )}</td>
<td>136.26</td>
<td>3497, 3441, 3422</td>
</tr>
<tr>
<td>{( V_{Ca_OH_{O2}_OH_{O2}_OH_{O1} )}</td>
<td>149.91</td>
<td>3623, 3664, 3674, 3702</td>
</tr>
</tbody>
</table>

(a) These configurations repeatedly fail to converge to a reasonable minimum (see text), (b) is the "cross" configuration and (c) is the "face" configuration.
4.4.3.2 Embedded-cluster calculations

Electronic structure calculations of hydrogen defects in forsterite have previously been undertaken by Brodholt and Refson (2000) and Braithwaite et al. (2003). This work takes the study further by considering the effect of the use of the hybrid B3LYP functional. It is not practical to consider all of the 38 conformations using the embedded cluster approach, instead only four low energy configurations \((OH_{O3})^\times, \{V_{Mg}\}OH_{O2}\}^\times, \{V_{Mg}OH_{O2}OH_{O2}\}^\times\), and \{V_{Si}OH_{O3}OH_{O3}OH_{O3}\}^\times\), see figures 4.7 and 4.8) were selected for study. A few other arrangements have been studied by Braithwaite et al. (2003), here calculations using the B3LYP functional and an analysis of basis set effects are presented.

Figure 4.7: Mg centred QM cluster looking along [010] with [001] orientated up the page and [100] across the page containing a \(\{V_{Mg}\}OH_{O2}\}^\times\) (part a) and a \(\{V_{Mg}2OH_{O3}\}^\times\) defect complex (part b).
Figure 4.8: Defects in the larger QM cluster. Part a shows the $OH_{O3}^*$ defect looking along [001] with [100] aligned in the vertical, the OH dipole lies in the (100) plane. Part b shows the $\{V_{s}OH_{O2}OH_{O3}OH_{O3}\}$ defect cluster in the same orientation. O1 (on the right hand side), O2 (top) and one of the O3 ions have OH dipoles along the edges of the vacant tetrahedra while the second O3 OH dipole points out of the tetrahedra.

Experimentally, the principal method for the identification of OH defects is infrared spectroscopy and it is possible to calculate the IR spectra of likely defects providing a route to make contact between experiment and simulation. This technique has recently been successfully applied by Braithwaite et al. (2003) whose suggestion that IR peaks at $\sim 3200 \text{ cm}^{-1}$ were due to hydroxyl associated with silicon vacancies and peaks at $\sim 3600 \text{ cm}^{-1}$ are associated with magnesium vacancies was supported by a recent experimental study (Lemaire et al., 2004). In order to extend the study of Braithwaite et al. (2003) further calculations have been undertaken where the number of atoms included in the calculation of the dynamical matrix (the “active atoms”) is increased from only the hydrogen atoms. This will only have an effect if the off diagonal elements in the dynamical matrix are large, in other words if moving one atom a short distance exerts a force on another (in this case a hydrogen) atom. The results indicate that only the oxygen atoms bonded to the hydrogen atoms behave in
this way and further increase in the number of active atoms has a very small effect. It is also worth noting that the change in calculated frequency accounts for the majority of the disagreement between experiment and calculation. Ideally, frequency calculations would be carried out for all clusters but this is not possible due to computational constraints. In any case, it seems that these calculations give further weight to the assignments of OH frequencies described by Matveev et al. (2001), Braithwaite et al. (2003) and Lemaire et al. (2004).

Table 4.13: Effect of changing the number of atoms in the active region on calculated IR active OH stretching modes.

<table>
<thead>
<tr>
<th>Active atoms</th>
<th>Calculated frequencies (cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only hydrogen atoms (2)</td>
<td>3141, 3150</td>
</tr>
<tr>
<td>Hydroxyl groups (4)</td>
<td>3235, 3243</td>
</tr>
<tr>
<td>All atoms in QM cluster (24)</td>
<td>3236, 3243</td>
</tr>
</tbody>
</table>

Analysis of the energetics of hydrogen incorporation

Analysis of the defect energies presented in table 4.12 can be undertaken in several ways. The first approach is to calculate the energy associated with forming the point defects by reactions with water. The positively charged hydrogen must be charge balanced by defects with net negative charge. In pure forsterite these are likely to be cation vacancies, so the reactions are:

\[
\text{Si}_i^* + 2H_2O \rightarrow \text{SiO}_i^* + V_{Si}^* + 4(OH)_o^*,
\]

\[
E = 4E(OH_{o3}^*) + E(V_{Si}^*) + U_{SiO} + 2E_{pt} = +18.83 \text{ eV }, \tag{4.4}
\]

for the formation of silicon vacancies and:

\[
Mg_{Si}^* + H_2O \rightarrow MgO + V_{Mg}^* + 2(OH)_o^*,
\]

\[
E = 2E(OH_{o3}^*) + E(V_{Mg}^*) + U_{MgO} + E_{pt} = +6.51 \text{ eV } \tag{4.5}
\]

17 The computational resources needed to calculate the dynamical matrix scale with the cube of the number of atoms -- and for the larger cluster, a single atom calculation requires approximately 1 week of CPU time. In principle this problem could be over come by utilising HPCx, the UK’s large scale research computing resource but, unfortunately this is not possible because of a incompatibility between the current version of IBM’s Fortran compiler and the GAUSSIAN code. This problem has been reported to IBM and GAUSSIAN inc. by the management of HPCx and a solution is expected to be included in the next version of IBM’s Fortran compiler, which is currently under development.
for the formation of magnesium vacancies. $E_{pt}$ is defined as the half of the energy of the gas phase reaction:

$$H_2O_{(g)} + O^{2-}_{(g)} \rightarrow 2OH^{-}_{(g)}, \quad (4.6)$$

which was estimated by Wright et al. (1994) to be $-9.74$ eV (see table 4.15).

The hydrogen need not be isolated and could be closely bound to the vacancies with the resulting equations:

$$Si^{x+}_{Si} + 2H_2O \rightarrow SiO_2 + \{V_{Si}, 4(OH)_{O}\}^x$$

$$E = 2\{V_{Si}, 4(OH)_{O}\}^x + U_{SiO_2} + 2E_{pt} = 1.77 \text{ eV} \quad (4.7)$$

and:

$$Mg_{Mg}^{x+} + H_2O \rightarrow MgO + \{V_{Mg}, 2(OH)_{O}\}^x$$

$$E = \{V_{Mg}, 2(OH)_{O}\}^x + U_{MgO} + E_{pt} = 3.08 \text{ eV} \quad (4.8)$$

It is perhaps most meaningful to calculate the binding energy between the hydrogen ion and the point defect, which is just defined as the difference between, for example, equation 4.4 and 4.6, or equivalently by reactions like:

$$V_{Si}^{x+} + 4(OH)_{O} \rightarrow \{V_{Si}, 4(OH)_{O}\}^x$$

Use of the binding energy conveniently removes the energy of the water molecule and solid phase. This removes a major source of error, particularly for the embedded cluster results where the energy of MgO and SiO$_2$ is very hard to estimate (see Braithwaite et al., 2003). The binding energies are presented in table 4.14.
Table 4.14: Binding energies (eV) between hydrogen and cation vacancies in forsterite.

<table>
<thead>
<tr>
<th>Method</th>
<th>$V_{n} + {\text{OH}}<em>{o}$ $\rightarrow {n</em>{o} \text{OH}}_{o}$</th>
<th>$V_{n} + 2{\text{OH}}<em>{o}$ $\rightarrow {2n</em>{o} \text{OH}}_{o}$</th>
<th>$V_{n} + 4{\text{OH}}<em>{o}$ $\rightarrow {4n</em>{o} \text{OH}}_{o}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mott-Littleton</td>
<td>-2.28</td>
<td>-3.43</td>
<td>-17.06</td>
</tr>
<tr>
<td>HF/6-31G //HF/6-31G</td>
<td>-5.46</td>
<td>-2.14</td>
<td>-21.96</td>
</tr>
<tr>
<td>HF/6-31G //HF/6-31G+G*</td>
<td>-2.62</td>
<td>-2.11</td>
<td>-23.13</td>
</tr>
<tr>
<td>PW91/6-31G //PW91/6-31G</td>
<td>-3.04</td>
<td>-2.60</td>
<td>-21.56</td>
</tr>
<tr>
<td>PW91/6-31G //PW91/6-31G+G*</td>
<td>-2.74</td>
<td>-2.34</td>
<td>-21.90</td>
</tr>
<tr>
<td>B3LYP/6-31G //B3LYP/6-31G</td>
<td>-2.93</td>
<td>-2.48</td>
<td>-21.83</td>
</tr>
<tr>
<td>B3LYP/6-31G //B3LYP/6-31G+G*</td>
<td>-2.70</td>
<td>-2.25</td>
<td>-22.55</td>
</tr>
</tbody>
</table>

These equations show that hydrogen will tend to form various hydroxyl - defect associates (at least if the temperature is low enough and concentration high enough that entropy can be neglected). This is simply a reflection of the decrease in energy of the polarization of the lattice at long range when the net charge on the point defects is reduced. The extent of this binding, and the tendency for the OH groups to be found on their preferred position will depend on a wide range of variables not considered in these calculations, entropic effects have already been noted but the tendency to form silicon and magnesium vacancies will depend on the external chemistry (enstatite activity) as well as the presence of other defects, including iron (which is expected to be important, see Wright and Catlow, 1994, 1996). It is perhaps more promising in these situations to attempt to use the calculations to “fingerprint” particular defect clusters rather than make predictions based on the energetics of rather idealised reactions.

4.4.4 Iron

Finally, the incorporation of iron(II) and iron(III) onto the M1 and M2 sites is considered using the Mott-Littleton method (see table 4.7). Iron, in both oxidation states, is found to lie preferentially on the M2 site. Electron paramagnetic resonance (EPR) spectroscopy provides experimental support for this result, which is for a dilute system. The spectroscopic studies suggest that
iron III in forsterite partitions into the Mg2 site at low total iron concentrations while at higher concentration iron(III) is found on both the silicon and Mg2 sites in a 60:40 ratio (Chatelain and Weeks, 1971; Niebuhr, 1975; Gaite and Hafner, 1984; Gaite and Ranger, 1997).

4.5 Modelling point defects - discussion

The defect energies presented in the preceding sections are the internal energy contribution needed to remove an ion from the lattice to the gas phase to form a vacancy and to bring an ion from the gas phase into the lattice to form an interstitial. The energies are therefore of little relevance to reactions under geophysical conditions but it is possible to use the energies to construct energetic cycles to derive the enthalpy\(^\text{18}\) of relevant defect reactions. This section concentrates on deriving the enthalpies of relevant reactions based on the work of Smyth and Stocker (Smyth and Stocker, 1975; Stocker and Smyth, 1977). A second issue is that the structures of some of the defects are of interest, this is particularly true of the hydroxyl defects where the calculations can be used to attempt to understand hydrogen speciation in mantle minerals.

Before considering the defect reactions, it is well worth considering the accuracy of the calculations. As far as I am aware, there is no experimental data that directly constrains the defect thermodynamics although some studies reported below give important insights. Instead of considering agreement with experiment, the consistency of the present results with previous studies will be considered. The results from the potential model presented in tables 4.7 and 4.12 are in excellent agreement with previous work using the same potential model, this includes work that utilised the super-cell approach (Richmond and Brodholt, 2000) as well as the Mott-Littleton method used here (Wright and Catlow, 1994; Jaoul et al., 1995). This is hardly surprising — indeed disagreement would have suggested errors in one or more of the codes used to do the calculations. An additional test of the accuracy of the potentials based

---

\(^{18}\) The energies calculated are strictly the internal energy of the reaction excluding zero point motion, by definition this is the same as the enthalpy at zero Kelvin (still neglecting the zero point motion). As enthalpy is only weakly dependent on temperature this value can be used as an approximation for the enthalpic contribution to the finite temperature free energy of the reaction.
defect energetics is comparison with the results of electronic structure calculations. In general, as was pointed out in the results sections, good agreement with such calculations is observed, in particular the defect formation energies calculated using the embedded cluster methodology by Braithwaite et al. (2002; 2003), as well as the additional results presented here, are in general agreement with the calculated atomistic values. Discrepancies between results obtained using the potential model and electronic structure methods, and between results from the embedded cluster method and periodic DFT calculations, have been noted by Braithwaite et al. (2003) and Brodholt (1997). The largest error is associated with the formation of a vacancy on the silicon site. The first reason for this error is likely to be that the potential model is unable to describe the resultant five co-ordinate silicate species. A second consideration is that the charge on the silicon vacancy is the largest considered – resulting in the largest correction terms for the long-range polarization and the largest electronic polarisation, especially of the oxygen ions. It is possible that the basis set is not sufficient for modelling this polarisation. Although some of this difference can be attributed to the small size of the super-cell it likely that an additional factor is caused by the way electronic polarisation around the defect is treated in the atomistic and density functional calculations. In the DFT study it is likely that the polarisation is under-estimated around the highly charged silicon vacancy because of an inadequate plane wave basis that was only converged with respect to bulk olivine (this would destabilise the defect, as it would have a larger effective charge).

Some additional energies are needed in order to consider the defect reactions, these include the enthalpy of formation of a number of minerals that will be the source or sink of the ions from the defect and a number of other standard energies. These are given in table 4.15, with formation energies calculated using the same interatomic potential model used to calculate the defect structures and energies. These energies are not available for comparisons with the embedded cluster calculations because of the difficulty in producing results for a periodic solid using a computer code designed for studies of finite clusters (see Braithwaite et al., 2002).
Table 4.15: Lattice energies of a range of crystal structures calculated using the interatomic potential model presented in table 4.1. Energies are for formation of infinite crystal from isolated ions.

<table>
<thead>
<tr>
<th>Crystal Structure</th>
<th>Energy p.f.u (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mg$_2$SiO$_4$ (forsterite)</td>
<td>-212.49</td>
</tr>
<tr>
<td>Fe$_2$SiO$_4$ (fayalite)</td>
<td>-211.24</td>
</tr>
<tr>
<td>MgSiO$_3$ (enstatite)</td>
<td>-170.97</td>
</tr>
<tr>
<td>Fe$_2$O$_3$ (hematite)</td>
<td>-124.83</td>
</tr>
<tr>
<td>SiO$_2$ (α-quartz)</td>
<td>-128.66</td>
</tr>
<tr>
<td>MgO (periclase)</td>
<td>-41.31</td>
</tr>
</tbody>
</table>

(a) $E_{OD}$ is the dissociation energy of an oxygen molecule, $E_{A2}$ is the sum of the first and second electron affinities of oxygen and $E_{Fe(III)}$ is the third ionisation energy of iron. Data marked (b) are from the CRC handbook, (c) is from Wright et al. (1994) and (d) is from Catlow (1977a).

### 4.5.1 Intrinsic defects in forsterite

Following the development of Smyth and Stocker (1975), the possible defect reactions in a closed system containing forsterite are considered, followed by possible reactions in an open system and those that involve the formation of free electrons or electron holes. For convenience, any electron holes are assumed to become localised on lattice iron, oxidising it to iron(III) and any free electrons reduce iron(III) back to iron(II). It has been argued that such a procedure is justified (Smyth and Stocker, 1975) given the large band gap in forsterite (Smyth and Stocker quote an approximate value of 8.8 eV) and the smaller energy required to trap an electron hole on iron on the Mg$_2$ site (8.38 eV using the data in tables 4.7 and 4.15). Calculations of the energetics, structure and properties of other defects with trapped electrons are possible using the embedded cluster model and is the subject of ongoing investigation. However, the defect reactions using the calculated defect energies from the potentials based approach are presented and, where appropriate, indicate the effect of using the more accurate but sparser embedded cluster results.
4.5.1.1 Closed system

As discussed in chapter two, crystals at thermodynamic equilibrium contain a number of point defects because the entropy gained in forming the defects outweighs the energetic penalty of forming the defect. For simple uncharged defects the defect concentration at a given temperature can be calculated in a straightforward manner by explicitly writing the free energy change in terms of the enthalpy of the formation of the point defect and the configurational entropy gained as a function of defect concentration and finding the minimum free energy to give the equilibrium defect concentration. Such a procedure is much more complex in multi-component ionic systems because there are a range of possible defect types, in principle one should minimise the free energy numerically taking into account the enthalpic and entropic contribution from all possible defect species under an imposed condition of charge neutrality (Ashcroft and Mermin, 1976). The first stage requires the calculation of possible reactions resulting in the formation of intrinsic defects. This is undertaken here. Using the results of the Mott-Littleton calculations gives the energy of a full Schottky defect where a full formula unit of forsterite vacancies is formed and the ions are moved to the surface as:

\[ \text{Si}^{2+}_\text{Si} + 2\text{Mg}^{2+}_{\text{Mg}} + 4\text{O}^{2-}_{\text{O}_3} \rightarrow \text{V}^{\text{Si}^{2+}} + 2\text{V}^{\text{Mg}^{2+}} + 4\text{V}^{\text{O}^{2-}} + \text{Mg}_2\text{SiO}_4 \]

\[ E(\text{Schottky}) = E(V^{\text{Si}^{2+}}) + 2E(V^{\text{Mg}^{2+}}) + 4E(V^{\text{O}^{2-}}) + U_{\text{Mg}_2\text{SiO}_4} = 35.44 \text{ eV} \]  \hspace{1cm} (4.9)

or 5.06 eV per defect. This can be compared with a value of 30.25eV given by GGA calculated using a super-cell containing 56 atoms (Brodholt, 1997). Additional calculations using a fully converged super cell results with the potential model results in better agreement with the Mott-Littleton calculations (36.4 eV) and decreasing the size of the super cell will tend to reduce\(^9\) this value explaining the discrepancy.

---

\(^9\) This is counterintuitive but correct. It may be expected that a super cell that is too small would give energies of individual defects that are too high because the cell constrains the relaxation around the defect. This is what is seen for the Mot-Littleton calculation, and in general for many types of calculation. For charged vacancies another effect predominates, the reduction in the electrostatic potential on the site of the defect due to the absence of ions in the adjoining cells. This tends to reduce the energy of the smaller super cells.
The second major type of intrinsic defect is the Frenkel defect where a vacancy is charge balanced by an interstitial of its own type. In principle, Frenkel defects can form on any of the three sublattices. The oxygen Frenkel:

\[ O_{O3}^e \rightarrow V_{O3}^o + O_{12}^o, \]

\[ E(O-Frenkel) = E(O_{12}^o) + E(V_{O3}^o) = 8.43 \text{ eV}, \]  

(4.10)
gives a defect energy of 4.22 eV per defect formed, the magnesium Frenkel:

\[ Mg_{Mg}^e \rightarrow V_{Mg}^o + Mg_{Mg-split}^o, \]

\[ E(Mg-Frenkel) = E(V_{Mg}^o) + E(Mg_{Mg-split}^o) = 6.73 \text{ eV}, \]  

(4.11)
yields an energy of 3.37 eV per defect, the silicon Frenkel defect

\[ Mg_{Si}^e \rightarrow V_{Si}^o + Si_{Si}^{***}, \]

\[ E(Si-Frenkel) = E(V_{Si}^o) + E(Si_{Si}^{***}) = 24.21 \text{ eV}, \]  

(4.12)
or 12.10 eV per defect. Clearly the Mg Frenkel defect will be the predominant intrinsic defect, in agreement with the suggestion of Smyth and Stocker (1975), but this does not rule out the possibility of other intrinsic defects (indeed they are required to minimise the free energy).

5.4.1.2 Open system

The second class of defect reactions considered by Smyth and Stocker (1975) were those which change the stoichiometry of the olivine crystal, these may still be considered intrinsic defects but in assessing the free energy change of the defect reactions account should be taken of the external chemical potential. To completely define the thermodynamic state it is necessary to specify the temperature, pressure, and the activities of two components, usually these are the oxygen and enstatite activities. The defect reactions result in a deviation from stoichiometry, which can be usefully expressed as an excess or deficiency in silicon content:

\[ \xi = \frac{n_{Si}}{n_{Si} + n_{Mg} + n_{Fe}} - \frac{1}{3}, \]  

(4.13)
or an excess or deficiency in oxygen content:

\[
\overline{\eta} = \frac{n_O}{n_{Si} + n_{Mg} + n_{Fe}} - \frac{4}{3},
\]

where \( n_X \) is the abundance of element \( x \) and deviation of \( \xi \) or \( \eta \) from zero indicates non-stoichiometry (see Nakamura and Schmalzried, 1983; Tsai and Dieckmann, 2002).

Stocker and Smyth (1977) considered a range of majority defect types and arrived at equations for the change in defect concentration with oxygen partial pressure and enstatite activity. These relationships are of direct use for analysing the effect of chemistry on point defect diffusion (because the flux of the diffusing species is proportional to the concentration of point defects) and to the interpretation of thermogravimetric studies (Tsai and Dieckmann, 1997). Such studies were undertaken by Nakamura and Schmalzried (1983) and later by Tsai and Dieckmann (2002).

Smyth and Stocker (1975) considered a series of reactions based on olivine dissolving or precipitating MgO or SiO\(_2\). Rather than evaluating the energies of these reactions the energetics of reactions involving enstatite are considered. These reactions are more realistic of the upper mantle where neither MgO nor SiO\(_2\) are stable. The first possible reaction is likely in regions of silicon excess:

\[
Mg_2SiO_4 + 2MgSiO_3 \rightarrow 3Mg_2SiO_4 + 2V_{Mg}^- + 2V_O^+.
\]

\[E = 2U_{Mg_2SiO_4} + 2E(V_{Mg}^-) + 2E(V_O^+) - 2U_{MgSiO_3} = 15.00 \text{ eV}.\]  

(4.15)

A second reaction is also possible involving the formation of defects and precipitation of enstatite; this is particularly likely under conditions of silicon deficiency. The reaction for this is:

\[
Si_{Sl}^+ + 2Mg_{Sl}^+ + 3O_{Sl}^- \rightarrow V_{Sl}^+ + 2V_{Mg}^- + 3V_{O3}^- + Mg_2SiO_3,
\]

\[E(Mg_2SiO_3 - \text{Schottky}) = E(V_{Sl}^+) + 2E(V_{Mg}^-) + 3E(V_{O3}^-) + U_{Mg_2SiO_3} = 27.94 \text{ eV}.\]  

(4.16)

At high oxygen partial pressure, in the oxygen excess region it is possible to add oxygen to the forsterite lattice. Two electrons from the forsterite crystal are needed to allow the oxygen molecule to dissociate into a pair of \( O^2^- \) ions which
then form interstitial species. This may be mediated by the oxidation of iron(II) to iron(III). A plausible reaction for this process is:

\[ 2Fe_{Mg}^* + \frac{1}{2} O_{2(l)} \rightarrow 2Fe_{Mg}^* + O_t^* \]

\[ E = 2E\left(Fe_{Mg}^* \right) - 2E\left(Fe_{Mg}^* \right) + 2E_{Fe(II)} + E\left(O_t^* \right) + \frac{1}{2} E_{OD} + E_{A_2} = 6.17 \text{ eV. (4.17)} \]

Where \( E_{OD} \) is the dissociation energy of a oxygen molecule, \( E_{A2} \) is the sum of the first and second electron affinities of oxygen and \( E_{Fe(II)} \) is the third ionisation energy of iron. The values of these energies are taken from literature sources reported in table 4.15. This reaction is similar in character to those recently considered by Tsai and Dickmann (2002) although these authors explicitly chose not to consider the presence of defects in the oxygen sub lattice in order to simplify their analysis. The second possible mode of oxygen incorporation is via the formation of new forsterite lattice and vacancies on the cation sublattices. The reaction for this process is:

\[ 8Fe_{Mg}^* + 4O_{2(l)} \rightarrow 8Fe_{Mg}^* + 4O^* + 2V_{Si}^* + V_{Si}^{**} \]

\[ E = 8E\left(Fe_{Mg}^* \right) + 2E\left(V_{Si}^{**} \right) + E\left(V_{Si}^{**} \right) + U_{Mg, SiO_4} = 32.71 \text{ eV, (4.18)} \]

or 8.19 eV per oxygen ion.

Under reducing conditions it may be possible for oxygen to leave forsterite creating an oxygen deficit. Smyth and Stocker (1975) suggest two reactions for this process, each liberating excess electrons as the \( O^2 \) species become neutral atoms to form free molecular oxygen. I assume these electrons then reduce iron III to iron II in the lattice. The first reaction results in oxygen vacancies:

\[ 4O^* + 8Fe_{Mg}^* \rightarrow 2O_2 + 4V_{O_3}^{**} + 8Fe_{Mg}^* \]

\[ E = 8E\left(Fe_{Mg}^* \right) - 8E\left(Fe_{Mg}^* \right) - 8E_{Fe(II)} - 4E_{OD} - 4E_{A_2} = 9.06 \text{ eV, (4.19)} \]
(4.53 eV per oxygen ion), while the second involves the dissolution of part of the forsterite crystal with the formation of interstitial metal ions:

\[
4O_{2}^{*} + 2Mg^{*}_{Si} + Si^{*}_{Si} + 8Fe^{*}_{Mg2} \rightarrow 2O_{2} + 2Mg^{*}_{Si} + Si^{*}_{Si} + 8Fe^{*}_{Mg2},
\]

\[
E = 2E(Mg^{*}_{Si}) + E(Si^{*}_{Si}) + 8E(Fe^{*}_{Mg2}) - 8E(Fe^{*}_{Mg2}) = 10.01 \text{ eV},
\]

or 5.01 eV per oxygen ion.

**4.6 Conclusions**

The material presented in this chapter is in some ways only preliminary to the calculations presented in chapter 5, however some important results stand out. The first is that in pure forsterite the majority intrinsic defect species is predicted to be the magnesium Frenkel defect. This does not rule out the possibility of defects on the silicon or oxygen lattices – indeed these are required at equilibrium – but does indicate that electrical conductivity, for example, may be controlled by these defects. There has been at least one theoretical study of the intrinsic conductivity of forsterite (Morin et al., 1977, 1979), this analysed the likely introduction of bands in the band gap in forsterite on the basis of a comparison with MgO and quartz. The conclusion was that intrinsic conductivity can be explained by postulated magnesium interstitials on the unoccupied octahedral site. The results given above suggest that such defects do not exist and instead the split interstitial defect should predominate, the need for a re-evaluation of intrinsic conductivity data is therefore suggested.

In any case, electrical conductivity of olivine under mantle conditions is likely to be controlled by hydrogen diffusion; this is one reason for the major interest in hydrogen speciation in upper mantle rocks. As discussed in section 4.5.3, the results of the calculations allow insights into the mechanism of hydrogen incorporation in forsterite with results that allow the assignment of particular observed bands in the infrared spectra to particular defect structures. Hydrogen is also seen to promote the formation of silicon vacancies, an observation which may help to explain the weakening of olivine in the presence of water as suggested by Brodholt and Refson (2000). Further work is needed to develop a complete model of point defects in the mantle and 4.4.1.2 only begins this. In
addition to the energies of defects in forsterite, the defect states in the surrounding minerals should be considered along with the temperature and pressure. Such an analysis is beyond the scope of the current work, but the importance of the oxidation and reduction of iron in the lattice is established.

A further interesting observation is that oxygen ions seem to be able to form interstitial defects in the vacant octahedral sites in the olivine structure but, perhaps surprisingly, magnesium ions are unstable in this environment. The reason for this seems to be at least partially due to the electrostatic interactions between the defect and the rest of the crystal, best described by the electrostatic potential on the site. This is positive, so negatively charged oxygen ions are stabilised by the electrostatic potential while positively charged magnesium ions on the site are penalised. This effect has consequences for magnesium and oxygen diffusion, which is described in the following chapter.
Chapter 5: Diffusion in olivine

As discussed in the introductory chapter, the mobility of the atoms that make up mantle minerals has a great influence on mantle processes. Diffusion can control the rate of formation of new phases as minerals move between stability fields and may have an effect on the electrical conductivity of the mantle. It is also crucial in determining the rheology of the mantle. In this context the diffusivity of all the major components and a number of impurity species in olivine have been measured by laboratory experiments but, in general, little is known about the atomic scale diffusion mechanism under mantle conditions. Indeed it has been recognised (e.g. Béjina and Jaoul, 1997) that to understand the diffusion process on the atomic scale computational modelling must be used.

The aim of the work described in this chapter is to apply the computational methods described in chapter 3, to the problem of understanding the mechanism of diffusion on the atomic scale in olivine. As well as the intrinsic interest, such an understanding is essential if laboratory results are to be reliably extrapolated to mantle conditions and geological timescales, because of the danger of the mechanism changing under the different conditions in the laboratory and the mantle. Results presented in the chapter include models of oxygen diffusion and magnesium diffusion. Details of a study of hydrogen diffusion carried out in collaboration with A. Gatzemeier and S. M. Woodley will be described in a forthcoming publication. The basic theoretical background to diffusion was outlined in chapter 2 but before describing the results of work on diffusion in olivine the basic methods available to study diffusion using computer models are described.

5.1 Modelling diffusion

The ideal method to model diffusion in solids is to perform a “computer experiment” using long timescale first principle molecular dynamics and monitor the movement of the diffusing species. The approach is then to calculate the root mean square (RMS) displacement (averaged over all particles
and time origins) of the ions of interest (figure 5.1). The gradient gives
the diffusion coefficient via Einstein’s equation. However, in solids, the hopping
events leading to diffusion are rare and modelling diffusion via molecular
dynamics (using either parameterised potentials models or an electronic
structure approach) is generally beyond current computer resources. Although
some significant advances have been made in increasing the time scales that can
be probed (Voter, 1997b, 1997a, 1998; Sorensen and Voter, 2000) and applied
to some ionic materials (e.g. Harris et al., 2004) the approach used in this thesis
is to extract information from a study of the potential energy surface derived
from static calculations using a parameterised potential model.

![Simulation time graph](image)

**Figure 5.1:** Idealised cartoon of the single atom RMS displacement for various
behaviours not averaged over time origins. (a) Liquid like behaviour where an
atom continuously makes progress away from its starting point through
simulation time. (b) Behaviour of a solid where the atom vibrates on its lattice
site but fails to diffuse on the time scale of the calculation giving a flat RMS
displacement plot. (c) Intermediate behaviour where occasional hopping events
are caught in the simulation giving jumps in the atoms position, averaging this
behaviour over different time origins and a large number of atoms gives results
similar to (a).

The use of static techniques implies that the diffusion mechanism is sufficiently
slow for the diffusing ion to be re-equilibrated between each hop. In this case,
and in the absence of any quantum tunnelling effects (which are limited to low
temperatures and light atoms), the structure of the potential energy surface can
give the activation energy for intrinsic diffusion. The task is to find the pathway
between two energy minima (representing the start and end points for a hop)
with the lowest energy maximum. Such a pathway is pictured in two
dimensions in figure 5.2. The maximum energy found on the minimum energy
pathway is characterised by a saddle point or first order stationary point on the
potential energy surface. In common with the familiar one-dimensional case,
the stationary point is characterised by the first derivatives of the potential
energy with respect to the atomic co-ordinates tending to zero. In addition, the
second derivative matrix will indicate that in two directions the energy
decreases while the energy increases in all other directions – the second
derivative matrix will yield exactly one imaginary vibrational modes with
associated energy lowering atomic displacements.

Figure 5.2: 2 dimensional contour representation of a potential energy surface,
with surface overlay. The energy minima are at (35.5, 0.5) and (35.5, 8.5) and
the transition state is at (34.5, 4.5), the structure of the transition state can be
seen in the red surface above the contour plot.
In contrast to the location of minima and maxima, the location of stationary points of a multi-dimensional function is difficult. The mode following RFO approach offers one possibility, however the starting geometry must be close to the desired transition state. Other approaches, which require the geometry of the start and end points of the pathway to be known in advance, include the “nudged elastic band” method (Henkelman et al., 2000) and the linear stationary transit and quadratic stationary transit methods (Bell and Crighton, 1984; Fischer and Karplus, 1992). The dimer method (Henkelman and Jonsson, 1999) offers a route to the location of transition states to be found from the starting geometry only. A simple approach is to search for the transition state by constraining a series of geometry optimisation calculations such that the diffusing atom is fixed on a point away from its lattice site. The variation of energy with the position of the fixed ion can be used to build up a picture of the potential energy surface in order to locate the transition state. Although less powerful than the molecular dynamics, dimer or stationary transit methods, mapping of the potential energy surface remains the dominant approach for studies of slow diffusion in ionic materials (e.g. Islam, 2000; Woodley et al., 2003).

Once the transition state and activation energy is found, it is also possible to estimate the pre-exponential factor for diffusion. If molecular dynamics simulation has been used at a range of temperatures then the diffusivity as a function of temperature is known and the extraction of the activation energy and pre-exponential factor is achieved by fitting to the Arrhenius relationship as discussed in chapter 2. When static calculations have been used the process is not so straightforward. It is necessary to calculate the expected rate of each hopping event that passes through a transition state using Vineyard theory (Vineyard, 1957). If all the transition states are equivalent by symmetry, for example for oxygen diffusion in MgO, then the pre-exponential factor can be found (Vocadlo et al., 1995), but otherwise additional calculations (e.g. Kinetic Monte-Carlo simulation) are needed to resolve the diffusion pathway and give details of the absolute diffusivity.
5.2 Oxygen diffusion

Although high quality quantitative experimental data exists for oxygen diffusion in both synthetic forsterite (Sockel and Hallwig, 1977; Jaoul et al., 1980; Reddy et al., 1980; Sockel et al., 1980; Ando et al., 1981; Jaoul et al., 1983; Oishi and Ando, 1984) and in natural olivine, such as gem quality Fo90 from San Carlos, Arizona (Houlier et al., 1988; Gérard and Jaoul, 1989; and Ryerson et al., 1989), there is no general agreement on the nature of the atomic scale diffusion mechanism. Such knowledge is essential if we are to extrapolate laboratory data to mantle conditions. In order to understand this mechanism, the inherent atomic resolution of atomic modelling has been used to systematically calculate the activation energy of diffusion by a number of well defined possible competing mechanisms in forsterite. This allows the prediction of the most favourable diffusion mechanism in a perfect forsterite crystal. The study is then taken further to examine the effect of iron and hydrogen on oxygen diffusion and attempt to understand the importance of such defects in natural olivine. The results of the study reported in this section have been previously reported in the journal Physics and Chemistry of Minerals (Walker et al., 2003). Before the methodology and results of study are described, the available experimental data on oxygen diffusion in olivine are reviewed. It is worth noting here that the results of this study can be used to explain all of the existing experimental data and to make predictions regarding the oxygen diffusion mechanism that were not possible from the experiments alone.

5.2.1 Experimental background

The activation energy for oxygen diffusion has been measured in synthetic forsterite by Jaoul et al. (1980; 1983) (320 and 293 kJmol\(^{-1}\) respectively), Reddy et al. (1980) (372 kJmol\(^{-1}\)) and Ando and co-workers (Ando, et al. 1981; Oishi and Ando 1984) (415 kJmol\(^{-1}\)). In natural olivine, the activation energy has been measured in samples of San Carlos olivine by Gérard and Jaoul (1989) (318 kJmol\(^{-1}\) parallel to [001]) and Ryerson et al. (1989) (266 kJmol\(^{-1}\) parallel to [100]). More recently the diffusion has been measured in olivine of mantle origin by Dohmen et al. (2002), this study yields an activation energy of 338 kJmol\(^{-1}\) parallel to [001]. These measurements utilise natural samples with
The earliest experiments on oxygen diffusion in olivine utilized synthetic samples and were carried out at only one temperature (Sockel and Hallwig, 1977; Sockel et al., 1980). For synthetic forsterite, there is compelling experimental evidence that the activation energy for oxygen diffusion is isotropic; both Reddy et al. (1980) and Jaoul et al. (1983) measured diffusion in all three directions and at various temperatures and did not observe a difference in activation energy with direction of diffusion. However, diffusion has been observed to be fastest along the [010] direction (Reddy, et al. 1980). As the activation energy is equal in each direction, the anisotropy of the diffusion coefficient must be an expression of an underlying inequivalence of the pre-exponential factor for diffusion.

In natural olivine the activation energy has only been reported for diffusion parallel to the [100] and [001] axes. However, some data is available for diffusion parallel to [010] in natural olivine; summarised in figure 14 of Ryerson et al. (1989), reproduced in figure 5.3. At low oxygen partial pressures (with an Fe-FeO oxygen buffer) this figure shows that the variation in diffusion rate with temperature appears to be isotropic, but at high oxygen partial pressure (Ni-NiO buffer) the apparent activation energy parallel to [010] is 71 kJmol\(^{-1}\) higher than parallel to [100]. It is in this context that this study is particularly relevant, since it is possible to explore atomic transport directly and with relation to one axis, or indeed any combination of axes.
Figure 5.3: Outline of figure 14 of Ryerson et al. (1989) showing the variation of diffusivity (m/s, y-axis) with temperature (°C, x-axis), direction and oxygen fugacity. Fe-FeO buffer gives low oxygen partial pressure and Ni-NiO buffer gives high partial pressure. Reproduced without permission.

As well as being temperature dependent, the diffusion coefficient also varies with the concentration of the defects that permit diffusion. This can provide a powerful method to deduce the diffusion mechanism, if the concentration of the diffusing defects can be varied in a controlled way by, for example, altering the chemistry of the sample's surroundings. In olivine, the variation in the concentration of point defects with oxygen and enstatite activity has been studied theoretically by Stocker and Smyth (1977). This study has been used as a basis for the analysis of diffusion data collected with differing oxygen partial pressures. Analysis of this data is carried out by performing multivariate least squares regression onto equation 2.7 with $D_0$ re-defined by equation 5.1.

$$D_0 = C_0 f_{O_2}^p$$

(5.1)

The oxygen fugacity $f_{O_2}$ is known from the experiment, $C_0$, and $p$, are assumed to be constants found by the regression, $p$ describes the variation of diffusion rate with oxygen partial pressure and $C_0$ includes all the other factors contained within $D_0$ in equation 2.7. With luck, $p$ can be used to determine the nature of the diffusing species (Stocker and Smyth 1977). To simplify the argument, on an increase of the oxygen partial pressure, the concentration of oxygen vacancies would be expected to drop and the concentration of oxygen interstitial ions would be expected to rise. If a vacancy mechanism operates then the
decrease in vacancy concentration with increasing oxygen partial pressure would lead to a decrease in the value of $D$. Conversely, if an interstitial mechanism operates then the opposite would be true. To use this argument effectively, the ‘charge neutrality condition’ (balance of charged defects in the crystal), which can be found from a knowledge of the majority point defect type, must be known (Stocker and Smyth 1977). $p$ has been found to be 0 in synthetic forsterite (Jaoul, et al. 1983) and approximately 0.3 in natural olivine (Gérard and Jaoul 1989; Ryerson, et al. 1989). This positive, non-zero value of $p$ in natural olivine that has been used to suggest that oxygen diffusion operates by the interstitial mechanism.

Although, on simple geometrical and space filling grounds, in any close packed oxide, oxygen would be expected to diffuse by a vacancy mechanism, the experimental information has been used to argue that other mechanisms are important. For $\text{Fo}_{100}$, oxygen diffusion is independent of oxygen partial pressure; unfortunately, this does not provide an unequivocal identification of the diffusion mechanism. On the other hand the positive dependence on oxygen partial pressure found in San Carlos olivine precludes a vacancy mechanism for all the charge neutrality conditions considered by Stocker and Smyth (1977). Suggestions regarding the oxygen diffusion mechanism include a simple interstitial mechanism (Houlier et al., 1988), an interstitial mechanism where the diffusing species has a charge of $-1$ (Gérard and Jaoul, 1989), a coupled counter vacancy mechanism (Ryerson et al., 1989) or diffusion via cooperative motion of O3 and O2 ions (Brodholt, 1997). No experimental methods to discriminate between these mechanisms have been proposed. While analysis of San Carlos olivine provides an experimentally accessible route to study the oxygen diffusion mechanism, it is far from clear that the results are applicable to any other olivine sample, be it natural or synthetic. The diffusion mechanism may well depend on the concentration of other point defects (e.g. iron or hydrogen) in the sample and in order to understand oxygen diffusion under mantle conditions a range of different olivine samples should be studied.

5.2.2 Methodology

In order to investigate the oxygen diffusion mechanism in olivine, the activation energies for diffusion by the simple vacancy hopping and interstitial
mechanisms parallel to all three crystallographic axes were systematically calculated in forsterite. For each mechanism, diffusion is broken down into a series of "hops" between adjacent sites. Each hop is associated with a migration energy barrier. By making a series of hops, the diffusing ion may cross the unit cell. Under the assumption that consecutive hops are uncorrelated, the maximum migration energy required to achieve movement in a particular direction is the activation energy for diffusion in that direction. It is often possible for the ion to diffuse by more than one pathway across the cell; in these cases, the activation energy is the lowest activation energy of these routes. The mechanism with the lowest activation energy will be the dominant mechanism for diffusion. The activation energy found by this method assumes that there is no energetic cost associated with forming the diffusing defect, it is therefore the activation energy for extrinsic diffusion. For diffusion in other regimes, additional energy contributions must be included as described in chapter 2. These energy contributions are made up of other defect energies as described below.

In order to obtain the migration energies for each hop the defect energies of a number of point defects and defect clusters are calculated in the static limit using an atomistic model of the crystal. The atomistic model used is the same as that described in chapter 4, except that the original O-H Morse potential of Schröder et al. (1992) is used. This model was used to study defects corresponding to diffusing species in their minimum energy configuration and in the activated state (the maximum energy configuration on the minimum energy pathway between two points). The migration energy is found by subtracting the energy of the initial, minimum energy, state from that of the activated state. All energies are calculated using atomistic simulation techniques and the Mott-Littleton method implemented within the GULP code (Gale, 1997). Because a formal charge model is used for the oxygen ions, the coulombic interactions are likely to be too large. Because of the way in which the fitting is performed, this is counterbalanced at energy minima by the short range potential. However, away from the minimum where experimental data was not used to constrain the model, features on the model potential energy surface are likely to be over-expressed when compared to the real potential.
energy surface. The model will therefore provide an upper bound on the activation energies calculated using it.

Minimum energy configurations are found using standard quasi-Newtonian methods (section 3.3). The activated state is found utilising a transition state search algorithm based on the Rational Function Optimization (RFO) procedure described by Benerjee et al. (1985) and implemented within GULP. This method is capable of locating saddle points on an energy surface from an arbitrary starting point. The saddle point found by the RFO method will only correspond to the relevant transition state if the initial defect geometry is close to the true transition state for the hop of interest. This is because the energy surface will contain a large number of saddle points for alternative hops and, possibly, for hops between intermediate states along the migration pathway. In order to ensure that the relevant saddle point was found, an initial survey of the energy surface was carried out via the technique of adiabatic mapping. This technique involves fixing the diffusing ion in a series of positions on a three dimensional grid between the initial and final positions, and minimising the energy of the rest of the structure. A trial transition structure is selected from this grid and used as an initial configuration for the RFO transition state search.

A problem was encountered when searching for the structure of the transition state that was found to originate from the shell model representation of the oxygen ion. This model allows the positively charged oxygen core to be displaced from the negatively charged oxygen shell in order to simulate the polarisability of the O\(^2\) ion. As the ion moves away from its ideal minimum energy site it experiences an increasing polarisation and, in regions where the anion is in an activated state, the electronic deformation of the ion may have strong multipolar components. These are inadequately represented by the dipolar shell-model. This deficiency is often manifested within potential models like that outlined above by an excessive polarisation of the ion; the shell may move too far and become 'disassociated' from its core resulting in a dramatic, unphysical, decrease in defect energy. This problem is similar to that described by Wojcik and Hermansson (1998) and is overcome by adding a large \(k_4\) term that varies with the core – shell separation distance raised to the fourth power, this severely penalises large separations but has only a small effect for
small core – shell separations. This term is removed once a trial transition state is found and the energies reported below do not include this term.

5.2.3 Results

Perfect lattice and defect energies given by the model described above have been presented in chapter 4 (table 4.7). The defect energies mark minima on the potential energy surface that are the start and end point for each hop; they are the reference energy used to calculate the height of the energy barrier. These results also allow the consideration of diffusion outside of the extrinsic domain when corrections for the formation of the diffusing defect need to be made.

The broad agreement between electronic structure studies and the results from the Mott-Littleton methodology used here is described in chapter 4 and suggests that the potential model used in this study is appropriate for the description of defects in forsterite. This suggests that the potential model will give qualitatively correct activation energies and the relative energetics are sufficiently accurate to be used in the interpretation of experimental data.

In chapter 4, a number of methods of producing point defects on the oxygen sub lattice were considered. Before moving onto a description of the results of modelling diffusion, it is worth recalling some important conclusions. Firstly the formation of vacancies on the oxygen sub lattice is more likely to be produced by intrinsic MgO-Schottky defects then O-Frenkel defects and the energetic cost associated with introducing the Schottky defect was found to be 7.71 eV. The second important observation is that oxygen interstitials could be formed at high oxygen partial pressure charge balanced by reduction of Fe III to Fe II with an energetic cost of 6.17 eV. The suggestion is that for pure forsterite a small concentration of intrinsic oxygen vacancies is expected but with the presence of iron in solid solution and at sufficiently high oxygen partial pressure these vacancies may be replaced by oxygen interstitials.

5.2.3.1 Vacancy mechanism

One additional result of particular note is the large difference (at least 0.66 eV) between the energy required to form vacancies on the three symmetry inequivalent oxygen sites. This was previously reported by Brodholt (1997) and was used to argue that almost all oxygen vacancies are to be found on the O3
site. This leads to a considerable simplification when considering the vacancy mechanism as we only need consider four hops between O3 sites rather than the 15 symmetry inequivalent hops between adjacent oxygen sites in the HCP sub-lattice. These four hops are shown in figure 5.4. Three of these hops are between adjacent O3 sites in the sub-lattice, the fourth (hop C) is a two-stage hop through interstitial I1. Two of the three simple hops (A and D) are hops between different Si04 tetrahedra, hop A resulting in a movement along [100] and [010] and hop D only having a component along [001]. Hop B is the hop between O3 sites belonging to the same Si04 unit.

**Figure 5.4:** Looking down [100] to four unit cells showing the possible routes for diffusion along [010] and [001] and all 4 hops assuming a vacancy mechanism where only O3 vacancies are involved. Colours of species, and locations of the various oxygen ions are described in figure 1.3.

Inspection of figure 5.4 gives the possible routes that a diffusing O3 vacancy may take to cross the unit cell in each direction. Diffusion parallel to [100] is only possible by repeated occurrences of hop A or occurrences of hop A alternating with hop B. In either case for diffusion parallel to [100], the (extrinsic) activation energy will equal the migration energy of hop A. Diffusion parallel to [010] can be achieved by repeated occurrences of hop A, B and either hop C or D (the activation energy may be equal to the migration...
energy associated with any of these hops). Diffusion parallel to [001] requires either hops C and D, C and A, or D and A. The migration energy calculated for each hop is reported in table 5.1. In each case diffusion will be rate-limited by hop A. This gives an extrinsic activation energy of 119 kJmol\(^{-1}\) independent of the direction of diffusion.

**Table 5.1:** Migration energies for vacancy and interstitial diffusion hops. Hops A – D are between O3 sites and hops E – I are for interstitial diffusion. Hops E and F are between adjacent I2 sites, hops G – H are from I2 to I1 sites and hops G’ - H’ are from I1 to I2 sites. The approximate hop distances are for straight lines measured in the perfect, undistorted, forsterite lattice.

<table>
<thead>
<tr>
<th>Hop</th>
<th>Approximate hop distance (Å)</th>
<th>Defect energy of initial state (eV)</th>
<th>Defect energy of activated state (eV)</th>
<th>Migration energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3.0</td>
<td>24.54</td>
<td>25.765</td>
<td>1.23</td>
</tr>
<tr>
<td>B</td>
<td>2.6</td>
<td>24.54</td>
<td>25.260</td>
<td>0.72</td>
</tr>
<tr>
<td>C</td>
<td>3.4</td>
<td>24.54</td>
<td>25.223</td>
<td>0.69</td>
</tr>
<tr>
<td>D</td>
<td>3.3</td>
<td>24.54</td>
<td>31.270</td>
<td>6.73</td>
</tr>
<tr>
<td>E</td>
<td>3.0</td>
<td>-16.11</td>
<td>-14.942</td>
<td>0.97</td>
</tr>
<tr>
<td>F</td>
<td>3.0</td>
<td>-16.11</td>
<td>-15.144</td>
<td>1.17</td>
</tr>
<tr>
<td>G</td>
<td>3.5</td>
<td>-16.11</td>
<td>-13.542</td>
<td>2.57</td>
</tr>
<tr>
<td>G’</td>
<td></td>
<td>-14.37</td>
<td></td>
<td>0.83</td>
</tr>
<tr>
<td>H</td>
<td>3.3</td>
<td>-16.11</td>
<td>-14.265</td>
<td>1.86</td>
</tr>
<tr>
<td>H’</td>
<td></td>
<td>-14.37</td>
<td></td>
<td>0.11</td>
</tr>
<tr>
<td>I</td>
<td>3.4</td>
<td>-16.11</td>
<td>-13.177</td>
<td>2.93</td>
</tr>
<tr>
<td>I’</td>
<td></td>
<td>-14.37</td>
<td></td>
<td>1.19</td>
</tr>
</tbody>
</table>

As the energy of the activated state for hop A is higher than the energy minimum corresponding to a vacancy on the O2 site it is possible that there is a pathway that involves diffusion through the O2 site. Such a pathway was suggested by Brodholt (1997), but the activation barrier was not calculated. In order to test this possibility the migration energy for hops from O3 to O1 and O2 sites within the same SiO\(_4\) tetrahedral unit were calculated. The energies of the activated states corresponding to these hops are 28.001 eV (and a barrier of 3.46 eV) for the O3 to O1 hop and 27.159 eV (and a barrier of 2.63 eV) for the
O3 to O2 hop (hop distances are 2.8 and 2.5 Å respectively). Therefore, the O3 to O3 hopping mechanism via hop A is the most energetically favourable mechanism.

5.2.3.2 Interstitial mechanism

For the interstitial mechanism it is not possible to automatically eliminate hops using the assumption of defects only occurring on I2 sites as was possible for the vacancy mechanism using the assumption of defects only occurring on O3 sites. This is because this would preclude diffusion along [010] by the interstitial mechanism. Instead, it is assumed that the unfavourable I1 sites are temporarily occupied by the diffusing ion. Although the activated state for hops from the I1 to the I2 site will be the same as for hops from the I2 to the I1 site, the energy of the initial configuration will be different. This means that the migration energy for the two types of hop will also be different. Hops involving interstitials are outlined in figure 5.5. Hops E and F are between adjacent I2 sites. Hops G, H and I are from I2 to I1 sites and G’, H’ and I’ are from I1 to I2 sites through the same activated state. Hops G, G’, H and H’ involve diffusion parallel to all three axes. The majority of movement achieved by hops I and I’ is made up of components along [010] and [001] directions. Diffusion parallel to [100] requires hop E and F, H and I’ or H’ and I. Diffusion parallel to [010] requires hop I, H or G and I’, H’ or G’. Diffusion parallel to [001] requires hop I, H or G and I’, H’ or G’, or hop E, or hop F. The calculated migration energies for these hops are presented in table 5.1. These results give an extrinsic activation energy parallel to [100] of 113 kJmol⁻¹ rate limited by hop F. Diffusion parallel to [010] is rate controlled by hop H with an activation energy of 178 kJmol⁻¹ and parallel to [001] the diffusion is controlled by hop E with an activation energy of 94 kJmol⁻¹.
Figure 5.5: Looking down [100] to four unit cells showing the possible routes for diffusion along [010] and [001] and all 5 hops assuming an interstitial mechanism where only I2 sites are preferred over I1 sites. Species are shown as in figure 1, interstitial position I2 is shown for clarity.

5.2.3.3 The effect of hydrogen and iron

In natural olivine, the oxygen diffusion mechanism may be modified by the presence of iron in solid solution or by point defects, including those formed by the incorporation of water into the lattice. When iron replaces magnesium in the unit cell the potential energy surface will be modified and there is also the possibility of redox reactions, where Fe^{2+} ions are oxidised to form Fe^{3+} ions. A key effect of such redox chemistry is likely to be enhancement of oxygen solubility in iron bearing olivine when compared to pure forsterite as suggested by equation 4.17. This could in turn lead to the extrinsic diffusion of oxygen interstitials. It has been shown by theory and experiment (e.g. Wright and Catlow, 1994; Rossman, 1996) that water will tend to disassociate and form OH point defects. If these defects diffuse more quickly then oxygen, oxygen diffusion may occur by the migration of these OH ions rather than by the migration of oxygen ions.

The magnitude of the effect of divalent iron in solid solution on the potential energy surface traversed by diffusing oxygen ions has been considered by calculating the lattice energy and defect energies for the initial state and transition state for hop A in pure fayalite; these are -844.49, 22.55 and 23.50 eV.
respectively. The activation energy barrier for vacancy diffusion is therefore decreased to 1.04 eV (100 kJmol⁻¹). This effect is small and is expected to diminish as the iron concentration decreases. The difference is probably too small to reliably detected by experiment.

As a first stage in the assessment of possible OH⁻ diffusion, the energy of a number of OH⁻ defects including hydrogen bound to the three inequivalent oxygen ions, bound to interstitial oxygen and associated with oxygen ions adjacent to oxygen vacancies have been calculated (table 5.2).

**Table 5.2:** energies of various hydroxyl defects and hydroxyl – oxygen defect clusters in forsterite.

<table>
<thead>
<tr>
<th>Defect</th>
<th>Defect energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{OH}^\prime_{o_1}$</td>
<td>18.05</td>
</tr>
<tr>
<td>$\text{OH}^\prime_{o_2}$</td>
<td>16.87</td>
</tr>
<tr>
<td>$\text{OH}^\prime_{o_3}$</td>
<td>16.57</td>
</tr>
<tr>
<td>$\text{OH}^\prime_{f(1)}$</td>
<td>-0.37</td>
</tr>
<tr>
<td>$\text{OH}^\prime_{f(2)}$</td>
<td>-3.51</td>
</tr>
<tr>
<td>$[\text{OH}<em>{o_1}V</em>{o_2}]^\prime$</td>
<td>44.66</td>
</tr>
<tr>
<td>$[\text{OH}<em>{o_1}V</em>{o_3}]^\prime$</td>
<td>43.41</td>
</tr>
<tr>
<td>$[\text{OH}<em>{o_2}V</em>{o_1}]^\prime$</td>
<td>46.09</td>
</tr>
<tr>
<td>$[\text{OH}<em>{o_2}V</em>{o_3}]^\prime$</td>
<td>42.61</td>
</tr>
<tr>
<td>$[\text{OH}<em>{o_3}V</em>{o_1}]^\prime$</td>
<td>43.07</td>
</tr>
<tr>
<td>$[\text{OH}<em>{o_3}V</em>{o_2}]^\prime$</td>
<td>41.76</td>
</tr>
<tr>
<td>$[\text{OH}<em>{o_3}V</em>{o_3}]^\prime$</td>
<td>42.74</td>
</tr>
</tbody>
</table>

It is then possible to calculate the binding energy between OH⁻ occupying oxygen sites and oxygen vacancies:

$$\text{OH}^\prime_{o_3} + V_{o_3} \rightarrow [\text{OH}_{o_3}V_{o_2}]^\prime,$$

$$E(\text{bind}) = E([\text{OH}_{o_3}V_{o_2}]^\prime) - \left(E(\text{OH}^\prime_{o_3}) + E(V_{o_3}^\prime)\right) = 0.65eV,$$  \hspace{1em} (5.2)

and the enthalpy of the reaction that results in a OH moving into an interstitial position, displacing the interstitial oxygen into the resulting vacancy:
Equation 5.2 shows that there is a thermodynamic driving force that acts to separate hydrogen ions bound to lattice oxygen from oxygen vacancies (this value is for a dilute system and gives an upper bound for real systems). This is due to electrostatic repulsion between the two defects with effective positive charge. The implication is that the presence of hydrogen in the olivine lattice will not affect the rate of oxygen diffusion by the vacancy mechanism. Equation 5.3 indicates that, if oxygen interstitials and hydrogen already co-exist, the hydrogen will tend to bind to the oxygen interstitial rather than any other lattice oxygen. If oxygen is diffusing by an interstitial mechanism this may result in a change in diffusion rate. Based on the same hops that we considered for oxygen interstitial diffusion the activation energy for interstitial OH diffusion can be calculated. For hop E the calculated energy of the activated state is -8.22 eV giving an activation energy barrier of 2.34 eV and for hop F the calculated energy of the activated state is -9.19 eV giving an activation energy barrier of 1.37 eV (interstitial OH diffusion has a higher activation barrier than the diffusion of isolated oxygen ions). If both species (O$^{2-}$ and OH$^-$) are present as interstitials, then oxygen diffusion will occur by the movement of oxygen interstitials rather than hydroxyl ions and the presence of hydrogen will not affect interstitial diffusion, if, however, all interstitial oxygen ions are hydrated then diffusion is expected to slow down.

5.2.4 Discussion

The results indicate that, of the mechanisms considered, vacancy diffusion should be the dominant oxygen diffusion mechanism in pure forsterite. However, the extrinsic activation energies calculated above are clearly not comparable to the activation energies determined by experiment. Given the high temperature of these experiments, it is probable that they sampled the intrinsic diffusion regime.

The intrinsic activation energy can be found by adding the fraction of the defect formation (e.g. Schottky or Frenkel) energy required to form the diffusing
species to the extrinsic activation energy. As Schottky defects will be the predominant intrinsic oxygen defect at high temperature we therefore add 3.86 eV to the extrinsic activation energy to give an intrinsic activation energy 5.09 eV (492 kJmol⁻¹). This energy is higher than all experimentally measured values. Reasonable explanations for this discrepancy include: (a) in the real crystals however there will be a population of other defects (point, line and planar in nature) that will interact with the diffusing defect, using the Mott-Littleton method we do not attempt to treat these interactions; (b) the experimental values for the activation energy may, at the lowest temperatures, include calculations which sampled the extrinsic regime - the credibility of this argument is enhanced when one considers that in the experimental study that yields the largest activation energy (Ando et al., 1981; Oishi and Ando, 1984) was conducted at, and sampled, the highest temperatures (and does not include temperatures below 1400 °C); (c) the potential model includes formal charges, this may result in the barrier height being over estimated.

In cases where oxygen vacancies have been replaced by interstitials, the model predicts an extrinsic activation energy of 113 and 94 kJmol⁻¹ parallel to [100] and [001] and a much larger activation energy (178 kJmol⁻¹) parallel to [010]. This anisotropy in activation energy is significant and may provide a method to experimentally discriminate between vacancy and interstitial mechanisms where diffusion is not sensitive to the oxygen partial pressure (i.e. in pure forsterite). Using only the results from this study, it is not possible to give any indication of the expected anisotropy in the diffusivity of oxygen, except to say that there need be none when the vacancy mechanism is operational and the interstitial mechanism certainly will express such anisotropy. For the vacancy mechanism, the details of any anisotropy in diffusivity will depend only on the pre-exponential and will not change with temperature but for the interstitial mechanism the anisotropy in diffusivity will obviously be temperature dependent as well.

5.3 Magnesium diffusion

The diffusion of magnesium ions has received more attention than the diffusion of oxygen with experimental studies of tracer diffusion and Mg-Fe
interdiffusion as well as some computational studies. As well as its significance for plastic deformation, the rate of magnesium diffusion in olivine is an important parameter in “geospeedometers” where, for example, the partitioning of magnesium and iron between olivine and spinel can be used to estimate the cooling rates of ultramafic rocks. Despite this large body of work, the atomic scale diffusion mechanism remains unknown.

5.3.1 Experimental background

There is a great deal of experimental data concerning the diffusion of cations in olivine including a series of studies of silicon diffusion (Sockel et al., 1980; Houlier et al., 1988; Béjina and Jaoul, 1996; Dohmen et al., 2002; Béjina et al., 2003) and many studies of the diffusion of the M site cations: including magnesium (Sockel and Hallwig, 1977; Sockel et al., 1980; Chakraborty et al., 1994), iron (Nakamura and Schmalzried, 1984; Bertran-Alvarez et al., 1993; Jaoul et al., 1995; Chakraborty, 1997) and cobalt (Morioka, 1980). When considering magnesium diffusion, these experiments can be separated into two types, those that measure tracer diffusion by diffusing $^{26}$Mg into an olivine sample of normal isotopic composition and those that measure the interdiffusion of magnesium and another element between two olivine samples of different compositions. The tracer experiments should yield results close to the true self-diffusivity of magnesium in olivine while interdiffusion experiments yield some average of the diffusivities of the two diffusing elements as described in chapter 2.

The most complete set of tracer diffusion experiments is that of Chakraborty et al. (1994) who performed experiments on synthetic crystals of forsterite as well as natural samples of San-Carlos olivine at temperatures between 1000 and 1300 °C under conditions of controlled oxygen fugacity ($pO_2$). They give activation energies for diffusion along [001] of $400(\pm 60)$ kJmol$^{-1}$ for magnesium diffusion in forsterite and $275(\pm 25)$ kJmol$^{-1}$ for San Carlos olivine and find that diffusion along [010] and [100] is slower than diffusion along [001] but do not give activation energies in the two slow directions. Magnesium diffusivity in San Carlos olivine was found to vary with $pO_2$ raised to the power of 1/6, this result, which is in fair agreement with previous studies (e.g. Nakamura and Schmalzried, 1984, found $1/5.5$). This suggests that the diffusing species is a
magnesium vacancy charge balanced by the formation of electron holes or by
the oxidation of iron but does not rule out the possibility of diffusion of
magnesium interstitial ions formed with a charge neutrality condition involving
magnesium vacancies and singly charged oxygen vacancies (Stocker and
Smyth, 1977). In the synthetic olivine, the effect of $pO_2$ is less clear, with
different samples giving different results, two possibilities are put forward, the
first invoking a $pO_2$ dependent change in mechanism and the second invoking
interstitial Fe(III) ions in the charge neutrality condition. The effect of pressure
on magnesium diffusion was also studied and the activation volume was found
to be small and positive (about 1 cm$^3$mol$^{-1}$) which is similar to the value derived
from interdiffusion experiments (Bertran-Alvarez et al., 1993; Jaoul et al.,
1995).

Chakraborty et al. (1994) interpret their activation energies as extrinsic
diffusivities (in the sense that they depend on the concentration of the
component ions) and argue that intrinsic diffusivity is unlikely. This argument
is based on two lines of evidence: (1) that the formation energy of a magnesium
vacancy is higher than their measured activation energy for diffusion and (2)
that the impurity concentration is sufficiently high to dominate the formation of
magnesium vacancies. This argument is not that the measured activation
energies are equal to the migration energy (their definition includes the
diffusion of defects formed by iron redox as extrinsic diffusivities) but is
critically flawed in two ways. The energy added to the migration energy to
arrive at the intrinsic activation energy is not the formation energy of the
intrinsic (Schottky or Frenkel) defect as argued by Chakraborty et al. (1994),
but is a fraction of this energy as described in chapter 2. Using the results
presented in chapter 4, and in the case of magnesium Frenkel defects, the energy
to be added to the migration energy is 3.37 eV not the $-8.33$ eV argued by
Chakraborty et al. (1994), although this is a large fraction of the measured
activation energy it is not large enough to rule out intrinsic diffusion in the
synthetic sample. The second argument rests on the assumption that each
impurity ion in the crystal will form a vacancy on the magnesium sublattice, but
various computational studies (Purton et al., 1997; Richmond and Brodholt,
2000) suggest that 3+ ions entering the forsterite lattice in dilute concentration
do so by simultaneous substitution for a silicon and magnesium ion and do not create magnesium vacancies. This is supported by experimental studies of chromium, aluminium and iron (Chatelain and Weeks, 1971; Niebuhr, 1975; Gaite and Hafner, 1984; Rager et al., 1988; Gaite and Ranger, 1997), which find evidence for 3+ ions in both octahedral and tetrahedral co-ordination.

Chakraborty (1997), in an analysis of Fe-Mg interdiffusion data that is consistent with the results of Chakraborty et al. (1994), goes further and suggests that the measured activation energy “are essentially the migration energy and that the contribution of the formation energy is smaller in magnitude than the uncertainty in the knowledge of the activation energy”. This conclusion is based on an estimate of the energetics of the formation of magnesium vacancies by iron redox reactions that suggests that the energetic cost of forming a M site vacancy in fayalite, charge balanced by the oxidation of iron, costs less than 12 kJmol⁻¹. Chakraborty (1997) then goes on to argue that parameterised potential functions are unlikely to be of use for computational studies of diffusion, mysteriously citing Lidiard (1981) and Gillan (1981) despite the lack of relevance of either reference and the abundance of successful studies of diffusion using just this method.

In contrast, the largely experimental study of Mg-Fe interdiffusion undertaken by Jaoul et al. (1995) makes use of computational results in order to assist in the interpretation of their experimentally determined activation energies and volumes. The experiment gives a measured activation energy for diffusion along [010] of 147(±58) kJmol⁻¹ with almost no pressure dependence (the activation volume was found to be 1 cm³mol⁻¹). If account is taken of the effect of $pO_2$ then this is reduced to 62(±58) kJmol⁻¹ in good agreement with their calculated migration energy of 58 kJmol⁻¹. Adding the formation energy of a magnesium Frenkel defect gives an intrinsic activation energy of 3.9 eV which is reported to be in agreement with previous high temperature experimental studies undertaken by Andersson giving activation energies between 3.7 and 4.4 eV.

Finally, it is worth commenting on the anisotropy of magnesium diffusion in olivine. All studies have concluded that diffusion along [001] is faster than diffusion along [010] and [100] but there is less data on the anisotropy of the
activation energy. One exception is the data of Sockel et al. (1980) whose limited data clearly show a much lower activation energy for diffusion along [001] compared to diffusion along [100] or [010].

5.3.2 Methodology

The methodology for studying magnesium diffusion by the vacancy mechanism was identical to that described above (section 5.3.1) but, because of the apparent complexity of the potential energy surface discovered in the search for the geometry of magnesium interstitial defects, a different approach was used for interstitial diffusion. The general approach is similar, the potential energy surface is first mapped to locate approximate saddle points and then a RFO transition state search is performed, but the method of locating the approximate saddle point is different. Rather than predetermining individual steps for diffusion, a large segment of the potential energy surface corresponding to moving the interstitial magnesium ion and relaxing the rest of the structure was evaluated.

Once the “complete” potential energy surface has been “mapped”, the problem becomes one of finding suitable approximate saddle points. Many approaches are possible and a simple one devised and implemented by Dr S. M. Woodley (personal communication) was used. This iterative approach is easily described, considering an analogous problem on a two dimensional topography with height representing the energy. The data gives the height on a regular grid across a rectangular area; this area is repeated in the nine surrounding rectangular areas. The first stage is to locate the lowest point in the central rectangle, achieved by searching over all data points. The problem is then to find the routes to the equivalent points in the adjacent rectangles that involve the lowest maximum heights. One way to find these routes is to pour water into the lowest point in the central rectangle, as the basin fills up the water will spill over from the lowest laying basin into other higher basins and eventually one of the surrounding equivalent lowest basins will flood. At this stage, the highest “wet” point is the transition state from one global minimum to the other. Adding more water will allow the transition states in the other direction to be found.
In the actual case of diffusion in a crystal the system is periodic in three directions so 27 surrounding cells have to be monitored, but apart from this the procedure is very similar with a 3D “bubble” defining the volume a diffusing atom with a certain amount of energy can access (or have a fixed probability of accessing); when this bubble makes contact with the global minimum in an adjacent unit cell diffusion in that direction is possible and an approximate transition state has been found. A grid of 20 points in x, 40 points in y and 20 points in z (i.e. ~0.25 Å grid point spacing), shown in figure 5.6, was used to cover the unit cell. The space group symmetry of forsterite was applied to limit the number of points mapped, reducing the 16000 calculations to a more manageable 2000. Once the approximate transition states have been located the RFO transition state search is started, this means that the data grid does not need to be too fine and allows for refinement of the transition structure with motion of the diffusing ion away from its grid point.

5.3.3 Results

Figure 5.7 shows the five inequivalent routes between magnesium vacancies that may be involved in magnesium vacancy diffusion. Hop A is between two Mg1 sites along [100] through the vacant octahedral I1 interstitial position while hop B is between two Mg2 sites along [100] through the vacant octahedral I2 position. Both of these hops exhibit high barriers to diffusion because of the low stability of magnesium ion in the I1 and I2 sites. Hop C is between two Mg2 positions with displacement mostly within a (001) plane. Hops D and E are from Mg1 sites to Mg2 sites with D mostly within the (100) plane and E with components in all three crystallographic directions. Hops D’ and E’ are the reverse hops from Mg2 sites to Mg1 sites. Hop F is between two Mg1 ions along [001].
Figure 5.6: Graphical representation of the 2000 data points used to estimate the location of the transition states for interstitial diffusion. Each data slice is for a different value of x with (a) = 0 and (e) = 0.5, values are in eV.
Figure 5.7: Magnesium diffusion by the vacancy mechanism (see text for details) projected onto the (100) plane. Hops A and B are not shown as they are perpendicular to the plane.

Migration energies for each of these hops is given in table 5.3, in addition the transition states for interstitial diffusion are also shown. There are several points to note, the first being the large barriers to diffusion through the vacant octahedra (hops A and B), this is a somewhat surprising result given the expectation that magnesium “prefers” an octahedral environment especially given that the transition state is found to be in close to the centre of the octahedron. As described in chapter 4, it seems likely that this is a coulombic effect that is also responsible for the lack of stable octahedrally co-ordinated magnesium interstitials. Vacancy diffusion along [001] is predicted to be via hop F with the low activation energy of 0.72 eV and diffusion along [100] and [010] is predicted to be via hops D and C with an extrinsic activation energy controlled by hop D with a value of 4.54 eV. Interstitial diffusion is not favoured over vacancy diffusion in any direction.
Table 5.3: Migration energies for magnesium vacancy and interstitial diffusion in forsterite.

<table>
<thead>
<tr>
<th>Hop</th>
<th>Defect energy of initial state (eV)</th>
<th>Defect energy of activated state (eV)</th>
<th>Migration energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>24.48</td>
<td>30.37</td>
<td>5.89</td>
</tr>
<tr>
<td>B</td>
<td>26.40</td>
<td>35.50&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9.10</td>
</tr>
<tr>
<td>C</td>
<td>26.40</td>
<td>27.87&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.47</td>
</tr>
<tr>
<td>D</td>
<td>24.48</td>
<td>26.46</td>
<td>1.98</td>
</tr>
<tr>
<td>D'</td>
<td>26.40</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>24.48</td>
<td>30.94&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6.64</td>
</tr>
<tr>
<td>E'</td>
<td>26.40</td>
<td>6.64</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>24.40</td>
<td>25.12</td>
<td>0.72</td>
</tr>
<tr>
<td>Interstitial // [100]</td>
<td>-17.75</td>
<td>-13.86</td>
<td>3.89</td>
</tr>
<tr>
<td>Interstitial // [010]</td>
<td>-17.75</td>
<td>-13.86</td>
<td>3.89</td>
</tr>
<tr>
<td>Interstitial // [001]</td>
<td>-17.75</td>
<td>-14.62</td>
<td>3.13</td>
</tr>
</tbody>
</table>

(a) These failed to converge in the RFO part of the calculation and so an estimate of the transition state is made from the initial search, where the moving ion is fixed and the rest of the structure relaxed, is used, the true energy of the activated state is not expected to be significantly different from this estimate.

5.3.4 Discussion

The migration energies presented above are for extrinsic diffusion (in the classical sense), and are within error of the experimental results of Jaoul et al. (1995), after their $pO_2$ correction. For intrinsic diffusion (pure Mg$_2$SiO$_4$ with thermally created point defects) an appropriate formation energy must be added. The results of chapter 4 suggest that the appropriate defect reaction is the magnesium Frenkel defect, and 3.37 eV should be added to the predicted migration energies to yield the intrinsic activation energy. This results in activation energies of 5.35 eV (513 kJmol$^{-1}$) along [100] and [010] and 4.09 eV (393 kJmol$^{-1}$) along [001], which is within the stated error of the results of Chakraborty et al. (1994) for higher temperatures where intrinsic diffusion would be expected. The sense of the anisotropy in activation energy is also correctly described although no experimental results for the activation energy along the slow directions have been presented.

For iron bearing olivine Chakraborty et al. (1994) extracted significantly lower activation energies, presumably because the magnesium vacancies are formed at lower energetic cost. One way to form magnesium vacancies is to charge balance their formation with the oxidation of iron. This process is represented by the reaction:
\[ 2Fe^{*}_{Mg} + Mg^{*}_{Mg} + \frac{1}{2} O_{2(g)} \rightarrow 2Fe^{*}_{Mg} + V^{-}_{Mg} + MgO_{(s)} \]

\[ E = 2E(Fe^{*}_{Mg2}) - 2E(Fe^{*}_{Mg2}) + 2E(Fe^{*}_{Fe13}) + E(V^{-}_{Mg1}) + \frac{1}{2} E_{OD} + E_{A2} = 5.45 \text{ eV}. \tag{5.4} \]

Adding one third of this (1.82 eV) to the [001] migration energy yields a predicted activation energy of 2.54 eV (243 kJmol\(^{-1}\)), a little lower than the energy measured by Chakraborty et al. (1994), but within error of that measured by Chakraborty (1997) for a more iron rich composition.

### 5.4 Conclusions

The absolute accuracy of the energy determinations will of course depend on the performance of the potential model and, especially its ability to describe defect structures and energetics. While the energy of stable defects calculated using this potential model is in fair agreement with available electronic structure calculations, there are no electronic structure studies that have attempted to compute the geometry of the activated state or the defect energy associated with the formation of interstitial oxygen. The high activation energies (when compared to experiment) reported here may be due to inaccuracies in the potential model’s description of the activated state.

By breaking diffusion down into a series of hops, evaluating the migration energy for these hops and then calculating the activation energy for diffusion in each direction some assumptions have made regarding the diffusion mechanism. The principle assumption is that between each hop the defect has time to re-equilibrate so that each hop is independent of the next. The methodology also limits the mechanisms that can be studied. In essence before the activated state can be found the exact sequence of atomic movements of the proposed mechanism must be written down. This is possible when only one or two atoms move but if cooperative motion occurs the number of possible “hops” (each involving the movement of several ions) increases dramatically and it rapidly becomes unfeasible to consider all permutations. A method that is not subject to these limitations is molecular dynamics. In principal, in a long molecular dynamics simulation diffusion could be observed directly. However, even at high temperature, the extremely long simulation time required to observe the
rare (less than one per nanosecond) diffusion events requires very substantial amounts of computer time and a highly stable potential model.

The first conclusion that can be drawn from this study is that in forsterite, oxygen diffusion will occur by the vacancy mechanism as would be expected on simple geometrical grounds. The simple atomistic model predicts that in the intrinsic high temperature domain the activation energy for this process will be 492 kJmol⁻¹ and diffusion will be isotropic. The second point is in agreement with what is found experimentally and the overestimation of activation energy may be due to inaccuracies in the potential model’s description of the activated state or may indicate that the experimental results show a mixture of intrinsic and extrinsic diffusion. A major and perhaps surprising conclusion is that oxygen diffusion is not affected by varying the hydrogen²⁰ or Fe²⁺ content of the sample.

Experiments on natural olivine have been used to preclude a vacancy mechanism. One possibility is that at high oxygen partial pressure oxygen interstitials may be formed, mediated in natural olivine by the reduction of iron III. If interstitial oxygen ions diffuse the model predicts that diffusion will be strongly anisotropic.

It is enlightening to compare these conclusions with figure 14 of Ryerson et al. (1989) shown in figure 5.3. This figure shows an Arrhenius diagram of oxygen diffusion in San Carlos olivine with two oxygen buffers. Two important observations can be made on increasing oxygen partial pressure by changing the oxygen buffer: (a) the diffusivity increases, (b) the activation energy becomes anisotropic with a difference between diffusion parallel to [100] and [010] of 71 kJmol⁻¹. The results suggest that this is because the diffusion mechanism changes from the vacancy mechanism to the interstitial mechanism with increasing oxygen partial pressure. The computed energy difference between diffusion parallel to [100] and [010] is 65 kJmol⁻¹, remarkably similar to the experimental value. However, currently available experimental data is

---

²⁰ This is in agreement with unpublished experimental findings, which show that oxygen diffusivity is identical in "wet" and "dry" olivine, but that oxygen diffusivity in some other minerals (e.g. quartz) is strongly dependent on the water content (S. Elphick, Personal Communication).
insufficient to confirm this and the study reported here provides an interesting hypothesis to be tested. An ideal test would be to measure the dependence of diffusion on oxygen fugacity at both low and high oxygen fugacity separately and to measure the activation energy along all three axes under both sets of conditions. The results could be used to evaluate such data and could also be used to interpret the diffusion mechanism in iron free olivine.

Turning to magnesium diffusion, the results are in agreement with previous computational studies (Jaoul et al., 1995) and agree to a remarkable degree with the experimental data. It seems that the data of Jaoul et al. (1995) and Chakraborty et al. (1994) for magnesium diffusion in San Carlos olivine can be interpreted as the diffusion of magnesium vacancies charge balanced by iron III (with the iron oxidation either corrected to extract “true” extrinsic activation energies or included in the calculation of the activation energy). The higher activation energy measured by Chakraborty et al. (1994) in synthetic forsterite seems to indicate that true intrinsic diffusion, with magnesium Frenkel defects providing the source of vacancies, was measured in that case. It seems that diffusion of magnesium interstitials is not favoured over vacancy diffusion, a conclusion reinforced by the many measurements of positive $pO_2$ dependence of magnesium diffusion in olivine.

In summary, the work in this chapter suggests that both vacancy and interstitial oxygen diffusion may operate in olivine but for magnesium diffusion, only the vacancy mechanism is likely. For both elements, the presence of iron has a significant effect that should be accounted for in modelling diffusion and in understanding mantle rheology. The following two chapters turn away from point defects as agents of crystal deformation and focus instead on line defects; the second class of defect that may be responsible for controlling the rheology of olivine.
Chapter 6: Atomic scale simulation of the cores of dislocations in complex ionic materials

For more than three decades, computational simulation methods have been used to overcome experimental limitations and study the atomic-scale detail of the cores of line defects in crystalline solids. However, these studies (see section 6.1) have so far been limited to crystals with very simple structures, namely the alkali halides and binary oxides with the rock salt structure, various metals and semi-conductors. The aim of the work described in this chapter is to develop a methodology that will enable the study of screw and edge dislocations in much more complex ionic and semi-ionic materials. This will enable the study of dislocation cores in a very wide rage of materials, particularly since a charge-charge pair potential is often included in models used to study materials that are usually considered as covalent.

An ability to model the fine structure of dislocation cores in these materials should open up diverse new avenues of research. For example the plastic deformation of crystalline solids is often controlled by the glide and climb of edge dislocations (Poirier, 1985), processes which are directly controlled by the structure of the dislocation core. Crystal growth is often enhanced by the emergence of screw dislocations at the crystal surface; a process that results in the formation of characteristic growth spirals and, in some materials, can lead to interesting long period polytypism (Nabarro, 1967). The presence of line defects in crystals can also open pathways for the rapid diffusion of point defects (Frost and Ashby, 1982), this has been measured in MgO (Narayan and Washburn, 1972).

The major problem when simulating dislocations is that all of the atoms in the crystal containing the dislocation are significantly displaced by its presence; the strain energy associated with the dislocation decays only logarithmically with distance from the dislocation line (Hull and Bacon, 1984). This means that modelling approaches used to study point defects and surfaces, where the atomic structure at long range from the defect is only slightly perturbed from, or
is taken to be identical to, the structure of the perfect crystal, is not immediately available. Before describing the methodology used here to overcome this problem, previous work in the field is reviewed. The final sections of the chapter outline several applications of the methodology to some “complex” ionic materials including MgO, which forms ~10% of the Earth’s lower mantle, and zeolite A, where the dislocation core forms a molecular helter-skelter. Examination of the structure of the cores of dislocations in forsterite is reserved for chapter 7.

6.1 Previous work

Some of the earliest work modelling dislocations, dating from the early 1970's, was performed on metals. An example is the work on α-iron (Sinclair, 1971; Gehlen et al., 1972; Sinclair et al., 1978) which also provided a convenient test bed for new methodologies. These studies used a similar approach to that described here, with a straight dislocation in a simulation cell with 1-D periodicity and a two-region approach to the calculation of dislocation energy. Various boundary conditions were developed to account for the behaviour of the crystal away from the dislocation core. If the long range behaviour can be found exactly before the simulation commences, and the non-linear core terms decay sufficiently rapidly, then rigid boundary conditions are appropriate, otherwise one of the flexible boundary conditions, e.g. Flex-I (Gehlen et al., 1972), Flex-S (Sinclair, 1971) or Flex-II (Hirth, 1972; Sinclair et al., 1978), must be used. Despite these successes, at the time of these studies interatomic potential models for metals were poorly developed. In contrast, potential models for simple ionic materials were well developed and generally reliable. Much of the early work was therefore performed on this class of materials.

21 Briefly, Flex-I involves calculating the forces applied on the atoms of the boundary region in order to move the atoms according to the solution of the displacements for a non-linear elastic body. These solutions are available for the isotropic case as used by Gehlen et al. (1972), but not for anisotropic systems. Flex-S expands the non-linear elastic displacement field as a Fourier series and splits the simulation into three regions; the forces on the middle region are used to calculate the coefficients of the expansion in order to calculate the displacements in the outer two regions. Flex-II and the more modern approach of Rao et al. (1998) solves the non-linear problem using Green’s functions. All of these approaches involve minimisation of the energy of the inner atomistic region then fixing this and moving the atoms that represent the elastic continuum, these two steps must be repeated until self-consistency is achieved.
(Hoagland et al., 1976; Puls and Norgett, 1976). Of particular note is the work undertaken by Puls and co-workers who developed the PDINT code for the simulation of dislocations in cubic ionic materials. Concentrating on MgO as a model system, this group first used a simple shell model with rigid boundary conditions to calculate the Peierls energy for the \((a/2)[110]\) edge dislocation (Puls and Norgett, 1976; Woo and Puls, 1976). A breathing shell model was then used with the Flex-II boundary conditions to recalculate the geometry of the dislocation core (Woo and Puls, 1977a) and to re-evaluate the Peierls energy barrier (Woo and Puls, 1977b). The code was then developed further in order to model point defect – line defect interactions, again in MgO (Puls et al., 1977; Puls, 1980, 1983). Further work involved comparisons between the behaviour of MgO, NaCl and NiO (Rabier and Puls, 1989; Rabier et al., 1990).

More recently, advances in interatomic potentials for metallic systems coupled with new flexible boundary conditions (Rao et al., 1998; Ohsawa and Kuramoto, 1999) have lead to work on technologically important metals. These include studies of dislocations in tantalum (Yang et al., 2001a, 2001b), iron (Rao et al., 1998) and aluminium (Fang and Wang, 2000). Fixed boundary conditions were used by Xu and Moriarty (1996) to study screw dislocations in molybdenum. Much of the recent atomic simulations of dislocations in metals have been focused on linking an atomic scale model of the dislocation core with a “bottom up” (Harding, 1997) mesoscopic description of plasticity and strength (Bulatov and Kubin, 1998; Li et al., 2003a). A pertinent example is the simulation of cross slip in copper (Rasmussen et al., 1997a; Rasmussen et al., 1997b) and recent simulations on very large systems of iron, which suggest a new picture of plasticity controlled by the cross slip of screw dislocations (Marian et al., 2004).

Work has also been undertaken that allows special 3-D cell geometries to be constructed in which the net force on the core of the dislocation from its periodic images is cancelled (Bigger et al., 1992; Cai et al., 2001). This has allowed for the use of standard plane-wave codes for the simulation of dislocations in metals such as molybdenum and tantalum (Ismail-Beigi and Arias, 2000) and in semi-conductors, where the suppression of dislocations is of interest to the computer chip industry. Studies of the structural and electronic
structure of dislocations in semiconductors include work on edge (Bigger et al., 1992; Liu et al., 1995) and screw (Cai et al., 2001) dislocations in silicon, studies of an extended defect structure in layered Si/Ge crystals (Mostoller et al., 1994) and the incorporation of hydrogen (Heggie et al., 2000) and other impurities (Kaplan et al., 2000) into edge dislocations. Comparisons of the structure of dislocations in silicon and carbon (diamond) have also been made (Heggie et al., 2000; Ewels et al., 2001; Heggie et al., 2002; Martsinovich et al., 2003). A limitation of using periodic boundary conditions is that the energy of the simulation cell with respect to a reference ideal crystal includes terms for dislocation – dislocation interactions analogous to defect – defect interactions discussed in chapter 2. While these can be calculated and corrected for (Cai et al., 2001) it is not at all clear that this technique is tenable when applied to ionic systems. This is because as well as the elastic interactions there will inevitably be electrostatic interactions between dislocation cores, which are expected to develop dipoles due to the rearrangement of their atoms.

Compared to studies of metals and semiconductors recent interest in ionic systems has been rather limited (see the review by Harding, 1990) with the with the notable exception of Watson and co-workers (see section 6.3). This group, using the METADISE code (Watson et al., 1996), were able to study screw dislocations in MgO (Watson et al., 1999) and the effect of these dislocations on the MgO [100] surface (Watson et al., 2001). The work of Sayle is also worth highlighting (Sayle, 1999), in this work one oxide is “deposited” on to the surface of another on the computer and, in some cases, edge dislocations are seen to form spontaneously in order to relieve strain caused by incompatibilities in the lattice parameter of the two materials.

Despite this large body of work there appears to have been no attempts to model dislocations in ionic materials, such as silicate minerals, zeolites and non-cubic ceramics that are structurally complex. This is despite the importance that dislocations must play in the behaviour of these materials and the regular application of computational models to understand the point defect and surface properties of complex systems. The major problems to be overcome are that the systems are generally of low symmetry, so anisotropic elasticity must be used. A further complication is that a typical simulation cell is likely to have a charge
or dipole across it. The following section outlines a methodology to address these issues.

6.2 Methodology

In this section the methodology used to find the structure of the core of edge and screw dislocations, and the energy cost associated with introducing them into an otherwise perfect ionic crystal are described. In general, this problem is complicated by the fact that dislocations can be curved (i.e. the dislocation line is not straight), can bifurcate, may be moving and their Burgers vector can vary with position along the dislocation line. The dislocation model adopted here follows most of those reviewed above and simplifies this situation by constraining the dislocation line to be straight and stationary with a constant Burgers vector, this allows the dislocation to be modelled in a cell that is periodic in one dimension (along the dislocation line), perpendicular to the dislocation line the modelled crystal is considered infinite and non-periodic. However, by considering the part of the crystal a long distance from the dislocation line to be under relatively small strain and using methods appropriate to the description of an elastic continuum, a two region approach to the calculation of dislocation formation energy can be constructed and the structure of the core can be found with a finite number of atoms in the simulation cell. The methodology is described in detail below; first the methods used to set up the simulation cell for edge and screw dislocations are described, this is followed by details of the actual simulation (which amounts to an energy minimisation problem, in this case performed using the latest revision of the GULP code (Gale and Rohl, 2003)), finally techniques used to analyse the properties of the resulting defect structure are considered.

6.2.1 The simulation cell

The first task is to construct a correctly orientated atomistic simulation cell of the appropriate dimensions. The cell is chosen to have a circular cross section approximately centred on the origin of the dislocation. The convention used
This work is for the cell to be periodic along the Cartesian z axis\textsuperscript{22} and non periodic along the other two orthogonal axes (x and y) where it is to be embedded into an elastic model of the infinite crystal. The dislocation line therefore lies along z. Screw dislocations must then have a Burgers vector parallel to z and edge dislocations have a Burger’s vector in the xy plane\textsuperscript{23}. The exact orientation of the cell with respect to the cell parameters is chosen to simplify the introduction of the dislocation (Steeds, 1973; Steeds and Willis, 1979) as described below. In cases where the crystal described by the chosen set of interatomic potentials is elastically anisotropic, it is necessary to calculate the elastic compliance tensor ($s_{im}$) for the bulk material expressed on the Cartesian frame of reference described above. The simulation cell may also need to be terminated in such a way to ensure that it is charge neutral and has no net dipole.

The requirement that the cell is charge neutral with no dipole is due to the Coulomb summation schemes described below. In practice, this is achieved by centring the cell on an axis of rotation parallel to the z axis and (if the cell does not consist entirely of charge neutral strings of atoms parallel to z) by building the cell out of “charge neutral units”. This is exactly analogous in one dimension to the zero dimensional approach used by Braithwaite \textit{et al.} (2002) in their study of point defects in forsterite.

### 6.2.2 Introducing the dislocation

Once the bulk-like cell is generated, the second step is to introduce a dislocation into it. In order to achieve this, the following virtual operations are performed on the perfect cell:

1. a small amount of material from the centre of the cell is removed,

2. a cut from the edge of the cell to the hole in the centre is made,

\textsuperscript{22} This differs from the convention used in GULP where the periodic repeat is along x. The transformation between the two reference frames is trivial, and the reference frame used here is much more convenient as it coincides with that chosen by Steeds.

\textsuperscript{23} If the crystal unit cell is not cubic, tetragonal or orthorhombic then, for some pure edge dislocations, the Burgers vector is not perpendicular to the dislocation line. This has the interesting effect; the atomic displacements associated with the introduction of, say, a screw dislocation along z include displacements along x and y, the displacements cannot be separated into edge and screw components.
3. the two sides of the cut are displaced by the Burgers vector such that the two sides remain parallel,

4. the cut is cemented together and the system is allowed to reach elastic equilibrium.

This procedure is shown diagrammatically in figure 6.1. In the present work, this deformation is not performed step-wise in the manner described, as this would require an iterative, finite element approach. Instead, a mathematical function that maps the location of a point in the bulk cell to the equivalent point in the dislocated cell is applied. This function is the solution to the linear elastic displacement field of the dislocation and is only a function of the Burgers vector, orientation of the dislocation and elastic properties of the crystal. Such a function can be exactly correct a large distance from the core but is only approximate close to the core. Subsequent energy minimisation, with atoms explicitly considered, corrects the geometry of the core and its surroundings.

Volterra provided a general, analytical solution to the displacement field needed to introduce the dislocation into an elastically isotropic material (Nabarro, 1967). In the case of a pure screw dislocation, in the frame of reference described above, the expression is trivial (Hull and Bacon, 1984):

\[ u_x = u_y = 0, \]

\[ u_z = \frac{b}{2\pi} \arctan \frac{y}{x}. \]  

Any point \((x, y)\) simply moves parallel to the dislocation line, such that a 360° rotation about the line corresponds to a displacement of 1 Burgers vector in a smooth fashion. In the case of a pure edge dislocation in an isotropic material the solution is more complex, it is given by Nabarro (1967) for the case of the Burgers vector along \(x\) as:

\[ u_x = \frac{b}{4\pi(1-v)} \left( \frac{xy}{x^2 + y^2} + \frac{b}{2\pi} \arctan \frac{x}{y} \right) - \frac{(1-2v)b}{8\pi(1-v)} \ln \frac{x^2 + y^2}{b^2} + \frac{b}{4\pi(1-v)} \frac{y^2}{x^2 + y^2}, \]

\[ u_y = 0, \]

where \(v\) is Possion’s ratio.
Figure 6.1: cartoon showing the process used to introduce a screw dislocation into a perfect crystal by Volterra’s procedure. The diagram shows a screw dislocation but an equivalent process occurs for an edge dislocation. (a) Represents the bulk-like cell generated by multiplication of the crystal unit cell, in (b) a portion of the crystal is removed, in (c) a cut is made and in (d) the two surfaces are moved past each other. In order to introduce an edge dislocation the movement in (d) would be at right angles to that shown. Volterra’s linear elastic displacement field maps points in (a) onto the equivalent point in (d) if the hole in the centre of the crystal is sufficiently large and there is no traction on its walls.
Unfortunately, crystals are very rarely elastically isotropic bodies; in order to be so the non zero elements in the elastic compliance matrix must be of the form:

\[ S_{11} = S_{22} = S_{33}; \]
\[ S_{12} = S_{13} = S_{23}; \]
\[ S_{44} = S_{55} = S_{66} = 2(S_{11} - S_{12}). \]

This means that, for almost all crystals, in order to find the correct displacement field needed to move an atom from its location in the perfect crystal into its location in the dislocated crystal, we must turn to the anisotropic elastic theory.

In the general case, with 21 independent elastic constants, analytical solutions are not readily available (Steeds and Willis, 1979). However, several approaches have been suggested (Stroth, 1958; Hirth, 1972; Asaro et al., 1973). In particular, symmetry simplifies the problem and several analytical solutions have been presented (Steeds, 1973). In the simplest case, where the dislocation is straight, \( z \) lies along a 6-, 4- or 2-fold axis and another 6-, 4- or 2-fold axis is perpendicular to it\(^{24}\) (chosen to lay along \( x \) or \( y \)) the analytical solutions equivalent to equations 1 and 2 are given by Steeds (1973) as follows. First, the elastic compliance matrix is converted into a reduced form (equation 6.4) in order to make use of the fact that the dislocation line is straight and there is no strain parallel to it:

\[ S_{lm} = S_{lm} - \frac{S_{3l}S_{3m}}{S_{33}}. \]

\(^{24}\) The presence of mirror planes can also lead to the same solutions for the anisotropic displacement field as a mirror plane can have the same effect on the elastic tensor as a rotation axis. The solution is valid if, on the axes given the matrix representation of the elastic constants tensor reduces to the form:

\[
\begin{bmatrix}
  c_{11} & c_{12} & c_{13} \\
  c_{22} & c_{23} & 0 \\
  c_{33} & 0 & 0 \\
  c_{44} & 0 & 0 \\
  c_{55} & 0 & 0 \\
  c_{66} & &
\end{bmatrix}
\]

or a more symmetrical form. Various combinations of 4- and 6-fold axes also have the effect of generating a matrix of this form (see Nye, 1957; Steeds, 1973).
The expression for the linear elastic displacement field for a screw dislocation is then:

\[ u_x = 0, \]

\[ u_y = 0, \]  

\[ u_z = \frac{b}{2\pi} \arctan \left( \frac{S_{45}}{S_{55}} \frac{y}{x} \right). \]  

(6.5)

This expression reduces to the isotropic case (equation 6.1) when \( S_{44} = S_{55} \). Otherwise, the effect is to change the pitch of the screw dislocation with the angle around the dislocation line.

For edge dislocations, the Burgers vector need not lie along one of the axes of the Cartesian frame of reference because the frame is chosen according to the symmetry of the crystal. First, the components of the Burgers vector along \( x \) and \( y \) (\( b_x \) and \( b_y \)) are calculated. Then, defining the anisotropic parameters:

\[ \lambda^4 = \frac{S_{22}}{S_{11}}, \]  

\[ 2\lambda = -\frac{2S_{12} + S_{66}}{S_{11}}, \]  

the displacement field for each component can be found. For the \( x \) component of the Burgers vector, with \( \lambda^4 < \Lambda^2 \), the expressions for the atomic displacements are:

\[ u_x = \frac{b_x}{4\pi} \left[ \arctan \left\{ \frac{2(\lambda^2 - \Lambda^2)\tan \theta}{1 - \lambda^2 \tan^2 \theta} \right\} - \left( \frac{S_{12}}{S_{11}} \right) \arctan \left\{ \frac{\sqrt{2(\lambda^2 - \Lambda^2)}\tan \theta}{1 + \lambda^2 \tan^2 \theta} \right\} \right], \]  

\[ u_y = \frac{b_x}{4\pi} \left[ -\frac{(\lambda^2 + S_{12}/S_{11})}{\sqrt{2(\lambda^2 - \Lambda^2)}} \ln r^2 \sqrt{\cos^4 \theta + \lambda^4 \sin^4 \theta - \frac{1}{2} \Lambda \sin^2 2\theta} \right] \]  

\[ + \frac{(\lambda^2 - S_{12}/S_{11})}{2\sqrt{-2(\lambda^2 + \Lambda^2)}} \left\{ \cos^2 \theta - \left( \Lambda - \sqrt{\Lambda^2 - \lambda^4} \right) \sin^2 \theta \right\}, \]  

(6.7)

\[ u_z = 0. \]
the \( y \) component of the Burgers vector gives:

\[
u_x = \frac{b_x}{4\pi^2} \left[ \frac{\left( \hat{\lambda}^2 + S_{zz}/S_{11} \right)}{\sqrt{2(\hat{\lambda}^2 - \Lambda)}} \ln r^2 \sqrt{\cos^4 \theta + \Lambda^4 \sin^4 \theta - \frac{1}{2} \Lambda \sin^2 2\theta} 
+ \frac{\left( \hat{\lambda}^2 - S_{12}/S_{11} \right)}{2\sqrt{-2(\Lambda + \hat{\lambda}^2)}} \ln \left\{ \frac{\cos^2 \theta - \left( \Lambda - \sqrt{\Lambda^2 - \Lambda^4} \right) \sin^2 \theta}{\cos^2 \theta - \left( \Lambda + \sqrt{\Lambda^2 - \Lambda^4} \right) \sin^2 \theta} \right\} \right],
\]

(6.8)

\[
u_y = \frac{b_y}{4\pi} \left[ \arctan \left\{ \frac{\sqrt{2(\hat{\lambda}^2 - \Lambda) \tan \theta}}{1 - \Lambda^2 \tan^2 \theta} \right\} + \frac{(\Lambda + S_{12}/S_{11})}{\sqrt{\Lambda^2 - \Lambda^4}} \arctan \left\{ \frac{\sqrt{-2(\Lambda + \Lambda^2) \tan \theta}}{1 + \Lambda^2 \tan^2 \theta} \right\} \right],
\]

\[
u_z = 0.
\]

In the case \( \Lambda^4 > \Lambda^2 \), the expressions for the \( x \) components are:

\[
u_x = \frac{b_x}{4\pi} \left[ \arctan \left\{ \frac{\sqrt{2(\hat{\lambda}^2 - \Lambda) \tan \theta}}{1 - \Lambda^2 \tan^2 \theta} \right\} 
+ \frac{(\Lambda + S_{12}/S_{11})}{\sqrt{\Lambda^4 - \Lambda^2}} \ln \left\{ \frac{\cos^2 \theta + \Lambda^2 \sin^2 \theta - \sqrt{\Lambda^4 + \Lambda^2}}{2} \sin 2\theta}{\cos^2 \theta + \Lambda^2 \sin^2 \theta + \sqrt{\Lambda^4 + \Lambda^2}} \sin 2\theta} \right\} \right],
\]

(6.9)

\[
u_y = \frac{b_y}{4\pi} \left[ \frac{\left( \hat{\lambda}^2 + S_{zz}/S_{11} \right)}{\sqrt{2(\hat{\lambda}^2 - \Lambda)}} \ln r^2 \sqrt{\cos^4 \theta + \Lambda^4 \sin^4 \theta - \frac{1}{2} \Lambda \sin^2 2\theta} 
+ \frac{\left( \hat{\lambda}^2 - S_{12}/S_{11} \right)}{\sqrt{2(\Lambda + \hat{\lambda}^2)}} \arctan \left\{ \frac{\sqrt{\Lambda^4 - \Lambda^2}}{\cot^2 \theta - \Lambda} \right\} \right],
\]

\[
u_z = 0,
\]

172
and the $y$ components become:

$$
u_z = \frac{b_x}{4\pi}\left[\frac{\left(\lambda^2 + S_{11}/S_{11}\right)}{\sqrt{2}\left(\lambda^2 - \Lambda\right)}\ln r^2 \sqrt{\cos^4 \theta + \lambda^4 \sin^4 \theta - \frac{1}{2}\Lambda \sin^2 2\theta}
+ \frac{\lambda^2 - S_{12}/S_{11}}{\sqrt{2}\left(\lambda^2 + \Lambda^2\right)} \arctan \left\{\frac{\sqrt{\lambda^4 - \Lambda^2}}{\cot^2 \theta - \Lambda}\right\}\right],$$

$$u_y = \frac{b_x}{4\pi}\left[\frac{\lambda^2}{\sqrt{\lambda^4 - \Lambda^2}} \arctan \left\{\frac{\sqrt{2(\lambda^2 - \Lambda) \tan \theta}}{1 - \lambda^2 \tan^2 \theta}\right\}\right] - \frac{(\Lambda + S_{11}/S_{11})}{\sqrt{\lambda^4 - \Lambda^2}} \ln \left\{\frac{\cos^2 \theta + \lambda^2 \sin^2 \theta - \sqrt{\frac{\Lambda + \lambda^2}{2} \sin 2\theta}}{\cos^2 \theta + \lambda^2 \sin^2 \theta + \sqrt{\frac{\Lambda + \lambda^2}{2} \sin 2\theta}}\right\},$$

$$u_z = 0.$$  

Derivations of equations 6.5 – 6.10 can be found in Steeds (1973). Analytical solutions for the displacement fields given by the anisotropic linear elastic theory are also available for the case of dislocation lines lying along a 3-fold axis with a perpendicular 2-fold axis and the case where the dislocation line is perpendicular to a 2-fold axis. Application of the above displacements to the atoms in the bulk-like cell result in the introduction of a dislocation, the atoms at a large distance from the dislocation line should be in the correct geometry but further work is required to find the structure of the dislocation core. Care must be taken in applying the elastic displacements, the elastic constants that are used must be those generated by potential model, the trigonometric functions must be used so as not to give discontinuities in the displacement field for angles of $\theta$ between 0 and 360°. It is not advisable to break up any molecular ions in the system by locating the origin of the displacement field (which inevitably includes a discontinuity) close to chemical bonds; if such a configuration is unavoidable then the displacement field must be smoothed across the core. A set of programs to set up the dislocation cell have been produced, they are outlined in Appendix 1.
6.2.3 Finding the geometry of the dislocation core

The method described in the previous section allows the generation of a simulation cell containing a dislocation. However, all the expressions used to introduce the dislocation (equations 6.1 to 6.10) assume that the displacements of the simulation cell can be described by linear elastic theory (the hole in figure 6.1 is, in practice, made infinitesimally small). There are three causes of error in the location of the atoms in the starting cell:

1. The displacements are taken to correspond to linear elastic theory; this is equivalent to assuming that the strains (i.e. the relative displacement of the atoms to one another) are small. This assumption is clearly invalid close to the dislocation origin where the strains can be very large. Non-linear effects are found to lead to an experimentally detectable increase in the crystal volume when the density of dislocations increases (Gairola, 1979).

2. The structure of the core is, by definition, not known from elastic theory. The elastic description of the dislocation has a singularity in energy at $r = 0$, and the atomic displacements are discontinuous. The method assumes that the structure of the core can be found from the starting approximation given by linear elastic theory.

3. Reconstruction of the crystal structure of the dislocation core may well be accompanied by a dilation, contraction or twisting of the core. This in turn applies forces from the core onto the surrounding crystal, which responds elastically. This effect is obviously not accounted for in setting up the cell.

Several solutions are available to this problem; if the structure of the core itself is not a concern, but its effect is, then non-linear elastic theory can be applied (see, for example, Gairola, 1979), if however the structure of the core is of interest then its constituent atoms must be explicitly treated. This is a particular problem since the geometry of the core depends on the position of the atoms surrounding the core and the position of the atoms around the core depends in turn on the geometry of the core. Fortunately, there is a solution to this problem: the position of atoms a long distance from the core can be found using
elastic methods given only a knowledge of the elastic constants, the direction of
the dislocation line and the Burgers vector. If enough atoms can be simulated
so that the non-linear elastic terms (associated with the large strains near to the
core) and the core dependent effect decays before the edge of the free atomistic
region is reached then so called fixed boundary conditions can be used. In such
a simulation, a large portion of the crystal around the dislocation core is relaxed
to minimise the energy of the whole cell while the atoms near the edge of the
simulation cell are held fixed at the positions determined by linear elasticity.
This is the approach used in this work. A check that the size of the relaxed
portion of the crystal is big enough to accommodate the non-linear and core
effects is to calculate the energy derivatives on the innermost atoms that are
held fixed during the relaxation. If these values are small then the free region is
big enough. In cases where not enough atoms can be included in the simulation
to meet this condition then one of the flexible boundary conditions (Sinclair,
1971; Gehlen et al., 1972; Sinclair et al., 1978; Rao et al., 1998) can be used. A
Green’s functions method such as Flex-II could also be used in the case where
an exact solution to the linear elastic problem is not available as the position of
the atoms in the elastic region would be expected to move towards their correct
location as determined by anisotropic elastic theory (R. G. Hoagland, personal
communication).

The methodology described so far is quite general and does not depend on the
form of the interatomic potential used. However, in order to model ionic or
partially ionic systems the cell described above must contain charged ions,
which interact via a $q_i q_j/r$ Coulombic potential. In principle this is not a problem
since in one dimension (in contrast with the two or three dimensional cases) the
Coulomb summation is absolutely, if slowly, convergent (Gale and Rohl, 2003).
The slow convergence makes a simple real space summation inappropriate for
these terms. However, several methods are now available which make the
problem tractable. In this work the Coulomb summations of either Saunders et
al. (1994) or of Wolf et al. (1999) are used. Both have been implemented in the
latest revision to the GULP code (Gale and Rohl, 2003), the choice of
summation method depends on the size of cell required.
For cells with a smaller radius the Saunders summation (Saunders et al., 1994), originally developed for the simulation of polymers is preferred. This summation is superior to that used by, for example, Hoagland et al. (1976) or Puls and Norgett (1976). Firstly, because the new approach does not require strings of atoms with a net charge of zero parallel to \( z \), and secondly because Watson et al. (1999) report that the new summation exhibited faster convergence. As the cell radius increases, the Saunders summation becomes increasingly inefficient. A better approach for these large cells is the real-space Wolf summation (Wolf et al., 1999). This has the advantage of allowing linear scaling with number of atoms in the simulation cell. However, the potential model should be fitted taking account of the fact that the Wolf summation is to be used. On a related note, the Madelung potential should be correct for all the atoms allowed to move during the simulation. This can easily be achieved by making the shell of atoms embedded in the elastic continuum thick enough to achieve convergence in dislocation energy.

In short, the simulation itself involves taking a cell containing a dislocation with the atoms near the centre of the cell incorrectly positioned. This is remedied by fixing the atoms around the edge of the cell (which are correctly positioned) and minimising the energy of the cell.

The current methodology introduces an additional approximation, as well as fixed boundary conditions, when introducing the dislocation. The strain of each crystal unit cell around the dislocation is finite and can be quite large, in general large strains of a unit cell result in relative displacement of the ions in it (except in cases where the atoms cannot move and preserve the cell’s symmetry). If the crystal is made up of covalently bonded molecules, one would expect the bonds within the molecules to be relatively unstrained while most of the strain of the cell is taken up in changing the molecules’ separation and orientation. The methodology outlined above results in all atomic separations being treated in the same way, the strain in the cell is homogeneous. Examples of the result of this approximation are described below (and in chapter 7) and possible solutions short of introducing flexible boundaries are discussed in chapter 8; the approximation introduces a discontinuity between the fixed and relaxed regions.
but it seems that this error is localised and does not affect the structure of the dislocation core.

### 6.2.4 Analysing the results

Once the geometry of the dislocation core and surrounding crystal has been established, physically meaningful parameters can be extracted from the model. The first point to note is that total energy of the simulation cell is not a particularly useful quantity. This is because this energy depends on the size of both the free and fixed portions of the simulation cell as well as on the exact method used to terminate the simulation cell. In contrast, the displacement field of the dislocation is of relevance. This is defined as the change in position of the atoms in the relaxed cell compared to the atoms' position in the bulk crystal (or the bulk like 1D cell). This vector field can be split into two components, the first is the linear elastic displacement field, which is already known; it was found analytically before the simulation (equations 6.1 – 6.10). The second component is the deviation of the linear elastic displacement field from the total displacement field caused by the three processes outlined above. This part will be termed the core displacement field (although not all contributions to it are due to the presence of the core); there are equivalent stress and strain fields.

A second property of interest is the “dislocation formation energy”, defined as the work done (per unit length of dislocation) by introducing the dislocation into an otherwise perfect crystal (Hull and Bacon, 1984). Expressing this is not a simple matter, as this energy does not converge with the size of the simulation cell (in contrast with models of point or planar defects). A quantity that can be expressed is the dislocation energy stored within a certain radius of the dislocation origin. The basic theory of dislocations splits this energy into two components (Hull and Bacon, 1984), the first stored within the dislocation core and the second stored by the elastic response of the rest of the crystal. As an example, in an elastically isotropic material the energy stored by a screw dislocation within a cylinder of radius \( r \) centred on the dislocation line is described by:

\[
E(r) = E(\text{core}) + \frac{Gb}{4\pi} \ln \left( \frac{r}{r_0} \right). \tag{6.11}
\]
$E_{\text{core}}$ is the energy stored within the core which has radius $r_0$, $b$ is the length of the Burgers vector and $G$ is the shear modulus. Similar expressions exist for anisotropic crystals replacing $G$ with another combination\textsuperscript{25} of the elastic constants (Steeds, 1973). $E(r)$ can be evaluated for values of $r$ directly from the atomistic model of the dislocation, so that many values of $E(r)$ can be found and these used to fit to equation 6.11. The core radius and energy can then be extracted. In order to evaluate $E(r)$ the simulation cell, with its atoms at an energy minimum, is split into two parts (figure 6.2). Region 1, centred on the origin of the dislocation, is circular with a radius $r$, region 2 is the rest of the cell. The total energy of the cell can now be split into three terms:

$$E_{\text{total}} = E_{11} + E_{12} + E_{21} + E_{22},$$

(6.12)

where $E_{11}$ is the total interaction energy between all pairs of ions in region 1, $E_{12}$ is the interaction energy between atoms in region 1 and atoms in region 2, $E_{21}$ is the interaction energy of atoms in region 2 with atoms in region 1 and $E_{22}$ is the interaction between pairs of atoms within region 2. $E(r)$ is then given by:

$$E(r) = E_{11} + E_{12}.$$  

(6.13)

The dislocation energy can be found by subtracting this value from the equivalent value found for the bulk cell.

\textsuperscript{25} This “energy factor”, which depends on the symmetry of the elastic constants tensor (and therefore the point group of the crystal) in a similar manner to the elastic displacement field, is derived for a number of point groups by Steeds (1973). See section 6.3.3 for an example.
Figure 6.2: cell setup for the evaluation of $E(r)$. The radius of region 1 is chosen to be equal to $r$, region 2 must then be sufficiently large to correctly yield region 1 – region 2 interaction energies. This limits the maximum value of $r$ that can be used to evaluate $E(r)$ for a particular simulation cell.

6.3 Dislocations in MgO

The mineral periclase (MgO) crystallises with the simple face centred cubic rock salt structure (space group $Fm\bar{3}m$). As the archetypical oxide, MgO has been subject to a vast number of computational studies and is the classic test system for models of ionic materials. There is also geophysical interest in the properties of MgO; as an oxide with one of the widest stability fields it is expected to form ~10% of the Earth’s lower mantle (Davies, 1999) alongside the perovskite phases. At this concentration, MgO could influence the large-scale properties of the lower mantle.

Although it has a simple structure, some of the properties of MgO are not easy to model. In particular, the elastic constants violate Cauchy’s relation, which states that, for simple cubic materials with central pair wise forces acting between ions, $c_{12}$ must be equal to $c_{44}$. This leads to the conclusion that there are
significant many body interactions in MgO or there are off-centred forces acting between pairs of ions. Parameterised potential models must consider this effect if they are to reproduce the elastic constants tensor of MgO.

The structure of MgO leads to several important slip systems, easy slip is observed in \( \{1\overline{1}0\}\{110\} \) and slip on \( \{0\overline{1}1\}\{100\} \) is observed when single crystals are loaded along the \( \{11\overline{1}\} \) axis to suppress easy slip and in polycrystalline samples at high temperature (Frost and Ashby, 1982). The \( \{100\} \) screw dislocation, while apparently not important in allowing deformation, has been inferred to be important for crystal growth (Watson et al., 2001).

### 6.3.2 Potential model

There are a wide range of potential models of MgO including those which make use of the shell model to describe the oxygen ions (e.g. Lewis and Catlow, 1985) and those that make use of a breathing shell model (e.g. Catlow et al., 1976; Matsui, 1998; Matsui et al., 2000). Use of the shell model alone results in good reproduction of the dielectric behaviour but fails to correctly describe the elastic properties. However, breathing shell models can correctly reproduce the Cauchy violation and give the correct elastic behaviour. For this reason the breathing shell model of Gale and Rohl (2003) was used for studies of dislocations in MgO. The model, described in table 6.1, gives a good description of the structure and the elastic and dielectric behaviour of MgO as shown in table 6.2 (see also Gale and Rohl, 2003).

**Table 6.1:** potential parameters used for modelling MgO (see text). Magnesium ions are assigned a charge of +2 while the charge on the oxygen is separated into a core charge of +0.8 and a shell charge of −2.8, the breathing shell radius is fixed at 1.2 Å. Short range potentials are truncated at 10 Å.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>( A ) (eV)</th>
<th>( \rho ) (Å)</th>
<th>( C ) (eVÅ(^6))</th>
<th>( k_{2x} ) (eVÅ(^2))</th>
<th>( k_{BSM} ) (eVÅ(^4))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mg – Obshell</td>
<td>28.7374</td>
<td>0.3092</td>
<td>0.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Oshell – Oshell</td>
<td>0.0</td>
<td>0.3</td>
<td>54.038</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ocore – Oshell</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>46.1524</td>
<td>-</td>
</tr>
<tr>
<td>Oshell – Obshell</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>351.439</td>
</tr>
</tbody>
</table>
Table 6.2: calculated and experimental properties of MgO. Experimental results of Zha et al. (1997b) compared to calculations based on the breathing shell model (table 6.1) and the shell model (table 3.1). The comparison is perhaps unfair as the breathing shell model is fitted to the experimental results while the shell model is only fitted to the structure of MgO.

<table>
<thead>
<tr>
<th></th>
<th>a (Å)</th>
<th>c_{11} (GPa)</th>
<th>c_{12} (GPa)</th>
<th>c_{44} (GPa)</th>
<th>ε&lt;sup&gt;δ&lt;/sup&gt;</th>
<th>ε&lt;sup&gt;α&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breathing shell model</td>
<td>4.2122</td>
<td>297.13</td>
<td>95.09</td>
<td>155.73</td>
<td>9.886</td>
<td>2.936</td>
</tr>
<tr>
<td>Shell model</td>
<td>4.1990</td>
<td>392.84</td>
<td>164.21</td>
<td>164.21</td>
<td>8.380</td>
<td>2.172</td>
</tr>
<tr>
<td>Experiment</td>
<td>4.212</td>
<td>297.0</td>
<td>95.2</td>
<td>155.7</td>
<td>9.86</td>
<td>2.96</td>
</tr>
</tbody>
</table>

### 6.3.3 The a[100] screw dislocation

The \(\langle100\rangle\) screw dislocation, and its effect on growth of MgO has been considered by Watson and co-workers (Watson et al., 1996; Watson et al., 1999; Watson et al., 2001). This work utilised the METADISE code and a rigid ion model of MgO. The description of the elastic behaviour is therefore not accurate in these calculations but they do constitute a useful comparison for the method described above. The methodology adopted for the simulation of an isolated screw dislocation is subtly different from the methodology outlined above. The major difference is in the way the dislocation energy was calculated, Watson et al. (1999) performed a series of geometry optimisations with varying sizes of the free atomistic region surrounded by a fixed size shell of atoms held in the location predicted by linear elasticity. This is in contrast to our approach, and the approach of previous workers such as Puls and Norgett (1976) and Hoagland et al. (1976) where one geometry optimisation is performed then this geometry is used to calculate the energy. The result of this difference is that any core effects are tightly constrained in the smaller radius cells of Watson et al. (1999) so that if, for example, the core expands with respect to the bulk material then at small radii there will be a high effective pressure on the core and as the radius is allowed to expand the effect will decrease. This means that the extracted dislocation energy will be calculated at some effective pressure and that the elastic constants can not be used to derive the elastic part of the dislocation energy. The energy factor extracted by Watson et al. (1999) is a fitted parameter from the data rather than a fixed value from the elastic constants. However, the geometry of the core is not expected to
differ from that described below as long as a large cell is selected for the description of the structure. A third difference is that the simulation cells of Watson et al. (1999) are square perpendicular to the dislocation line and the results described below use a round cell.

Convergence testing was carried out in order to select sizes for the relaxed and fixed regions in the calculation. Figure 6.3 shows the effect of increasing the thickness of the fixed region while keeping the free region at a fixed radius. It is clear that a shell of ions 15 Å thick is easily large enough to converge the electrostatic energy of the free region and isolate it from the surrounding vacuum. Convergence of the free region is not so straightforward because of the logarithmic dependence of energy on system size. A convenient test is to increase the radius of the free region until the outer-most atoms do not move during the energy minimisation; the difficulty is that such testing rapidly becomes inefficient. After some initial testing with smaller cells, it was found that a free region of 50 Å radius was thick enough to isolate the non-linear effects of the core from the fixed elastic surrounding region (see figure 6.7, for example). Hence, the radius of region 1 was 50 Å and the radius of region 2 (from the centre of the displacement field) was 65 Å. There are three obvious possibilities for the location of the origin of the displacement field termed the edge, ion and centre sites shown in figure 6.4 (Watson et al., 1999), the cells centred on these sites contained 8994, 8943 and 9000 species (Mg cores, O cores and O shells) respectively. All the cells were 4.212 Å thick. Introduction of the dislocation is relatively straightforward in this case as the anisotropic displacement field (equation 6.5) reduces to the isotropic form (equation 6.1). Energy minimisation was performed using the conjugate gradients method (section 2.3.1) and the Saunders (1994) Coulomb summation used.
Figure 6.3: convergence of dislocation energy with thickness of the fixed portion of the simulation cell. The radius of the inner region is 25 Å for the purposes of this convergence tests and the reported dislocation energy is the difference in energy between a dislocated and bulk like cell stored within the inner region, ionic relaxation is not taken into account. It is clear that convergence is achieved relatively rapidly.

Figure 6.4: three locations used to centre the 1D simulation cell and the elastic displacement field for models of the <100> screw dislocation in MgO.
The final energies of the simulation cells are \(-13.33\) eV/ion in each of the three cases (compared to a bulk cell value of \(-13.35\) eV/ion). The dislocation energy (difference between the energy of the bulk and dislocated systems) as a function of radius is presented in figure 6.6. This energy is fitted to the elastic solution to the dislocation energy. Figure 6.6a shows the critical importance of accounting for anisotropy in this procedure, two of the three lines represent fits using the Voigt and Reuss definitions of the shear modulus (133.8 and 128.0 GPa respectively) with equation 6.11, the gradient of the line on the log plot clearly shows that the isotropic elastic solution does not fit the results of the atomistic solution. Allowing the shear modulus to become an additional fitted parameter, the procedure used by Watson et al. (1999), results in a much better fit as does fixing the "shear modulus" as the value of the energy factor as derived by Steeds (1973). Thus the equation used to fit is:

\[
E(r) = E(\text{core}) + \frac{\sqrt{3}s_\text{a}s_\text{b}}{4\pi} \ln\left(\frac{r}{r_0}\right),
\]

(6.12)

with the symbols introduced in equation 6.4 and 6.11. Taking the anisotropic energy factor results in very good fits with just one independent parameter (obviously the core energy depends on the core radius). Fixing the core radius at 10 angstroms and only fitting to the data points from larger region 1 radii yields core energies of 4.2213, 4.4408 and 4.1916 eV/Å for the edge, centre and ion sites respectively. This is the same ordering of energies found by Watson et al. (1999) despite the different approach.

Figure 6.5 shows the final structure of the core of the [100] screw dislocation (all three cells give essentially the same geometry). In analysing this structure, it is useful to compare the final structure with the starting structure as derived by linear elastic theory; figure 6.7 shows this difference, which is termed the core displacement field. The core relaxation can conveniently be separated into two parts with distinct causes. The first effect is that the atoms tend to move in towards the core (figure 6.7 a). This mode of displacement is because a group of atoms, which originally formed a circle around the incipient dislocation line, form a helix after the introduction of the dislocation. As the elastic displacement field does not contain components perpendicular to the dislocation line, the radius of the helix and the original circle are the same, so that all the
Mg – O bond distances are increased by the process of introducing the dislocation. On relaxation, these bond distances tend to reduce towards their natural values in the bulk, which is achieved by the inwards relaxation. The second mode of displacement is motion parallel to the dislocation line (figure 6.7 b). This occurs because of the discontinuity at the origin of the elastic displacement field (for example an atom a short distance along the positive x-axis has a value of θ of 0, a short distance along the negative y-axis gives a value of 180 degrees, the displacement along z for these two points are 0 and 0.5 times the Burgers vector respectively, crossing the origin leads to a step of 0.5 b in the displacement field). In MgO this discontinuity is equal to the Mg – O separation so that like atoms are brought into contact across the discontinuity. This is clearly energetically unfavourable so the atoms close to the core move along z to remove the discontinuity. It is notable that the total displacement in z is equal to 2.4 angstroms, or half the Burgers vector. The resultant structure is reminiscent of the structure of one-dimensional ionic crystals grown inside carbon nanotubes (Wilson, 2002; Philp et al., 2003).

**Figure 6.5:** Relaxed simulation cell containing <100> screw dislocation in MgO (Mg red, O green). The smaller spheres represent ions held fixed in the elastic region and the larger spheres represent atoms that are allowed to relax during the energy minimisation.
Figure 6.6: Energy of MgO screw dislocation. Blue, red and black points are taken from the atomistic model for the centre, ion and edge choices of origin. Solid, dashed and dash and dot lines are fits to the data for the edge site using the anisotropic, Voigt and Reuss definitions respectively (see text for details).
Figure 6.7 is continued over page.
Figure 6.7: relaxation from elastic to atomistic solution for the <100> screw dislocation in MgO (centre site). (a) Ionic displacements parallel to the dislocation line smooth out the discontinuity at the origin. (b) Ionic displacements inwards towards the dislocation line shorten the bond lengths. (c) Oxygen polarisation is rapidly screened out.

6.4 Screw dislocations in zeolite A

Zeolitic materials are used extensively within the petrochemical and fine chemical industries, where the structure of the crystal and its integrity are often vital to the catalytic efficiency of that material (Lai et al., 2003). This porous class of materials contains cages and channels of molecular dimensions that permit selective diffusion within the host material, thus leading to their widespread use as molecular sieves and catalysts. The channel diameter ranges from ~4-15Å in “microporous” materials to greatly in excess of 30Å in “mesoporous” materials, offering the possibility of extensive tenability for specific applications. Modelling dislocations in zeolites is a more challenging example than modelling dislocations in MgO because of the increased structural complexity. When zeolites are chosen for selectivity towards a particular molecular dimension it is often presumed that they can be viewed as regular pores with uniform dimensions, with the exception of any chemical substitution within the framework topology. However, there is also the possibility that the
very channel structure itself may be distorted and displaced by extended defects leading to significant changes in transport behaviour. It is of particular interest to interpret how transport and reactivity within nanoporous materials is affected by the presence of defects whose length scale varies from a few to several hundred angstroms.

Point defects, such as hydroxyl nests have been well studied in zeolitic materials (e.g. Sokol et al., 1998; Sokol et al., 2000) whilst exotic defects, such as co-incidence boundaries seen in Zeolite L (Terasaki et al., 1984), have been observed and their dramatic effect of transport retardation has been demonstrated. Other extended defects, including stacking faults and growth spirals, have previously been studied in the bulk and at the surface. Surprisingly, almost nothing is known about what effect dislocations have upon the properties of zeolites. In this example, this point is addressed by determining the atomic structure of a screw dislocation in zeolite A. Brief details of this study have previously been published in Nature Materials (Walker et al., 2004), the discussion presented below provides some additional information and shows the flexibility of the method described in the earlier part of this chapter.

6.4.1 Experimental background

The motivation to study zeolite A (LTA) was inspired by the exceptional recent high-resolution AFM observation of growth spirals on its \{100\} surface (Dumrul et al., 2002). These growth spirals (figure 6.8), clearly show the presence of screw dislocations emerging from the interior of the crystal, thus providing steps on the surface that can act as very favourable sites for growth. Unlike the case of an ‘ideal’ crystal surface, where steps are extinguished when a full crystal layer is formed, the step is continually regenerated as layer after layer of material is added to the surface. This process avoids the rate-limiting step in crystal growth, which is normally considered to be the addition of a growth unit to the perfect surface (Nabarro, 1967).
Figure 6.8: AFM image of growth spirals on the \{100\} surface of synthetic zeolite A taken from Dumral et al. (2002). The spirals are caused by growth at steps that are due to the emergence of screw dislocations at the surface. The steps are reported to be 1.2 nm high and the field of view is 4 μm across. Reproduced without permission.

Whilst it is clear that dislocations play a vital role in controlling the growth of materials such as LTA (Dumrul et al., 2002), heulandite (Binder et al., 1996) and presumably many other zeolitic materials, we do not know what effect they have on structure and reactivity. Although the technique of AFM can reveal the presence of dislocations at the surface it cannot yield atomic scale resolution of the structure of the dislocation core, and thus the properties of this defect cannot be readily inferred or deduced. However, using the simulation method outlined above yields atomic resolution and results that allow an analysis of how the dislocation core structure affects zeolite transport and reactivity.

Zeolite A, first synthesised and characterised by Breck and co-workers (Breck et al., 1956; Reed and Breck, 1956), has the structural formula \([\text{Na}_{12}(\text{H}_2\text{O})_{27}]_8[\text{Al}_{12}\text{Si}_{12}\text{O}_{48}]_8\) and was one of the first zeolites to be commercially
exploited as, for example, an ion-exchanger within clothing detergents. It continues to be manufactured on a massive scale worldwide and, because of its utility, it has inspired widespread experimental (Agger et al., 1998; Agger et al., 2001) and computational (Agger et al., 1998; Agger et al., 2001; Slater et al., 2001) investigation. The structure consists of a three-dimensional interconnecting network of channels, which permits transport through the crystal, and two cage structures defined by the aluminosilicate framework: the alpha cage shown in figure 6.9a and beta (or sodalite) cage shown in figure 6.9b. Diffusion through the structure occurs via the channel system, where the narrowest points within the accessible void space are at each of the largest faces of the alpha cage, at an eight-membered ring with a diameter 6.8 Å (figure 6.10a). The narrowest point provides the largest barrier to transport and hence determines the diffusion rate for the non-defective structure.

**Figure 6.9:** Topology of zeolite A based on the silica framework, (a) highlights the alpha cage and (b) highlights the beta or sodalite cage. The alpha cages, joined by 8-rings, form the 3 perpendicular sets of channels, the sodalite cages joined by double 4-rings are relatively impermeable. *First published under licence by McMillan (Walker et al., 2004).*

### 6.4.2 Modelling details

In order to simplify the description of zeolite A, a purely siliceous model was used. The potential parameters, fitted by J. D. Gale, are chosen so as to decay smoothly to zero at inter-atomic separations of 8 Å. The real-space Wolf summation (Wolf et al., 1999) is used for the Coulomb part of the potential.
model, thus allowing a spatial decomposition scheme to be used for the total
energy summations. Using this approach, the calculations scale linearly with
system size; further details and parameters of the potential model are given in
table 6.3. The simulation was set up by generating a large cylindrical cell of
height 24 Å and radius 115 Å. This cell contains ~68000 ion centres. Possible
finite size effects were examined using a cell of approximately twice this radius
(~250,000 centres, 215 Å radius), there were no signs that the relaxation was
constrained in the smaller cell. Charge neutrality is achieved by building the
cell from neutral SiO₄ units (assigning each oxygen a charge of −1). Oxygen
ions within the cell, that belong to two units, then have a formal charge of −2, as
they ‘belong’ to two SiO₄ units, while those at the edge of the simulation cell,
which only co-ordinate one silicon ion, have charge of −1. The cell is
terminated only with oxygen ions, and all silicon ions are fully co-ordinated.
Enforcing this condition also minimises the dipole moment across the cell. The
screw dislocation was then introduced assuming linear elasticity (again the
equation 6.5 reduces to the isotropic form) with the origin of the elastic
displacement field located along the centre on one of the 8-ring channels for all
the simulation atoms. Finally, the total energy of the simulation cell is
minimised by varying the positions of atoms within 100 (or 200, for the larger
cell) Å of the centre of the cell using a conjugate gradients algorithm.

Table 6.3: Potential parameters used for the atomistic model of zeolite A.

<table>
<thead>
<tr>
<th></th>
<th>Core charge</th>
<th>Shell charge</th>
<th>Spring constant (eVÅ²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>4</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>0.84820</td>
<td>-2.84820</td>
<td>74.9200</td>
</tr>
<tr>
<td><strong>Buckingham</strong></td>
<td>A (eV)</td>
<td>ρ (Å)</td>
<td>C (eVÅ⁶)</td>
</tr>
<tr>
<td><strong>potentials</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Si – O</td>
<td>1434.65230</td>
<td>0.312249</td>
<td>10.66000</td>
</tr>
<tr>
<td>O – O</td>
<td>22764.0000</td>
<td>0.149000</td>
<td>27.88000</td>
</tr>
<tr>
<td><strong>Three body</strong></td>
<td>k (eVrad⁻¹)</td>
<td>θ₀(°)</td>
<td>ρ₁ (Å⁻¹)</td>
</tr>
<tr>
<td><strong>exponential</strong></td>
<td></td>
<td></td>
<td>ρ₂ (Å⁻¹)</td>
</tr>
<tr>
<td>O – Si – O</td>
<td>0.862528×10⁸</td>
<td>109.4700</td>
<td>0.180000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.180000</td>
</tr>
</tbody>
</table>
Table 6.4: Calculated structure and properties for α-quartz, as determined by both the model of Sanders et al. (1984) and the potential parameters given in Table 6.3. The experimental data is included for comparison. The percentage deviations of the present model are comparable to, if not better, than those of the previous model. Similarly, application of both models to siliceous LTA, where there is no experimental data to compare against, yields almost identical results. The model has also been tested on the low symmetry zeolite silicalite, where it performs comparably with the Sanders et al. model and gives excellent agreement with experimental data.

<table>
<thead>
<tr>
<th></th>
<th>Experiment</th>
<th>Potential of Sanders et al.</th>
<th>Present work</th>
</tr>
</thead>
<tbody>
<tr>
<td>a (Å)</td>
<td>4.9021</td>
<td>4.8707</td>
<td>4.9176</td>
</tr>
<tr>
<td>c (Å)</td>
<td>5.3997</td>
<td>5.3778</td>
<td>5.3794</td>
</tr>
<tr>
<td>C_{11} (GPa)</td>
<td>86.8</td>
<td>95.3</td>
<td>90.5</td>
</tr>
<tr>
<td>C_{33} (GPa)</td>
<td>105.8</td>
<td>112.9</td>
<td>110.6</td>
</tr>
<tr>
<td>C_{44} (GPa)</td>
<td>58.2</td>
<td>50.0</td>
<td>46.7</td>
</tr>
<tr>
<td>C_{66} (GPa)</td>
<td>39.9</td>
<td>40.4</td>
<td>39.6</td>
</tr>
<tr>
<td>ε^0_{11}</td>
<td>4.54</td>
<td>4.58</td>
<td>4.22</td>
</tr>
<tr>
<td>ε^0_{33}</td>
<td>4.64</td>
<td>4.88</td>
<td>4.55</td>
</tr>
<tr>
<td>ε^{*}_{11}</td>
<td>2.4</td>
<td>2.07</td>
<td>2.01</td>
</tr>
<tr>
<td>ε^{*}_{33}</td>
<td>2.4</td>
<td>2.09</td>
<td>2.03</td>
</tr>
<tr>
<td>ν_{max} (cm^{-1})</td>
<td>1162</td>
<td>1116</td>
<td>1134</td>
</tr>
</tbody>
</table>

In the readily synthesised form of zeolite A, with a Si:Al ratio of 1:1, it has been shown that aluminium and silicon strictly ordered with alternate Si – O – Al bonding in agreement with Löwenstein’s rule (Bell et al., 1992). It follows that only every other beta cage is equivalent by translation and the structure takes on a face-centred cubic (FCC) structure (space group Fm 3c) with cell parameter 24.61Å, thereby avoiding Al – O – Al linkages, which have been demonstrated to be energetically unfeasible on this length scale (Bell et al., 1992). According to standard dislocation theory (Hull and Bacon, 1984) the Burgers vector of any dislocation should be an integer multiple of one of the primitive lattice vectors where the elastic energy of the dislocation is proportional to the square of the length of the Burgers vector. Applying this to FCC LTA suggests that the lowest energy Burgers vector will lie along <110> with length $\sqrt{\frac{1}{2}}a$ (where a is
the lattice vector parallel to the a axis of magnitude ~24 Å) followed by a
Burgers vector along <100> with a length equal to a (figure 6.10a). However,
in an ordered framework material there is a possibility that the Burgers vector
may be commensurate with the topological symmetry, rather than that of the
crystal unit cell. In this instance, which is rather analogous to ‘superlattices’ of
ordered solid solutions such as beta brass (Nabarro, 1967), the shortest, and
hence lowest energy, Burgers vector would be the \( \frac{1}{2} a \) <100> screw dislocation,
which has length ~12 Å, as shown in figure 6.10b. As the introduction of this
dislocation has the lowest energetic penalty (because it has the shortest Burgers
vector) this is the only model to be fully relaxed. However, single point energy
calculations of the \( 1 a <100> \) screw dislocation confirm that its energy is very
much higher than the \( \frac{1}{2} a \) dislocation model described here. In dense materials,
the dislocation core greatly distorts the structure, but one recognises that in very
porous materials the deformation origin can be sited in empty space, thus
reducing the energetic cost of introducing the defect. In LTA, the most obvious
location for the dislocation is at the centre of an alpha cage, where the vector
will naturally run along the centre of the eight-ring channels (figure 6.9a)
responsible for diffusion. After the dislocation is introduced, the eight-
membered rings at the interface between alpha cages form a helix along the
dislocation line, shown in figure 6.11c. Note that the height of the dislocation
reported by Dumrul et al. (2002) is 1.2nm, which corresponds exactly with \( \frac{1}{2} a \),
equal to the length of the Burgers vector. Regular step or terrace heights on
apparently non-dislocated samples have also been measured to be 1.2nm by
Agger and co-workers (Agger et al., 1998). Another significance of this
distance is that it corresponds with the height of the alpha cage. From
consideration of previous surface structure determinations (Slater et al., 2001),
it is anticipated that when the screw dislocation terminates at the surface, a
complete alpha cage will be expressed.
Figure 6.10: Two possible screw dislocations in the zeolite A structure viewed perpendicular to the dislocation line. (a) Shows the $a <100>$ dislocation centred on a 8-ring channel, two 8-rings form an interpenetrating double helix shown in red and blue with a pitch of 24 Å. (b) Shows the $1/2 a <100>$ dislocation where two 8-rings fuse to form a single helix with a 12 Å pitch. *First published under licence by McMillan (Walker et al., 2004).*

6.4.3 Results and discussion

After relaxation of the structure, the atoms closest to the core are displaced away from the channel centre by up to ~0.15 Å. The displacements of atoms relative to their original position follow an approximately oscillatory pattern (figure 6.12), where alternate expansion and contraction of the radial coordination shells is observed. The silicon atoms are more strongly perturbed than the oxygens, and the magnitude of the displacements attenuate quite slowly, but beyond 20-30 Å the radial location of channels will correspond almost exactly to those of the non-defective material. Along the dislocation line, the displacements are more extensive with a maximum magnitude of ~0.4 Å and again, the silicon atoms relax more than the oxygen atoms. The net effect of these displacements is shown in table 6.5, but briefly, the length of the central helical structure is longer than the original channel. The uniform elastic approximation used to generate the initial co-ordinates for the dislocation leads primarily to extension of the Si-O bond lengths. On relaxation, the Si-O-Si bond angles become large in order to reduce the bond length elongation.
Figure 6.11 is continued over page.
Figure 6.11: Views of three 8-ring structures in zeolite A. (a) An 8-ring in perfect zeolite A viewed along <100>, note presence of 4-fold rotation axis in centre of the ring. (b) 8-ring opening onto the central channel of the dislocated zeolite, the 4-fold axis is lost (and becomes a 2-fold axis) and the opening becomes significantly deformed. Assuming Burgers vector lies along [100] this is a view along [010] or [001]. (c) A perspective view down the central channel (along [100]) containing the dislocation core, the 8-ring structure is cut and fuses with adjacent 8-rings to form a continuous helix but the symmetry is maintained. Numbers and letters refer to oxygen and silicon atoms listed in table 6.5. First published under licence by McMillan (Walker et al., 2004).
Figure 6.12: Displacements perpendicular to (a) and along (b) the dislocation line during atomic relaxation. This represents the difference between the displacement field predicted by linear elasticity and that including core effects. Oxygen cores are in black and silicon atoms are red. *First published under licence by McMillan (Walker et al., 2004).*

Figure 6.13 highlights the central core of both the non-defective and dislocated material displaying the surface of the *inaccessible* volume of the material, *i.e.* that which is impermeable to small molecules, by rolling a probe across the framework. In figure 6.13a, which corresponds to looking down on the [100] surface, one can clearly see the perfectly regular channel systems. Similarly, in figure 6.13b, which corresponds to a cross-section through the surface, regular, circular channels are seen. Turning to the dislocated system, in figure 6.13d where the core atoms that comprise the eight-ring helix are highlighted in pink, the central channel appears to be unperturbed, and clearly the channel extends into the host without obstructions. However, if we examine the ‘surface’ cross section in figure 6.13c, we see that the channel system has apparently become blocked by the presence of the central dislocation. The neighbouring channels, although deformed from a circular to oval cross-section, are not obstructed. A planar cross-section of the channel is shown in atomic detail in figure 6.11b and further data is given in table 6.5. The important feature of this elliptical distortion is that the channel has different minimum and maximum dimensions as well as a chirality. The minimum distance is reduced from 6.78Å to 5.95Å, whilst the maximum distance is increased to 7.80Å.
Table 6.5: Pore dimensions, bond distance and geometry of the 8-rings shown in figure 6.11.

<table>
<thead>
<tr>
<th></th>
<th>8-ring in bulk (figure 6.11 a)</th>
<th>8-ring adjacent to dislocation line (figure 6.11 b)</th>
<th>Central helix (figure 6.11 c)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>O – O distances across ring (Å)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – 5</td>
<td>6.78</td>
<td>5.95</td>
<td>8.87 (7.85)</td>
</tr>
<tr>
<td>2 – 6</td>
<td>6.59</td>
<td>7.36</td>
<td>8.75 (7.99)</td>
</tr>
<tr>
<td>3 – 7</td>
<td>6.78</td>
<td>7.85</td>
<td>8.87 (7.85)</td>
</tr>
<tr>
<td>4 – 8</td>
<td>6.59</td>
<td>6.88</td>
<td>8.75 (7.99)</td>
</tr>
<tr>
<td><strong>Si – Si distances across ring (Å)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A – E</td>
<td>8.11</td>
<td>7.18</td>
<td>9.81 (6.64)</td>
</tr>
<tr>
<td>B – F</td>
<td>8.11</td>
<td>9.34</td>
<td>9.81 (6.64)</td>
</tr>
<tr>
<td>C – G</td>
<td>8.11</td>
<td>9.21</td>
<td>9.81 (6.64)</td>
</tr>
<tr>
<td>D – H</td>
<td>8.11</td>
<td>7.44</td>
<td>9.81 (6.64)</td>
</tr>
<tr>
<td><strong>Si – O bond distances (Å)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A – 1 and E – 5</td>
<td>1.60</td>
<td>1.70</td>
<td>1.71</td>
</tr>
<tr>
<td>A – 2 and E – 6</td>
<td>1.61</td>
<td>1.69</td>
<td>1.70</td>
</tr>
<tr>
<td>B – 2 and F – 6</td>
<td>1.61</td>
<td>1.61</td>
<td>1.70</td>
</tr>
<tr>
<td>B – 3 and F – 7</td>
<td>1.60</td>
<td>1.57</td>
<td>1.71</td>
</tr>
<tr>
<td>C – 3 and G – 7</td>
<td>1.60</td>
<td>1.57</td>
<td>1.71</td>
</tr>
<tr>
<td>C – 4 and G – 8</td>
<td>1.61</td>
<td>1.65</td>
<td>1.70</td>
</tr>
<tr>
<td>D – 4 and H – 8</td>
<td>1.61</td>
<td>1.63</td>
<td>1.70</td>
</tr>
<tr>
<td>D – 5 and H – 1</td>
<td>1.60</td>
<td>1.72</td>
<td>1.71</td>
</tr>
<tr>
<td><strong>O – O nearest neighbours (Å)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – 2 and 5 – 6</td>
<td>2.55</td>
<td>3.15</td>
<td>2.90</td>
</tr>
<tr>
<td>2 – 3 and 6 – 7</td>
<td>2.55</td>
<td>2.38</td>
<td>2.90</td>
</tr>
<tr>
<td>3 – 4 and 7 – 8</td>
<td>2.55</td>
<td>2.32</td>
<td>2.90</td>
</tr>
<tr>
<td>4 – 5 and 8 – 1</td>
<td>2.55</td>
<td>2.87</td>
<td>2.90</td>
</tr>
<tr>
<td><strong>Si – Si nearest neighbours (Å)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A – B and E – F</td>
<td>3.08</td>
<td>3.29</td>
<td>3.32</td>
</tr>
<tr>
<td>B – C and F – G</td>
<td>3.13</td>
<td>2.99</td>
<td>3.67</td>
</tr>
<tr>
<td>C – D and G – H</td>
<td>3.08</td>
<td>3.15</td>
<td>3.32</td>
</tr>
<tr>
<td>D – E and H – A</td>
<td>3.13</td>
<td>3.39</td>
<td>3.67</td>
</tr>
<tr>
<td><strong>O – Si – O angles (°)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – A – 2 and 5 – E – 6</td>
<td>104.9</td>
<td>135.1</td>
<td>116.5</td>
</tr>
<tr>
<td>2 – B – 3 and 6 – F – 7</td>
<td>104.9</td>
<td>96.3</td>
<td>116.5</td>
</tr>
<tr>
<td>3 – C – 4 and 7 – G – 8</td>
<td>104.9</td>
<td>92.1</td>
<td>116.5</td>
</tr>
<tr>
<td>4 – D – 5 and 8 – H – 1</td>
<td>104.9</td>
<td>177.4</td>
<td>116.5</td>
</tr>
<tr>
<td><strong>Si – O – Si angles (°)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A – 1 – H and D – 5 – E</td>
<td>154.9</td>
<td>161.4</td>
<td>159.2</td>
</tr>
<tr>
<td>A – 2 – B and E – 6 – F</td>
<td>145.0</td>
<td>167.9</td>
<td>143.2</td>
</tr>
<tr>
<td>B – 3 – C and F – 7 – G</td>
<td>154.9</td>
<td>144.9</td>
<td>159.2</td>
</tr>
<tr>
<td>C – 4 – D and G – 8 – H</td>
<td>145.0</td>
<td>146.5</td>
<td>143.2</td>
</tr>
</tbody>
</table>

* Values in parentheses have been corrected to allow for pitch of dislocation and give the apparent width of the channel.
Figure 6.13: Channel systems in perfect and dislocated zeolite A. Parts a, b and c show the bulk system and d, e and f show the dislocated system. Parts a, b, d and e are the surface of the *inaccessible* volume viewed down (a and d) and across (b and e) the dislocation line. Parts c and f show the surface *accessible* volume, with the central helix highlighted in blue. *First published under licence by McMillan* (Walker et al., 2004).
To spatially image the cage systems that surround the dislocation, we plot the surface that bounds the *accessible* volume, to show where small molecules and ions with their hydration shell can diffuse through. In figure 6.13f the core is clipped to reveal the channel and cage system around the dislocation. Intriguingly, the effect of the dislocation is to almost completely equalise the diameter of the channel to give a helical tube with a prescribed handedness, seen in the growth spirals imaged in figure 6.8. The neighbouring channels are distorted, but are much more closely related to the perfect lattice channel-cage-channel system than the dislocation core, which exhibits a nearly perfectly regular channel diameter.

It is clear that the effect of the dislocation upon the channel system is dramatic, and we can expect that molecular transport will be considerably modified. Computer simulation of molecular transport in zeolites is a well-studied field, where very detailed and comprehensive investigations into jump dynamics at pinch points or bottlenecks within zeolites have been reported (Smit and Siepmann, 1994; Auerbach, 2000). It is beyond the scope and aim of this work to quantitatively analyse how the critical barrier heights for transport are affected; however, qualitative predictions can be made on simple geometric grounds. Since the diameter of the core channel is larger, access from the surface to the core interior will be easier than for neighbouring channels in the non-defective material, thereby permitting larger molecules to gain access. Diffusion parallel to the surface is, however, regulated by the distorted eight rings and thus transport of molecules with a kinetic diameter exceeding that of the eight-membered ring perpendicular to the dislocation core will be severely retarded. In LTA, because of the size of the dislocation vector, the channel systems become misaligned by $\frac{1}{2}a$, as is depicted in figure 6.13e. The guest molecule diffuses to the core region and then has to follow a branch to another channel to continue its traversal across the system. In effect the dislocation scatters diffusing species by rotating their direction of motion. It would seem likely that even at moderate partial pressures, molecules would have a relatively high-residence time in the vicinity of dislocation core, and this could lead to a substantial reduction of the gross ‘flux density’ of gas through the material. The overall reduction of the flux clearly depends on the dislocation density.
Although this is likely to be small, it has already been demonstrated that the structure is perturbed out to a range of 20 – 30 Å, and so the total volume influenced by the dislocation may be far from negligible.

Focusing on the shape of the dislocation core, given that it is helical it will locally be chiral and therefore different enantiomers will diffuse at distinct rates along it. The overall material may of course be racemic if an equal number of dislocations of opposite handedness exists. However, the enantiomers will become spatially partitioned within the material, which may be exploited.

Finally, turning to the question of how the reactive properties of the core may be different to those of the internal surface of the perfect crystal. Because the eight-membered ring is distended along a, it follows that all the Si-O (and Al-O in the real system) bond lengths are extended. Analysis of the correlation between Si-O bond length and the ‘reactivity’ is not simple; Si-O-Si bond lengths and angles can adopt a myriad of values with a remarkably minor energetic penalty (Henson et al., 1994), evident from the ~35 purely siliceous zeolitic structures, with extremely varied pore systems and reactive properties. However, given that the Si-O bond length is extended, it follows that the bond is weakened, and it will therefore be more vulnerable to attack by for example, hydrolysis and therefore dissolution or exchange of framework aluminium. It is also likely, therefore, that the core interior will be more acidic than an ideal channel, which will permit different chemistry to occur.

In summary, this example has revealed the atomic structure of a dislocation core in zeolite A. The considerable distortion of the local channel structure will bestow very different properties on the region surrounding the dislocation, especially with respect to molecular diffusivities. This may be especially important for aligned composite zeolite membranes, such as that recently described by Lai et al. (2003).

In this chapter, a methodology for studying dislocations in complex ionic materials has been described and used to examine screw dislocations in MgO and zeolite A. The structure of the cores of the two dislocations and the nature of the atomistic relaxation has been found to be rather different. MgO exhibits rather large atomic displacements resulting in reconstruction of a “twisted” bulk like structure at the core. In contrast, the atomic relaxations in zeolite A are a
small fraction of the Burgers vector and the core retains a unique helical structure (which is anticipated to have rather interesting properties). It seems likely that the difference in the behaviour of the two materials is due to their very different density and structure. The following chapter uses the methodology described above to examine the structure of the cores of some screw and edge dislocations in forsterite.
Chapter 7: Modelling dislocations in olivine

Although there have been a large number of experiments to determine the dislocation microstructure of olivine there are no results that shed light on the atomic scale structure of the dislocation cores. The deformation mechanisms of olivine were discussed in chapter 2. Briefly there are two proposed strain rate controlling processes, dislocation climb controlled by the diffusion of point defects and the cross slip of screw dislocations. Most workers prefer the former mechanism but there is no clinching argument for or against either of them. The observed slip systems are rather complex and vary with the strain rate and temperature, but both \([100](0kl)\) and \([001](hk0)\) have been widely reported (see Cordier, 2002 for a review). So far, the structure of the cores of dislocations in olivine are unknown, a situation which causes difficulty if we are to correctly assign a deformation mechanism to the Earth’s upper mantle because of the great importance of the dislocation core in the proposed deformation mechanisms.

This chapter reports the first computational studies of the cores of dislocations in any mantle silicate; in particular, models of the structure of the cores of some dislocations in olivine are described. In order to simplify matters only a subset of the observed dislocations have been modelled, namely the \(a[100]\) and \(c[001]\) screw dislocations and the \(a[100](010)\) edge dislocation. The screw dislocations are studied because of their potential importance in limiting deformation via cross slip while the edge dislocation has very recently been the subject of an extremely accurate experimental study (Johnson et al., 2004).

The modelling described in this chapter is insufficient to solve the problems relating to dislocation motion in olivine; it should be considered a proof of concept study to evaluate the possibility of a comprehensive study in the future and identify some of the possible pitfalls. The modelling methodology has been outlined in chapter 6 and this ground will not be covered again here, but the applicability of the parameterised model used in chapters 4 and 5, will be considered before describing the core structure of the individual dislocations.
7.1 Atomistic models

It is crucial that the potential model used to simulate dislocations can describe a large number of atoms, while retaining a good description of the elastic behaviour of the material. If the elastic constants tensor is correctly described then it is possible that the important non-linear effects may be captured by the model, but, if the elastic behaviour is incorrect in the limit of small strain, it is very unlikely that the non-linear effects will be captured. The major problem with the second requirement, that of the ability to model large systems, arises from the need to include a summation of the Coulomb interactions in the model.

In order to utilise the potential model described in table 4.1 the Coulomb summation of Saunders et al. (1994) must be used as the Coulombic interactions are subtly altered by the use of the Wolf summation (1999) and this must be corrected for in the short range part of the potential model. However, the Saunders summation is found to be computationally impracticable for large systems. A new model shown in table 7.1 was therefore derived, making use of the Wolf summation and using potential functions designed to smoothly decay at the short range cut off. It was hoped that this model would overcome some of the deficiencies of the previous model described in section 4.2 but, as can be seen from the comparison presented in table 7.2, the improvement is marginal. Interestingly, altering the charge of the model ions was found to lead to a major improvement – an observation that suggests possible new directions to improve the model.

The inaccurate representation of the elastic constants tensor, reported in chapter 4, persists in the new model. This suggests the possibility of errors in the calculated displacement fields and dislocation energies calculated using the model. However, the resulting structure is likely to be qualitatively correct, so both models have been used to calculate the structures of dislocations, but, because of the likelihood of inaccuracies in the evaluation of the dislocation energy, it has not been calculated. The calculated structure of screw dislocations is the same, regardless of the model used and thus only the new model was used for edge dislocations. In the following sections the calculated structure and modes of relaxation close to the core are presented.
**Table 7.1:** Parameters of the modified potential model used to study some dislocations. Sources and equations are given in the text. The pair potentials are truncated beyond 8 Å using a smooth cosine function extending over 1 Å and the three body interactions only operate between central silicon ions and the oxygen ions within 4 Å when the oxygen ion separation is less than 8 Å. Coulomb interactions are handled by the Wolf (1999) summation with a cut off of 16 Å and a shielding parameter (defining the width of the error function) of 0.2.

<table>
<thead>
<tr>
<th>Charges</th>
<th>Core - shell spring constant eV Å²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ions</td>
<td>Core</td>
</tr>
<tr>
<td>Mg</td>
<td>2.0</td>
</tr>
<tr>
<td>Si</td>
<td>4.0</td>
</tr>
<tr>
<td>O</td>
<td>0.84819</td>
</tr>
<tr>
<td>Buckingham potential</td>
<td>A (eV)</td>
</tr>
<tr>
<td>Si – O</td>
<td>1443.69990</td>
</tr>
<tr>
<td>O – O</td>
<td>22764.0</td>
</tr>
<tr>
<td>Mg – O</td>
<td>1396.95310</td>
</tr>
<tr>
<td>Exponential three-body</td>
<td>k (eV-rad⁻²)</td>
</tr>
<tr>
<td>O – Si – O</td>
<td>0.671854×10⁶</td>
</tr>
</tbody>
</table>

**Table 7.2:** Comparison of the performance of two atomistic models used to model dislocations in forsterite. Sources of experimental data are given in table 4.2.

<table>
<thead>
<tr>
<th></th>
<th>Experiment</th>
<th>Old model (table 4.1)</th>
<th>New model (table 7.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a (Å)</td>
<td>4.7534(6)</td>
<td>4.782</td>
<td>4.760</td>
</tr>
<tr>
<td>b (Å)</td>
<td>10.1902(15)</td>
<td>10.246</td>
<td>10.339</td>
</tr>
<tr>
<td>c (Å)</td>
<td>5.9783(7)</td>
<td>5.986</td>
<td>6.031</td>
</tr>
<tr>
<td>C₁₁ (GPa)</td>
<td>328</td>
<td>358.62</td>
<td>330.09</td>
</tr>
<tr>
<td>C₁₂ (GPa)</td>
<td>200</td>
<td>206.62</td>
<td>205.83</td>
</tr>
<tr>
<td>C₁₃ (GPa)</td>
<td>235</td>
<td>281.14</td>
<td>278.13</td>
</tr>
<tr>
<td>C₂₂ (GPa)</td>
<td>67</td>
<td>44.23</td>
<td>45.92</td>
</tr>
<tr>
<td>C₂₃ (GPa)</td>
<td>81</td>
<td>74.54</td>
<td>69.69</td>
</tr>
<tr>
<td>C₃₃ (GPa)</td>
<td>81</td>
<td>84.29</td>
<td>83.42</td>
</tr>
<tr>
<td>C₄₄ (GPa)</td>
<td>69</td>
<td>93.83</td>
<td>94.67</td>
</tr>
<tr>
<td>C₅₅ (GPa)</td>
<td>69</td>
<td>96.18</td>
<td>99.44</td>
</tr>
<tr>
<td>C₆₆ (GPa)</td>
<td>73</td>
<td>87.70</td>
<td>90.20</td>
</tr>
</tbody>
</table>

### 7.2 The a[100] screw dislocation

a[100] screw dislocations in forsterite have been modelled using the general methodology described in chapter 6. The simulation cell used was 75 Å in radius (12025 atoms) with the outer 25 Å held fixed during the atomistic
relaxation. The edge of the simulation cell is terminated with oxygen ions with partial charges assigned so that the cell can be considered as consisting of two types of change neutral units, [MgO₆] and [SiO₄]. The cell is centred on the M₁ sites to avoid any total dipole across it. Several sites were selected to act as the origin of the dislocation within this cell and single point calculations indicated that an origin within the M₁ “channel” yielded the lowest energy. However, as this site also removes the dipole across the cell (as described below) it is possible that this effect is not directly related to the structure of the core.

The nature of the effect of the dislocation on the electrostatic dipole across the simulation cell is rather interesting. If one considers a thin slice through the cell and calculates the components of its dipole perpendicular to [100] a finite dipole is found. This dipole is exactly cancelled by a dipole in the opposite direction in the equivalent slice displaced by half the lattice parameter along a. This arises because the M₁ site is an inversion centre. The interesting effect is when the screw dislocation is introduced – this offsets the two sides of the cell by 1/2a and removes the dipole in any and every slice because the string of inversion centres is replaced by a two-fold rotation axis.

The relaxed structure of the core is pictured in figure 7.1. The structure appears to be little altered by the presence of the dislocation but, in fact, the core takes on a rather unusual structure. In order to understand this, it is necessary to consider the HCP oxygen sub-lattice and the nature of the discontinuity in the displacement field at the core of the screw dislocation. Looking along the incipient dislocation line in the un-dislocated structure, one sees two triangles of oxygen ions “pointing” in opposite directions, where each triangle belongs to a different close packed layer and are offset from one another by 1/2a (figure 7.2a). Alternatively, these triangles form opposite faces of the M₁ octahedra. On the introduction of the dislocation the oxygen atoms are displaced by an amount that depends on the angle around the dislocation line. The effect of this translation is shown in figure 7.2b, the resultant structure is rather unusual in that it has a plane of four oxygen ions forming an approximate square separated by pairs of atoms, the magnesium ions migrate into this 6-coordinate, trigonal-prismatic environment.
Now turning to the core displacement field, figure 7.3 shows the components of this field perpendicular and parallel to the dislocation line. For the displacements along the line there are some large displacements of oxygen ions as they arrange themselves in the square conformation and a very large displacement of the central Mg ion so that it can maintain 6-fold co-ordination. It is worth noting that the magnitude of the displacements, as a fraction of the lattice vector, is much smaller than the displacements associated with the screw dislocation in MgO. The displacements perpendicular to the dislocation line are predominantly inwards; presumably, this is to reduce the strain at the core associated with the introduction of the dislocation.

**Figure 7.1**: view of core of a [100] screw dislocation in forsterite. Magnesium (blue), oxygen (red) and silicon (pink) atoms are shown and only minor deformation can be seen from this view. Numbered oxygen atoms refer to those in figure 7.2.
Figure 7.2: simple model of core of the a[100] screw dislocation in forsterite showing central octahedra, looking along the dislocation line. Part a shows undeformed octahedra with atoms in the upper plane in black and the lower plane shown as white circles, part b shows the arrangement after the introduction of the dislocation (atoms are numbered to allow comparison with figure 7.1).

Figure 7.3 is continued over page.
Figure 7.3: relaxation of the atomistic model after introduction of the $a[100]$ screw dislocation, movement along the dislocation line is shown in part a while movement perpendicular to the dislocation is shown in part b. Note the large displacement of a magnesium ion along $[100]$ at $r=0$ shown in part a.

7.3 The $c[001]$ screw dislocation

The methodology for modelling the $c[001]$ screw dislocation is identical to that described above for the $a[100]$ screw dislocation. Again, there are a limited number of sensible locations for the origin of the elastic displacement field avoiding the breaking of Si – O bonds. The lowest energy starting configuration has the origin located along a string of M1 sites, although the line does not pass directly through the octahedral faces.

Figure 7.4 shows the structure of the core following relaxation; it is particularly difficult to visualize this structure in a two-dimensional image, but the key change is that the string of inversion centres along the M1 sites becomes a two-fold rotational axis on the introduction of the dislocation. Examination of the structure reveals that all the magnesium ions are in octahedral coordination despite the presence of the dislocation core. The core displacement field is shown in figure 7.5 and the oxygen ions around the core move parallel to the dislocation line by up to ~0.7 angstroms or $1/8c$. The atoms also relax inwards, which corresponds to the expected strain relieving behaviour observed for the
a[100] screw dislocation in forsterite as well as screw dislocations in MgO. The $1/8a$ displacement parallel to the dislocation line results in a core structure that is very much like the undeformed crystal structure (see figure 7.6) by reversing the local effect of the screw dislocation.

**Figure 7.4:** view of the core of the c[001] screw dislocation in forsterite. The colours of the atoms correspond to those given in figure 7.1 and the structure is rather similar to the structure of bulk forsterite.
Figure 7.5: core displacement field associated with the $c[001]$ screw dislocation in forsterite. Part a shows displacements parallel to the dislocation line while part b shows displacements perpendicular to the dislocation line. It is clear that a larger cell is needed to converge to displacement field at the edge of the relaxed region.
Figure 7.6: Model for the structure of the core of a c[001] screw dislocation in forsterite. Parts a and b show the structure of the oxygen sub lattice close to the incipient dislocation core in the looking along [100] and [001] respectively. Black atoms are close to the observer while white atoms are further away. Part c shows the relative motion of the atoms on the introduction of the dislocation, assuming atom (1) is taken as fixed, atom (2) moves by approximately 1/4c, atom (3) moves by approximately 1/2c and atom (4) moves by approximately 3/4c. Part d shows the atomistic relaxation around the core with atoms in strings 1 and 3 taken as fixed, this means that rather than displacing all the atoms by 1/8c (see figure 7.5 a) two of the strings are displaced by 1/4c. The atomistic relaxation along the c-axis results in the recreation of octahedral coordination for the magnesium ions.
7.4 The $a[100](010)$ edge dislocation

In addition to the screw dislocations, the edge dislocation with Burgers vector $[100]$, dislocation line vector $[001]$ and glide plane $(010)$ has also been studied. It is particularly relevant to examine this dislocation as it has recently been subject to a study Johnson et al. (2004) using a new approach combining high resolution transition electron microscopy (HRTEM) with geometric phase analysis (Hütch et al., 2003). Details of the method are given by Hütch et al. (1998; 2003) and references therein but, in essence, the procedure involves taking two HRTEM images, one containing a dislocation and one in a dislocation free part of the crystal and by comparing the local Fourier transform of these images extracting the real space displacements needed to match the two images. This displacement corresponds to the total displacement field (linear elastic plus core contribution) introduced by the dislocation, and has resolution better than 0.1 Å. This section presents a model of the edge dislocation studied by Johnson et al. and compares the atomic scale results with the experimental determination of the dislocation displacement field.

Figure 7.7 shows the simulation cell used to model the edge dislocation after relaxation with the Burgers circuit highlighted and the Burgers vector shown in red. The simulation cell is identical to the one used to model the $e[001]$ screw dislocation but, on introduction of the edge dislocation additional material was removed from the perimeter of the cell. The final simulation cell with the dislocation has a radius of approximately 60 Å with the inner 30 Å allowed to move during the relaxation. This is a relatively small cell and improvements could be made given sufficient computational resources. The elastic solution used to set up the dislocation has been described in chapter 6 (equations 6.7 – 6.10). The cell is centred on Mg1 sites and is terminated in the same way as the screw dislocation cells. To avoid difficulties with the origin of the dislocation field lying on an atom, the origin was shifted with respect to the centre of the cell by 0.5 angstroms along the $a$ and $b$ axis. This leaves the origin within the octahedral channel but results in well-behaved values for the elastic displacement field for all atomic centres. Only the second Wolf-sum model described above was used to perform geometry optimisation; the conjugate gradients optimizer was used.
Figure 7.7: Simulation cell used to model the $a[100](010)$ edge dislocation in forsterite. Black arrows mark an incomplete circuit around the dislocation core while the red arrow indicated the direction of the Burgers vector. [100] is orientated horizontally, [010] is vertical and [001] is perpendicular to the plane of the paper.

The detail of the core structure after relaxation is shown in figure 7.8. This structure has an interesting resemblance to the Si$_2$O$_7$ unit in wadsleyite, and it is worth noting that, apart from this structure at the core, the SiO$_4$ tetrahedra are relatively unaltered by the presence of the dislocation. However, the magnesium sub-lattice has suffered much more disturbance but magnesium ions more than $\sim$10 Å from the origin of the displacement fields are in positions recognisable as M1 and M2 sites.
Figure 7.9 shows the components of the displacements during the relaxation in graphical form, as contour plots of the displacement along [100] and [010] and compares this to the experimental data of Johnson et al. (2004). Care must be exercised in making this comparison, as there are important differences between the experimental and computational approaches. In both cases, the displacements correspond to the difference between the displacement field prediction from anisotropic elastic theory and the observed or calculated displacement fields. In the computational case the displacements can only be measured for the individual atoms, in order to construct a plot of the displacement field the individual atomic displacements have been averaged as described in the figure caption. In the experimental case, there is background noise arising from imperfections in the two original images used to calculate the total displacement field (this results in the blotchy nature of the experimental figures away from the dislocation core) and there is a kind of spatial averaging arising from the resolution of the HRTEM images that do not achieve atomic scale imaging.

These caveats aside, there is a similarity between the calculated and experimental core displacement field, especially in the [100] components. Close to the core, there is a movement in the negative (to the left) direction in the upper right and lower left quadrants while there is movement to the right in the lower right and upper left quadrants. In the [010] direction the agreement is less convincing but there is a tendency for negative displacements on the right hand side of the core and positive displacements on the left. The magnitude of the displacements seem to be smaller for the atomistic model (maximum ~ 0.2 Å) than those reported from the experiment (maximum ~ 0.9 Å). This could be due to an inability of the conjugate gradients optimizer to locate a lower energy configuration than the predicted structure, could highlight a problem with the potential model or may be due to the use of fixed boundary conditions in the calculation. The use of fixed boundary conditions means that, in effect, the atomistic modelling presented here is performed under constant volume conditions while the correct comparison should use constant pressure conditions, this effect could be exacerbated by the relatively small simulation cell used. As a final note the experimental sample is San Carlos olivine (Fo90),
it is possible that precipitation of excess iron along the core of the dislocation could alter the displacement field.

**Figure 7.8:** View of core region of [100] edge dislocation. The (010) glide plane is horizontal and the view is along the dislocation line, which is parallel with the [001] direction. [100] is left – right and [010] is up – down. The image is just over 20 angstroms across and just under 20 angstroms high. Note the formation of a Si$_2$O$_7$ group reminiscent of the wadsleyite structure
Figure 7.9: Difference between calculated linear elastic displacement field and total displacement field determined by experiment (a and b) and the modelling (c and d). Parts a and c show the [100] components of the core displacement field while parts b and d show the [010] components. The displacements of each atom in the calculated system are used to construct an displacement field on a 50 × 50 grid by spatial averaging, the contours are then constructed around these grid points using a cubic function as implemented in the gnuplot software.

7.5 Discussion

The work presented in this chapter is the first attempt to model the cores of dislocations in forsterite and it allows some interesting observations. Given the radically different structure of the core of the a[100] and c[001] screw
dislocations, which belong to the two most observed slip systems in olivine, one may surmise that the behaviour of the two dislocations may be quite different. The $a[100]$ dislocation may be expected to dramatically alter the diffusivity of magnesium (although whether it would be enhanced or retarded is unclear) while the effect of the $c[001]$ dislocation is likely to be smaller. The ability of these two dislocations to cross slip is also likely to be different – an expectation that could be probed by undertaking a large scale simulation similar to that used by Marian et al. (2004) could be undertaken, however this would require a very substantial investment of computer resources. It is also possible that the different dislocations may have varied behaviour in the presence of water, an hypothesis that could be tested further by further explicit calculations. In any case, many of the properties of the dislocations will critically depend on the geometry of the cores, and these are in turn dependent on the original geometry chosen for the model. Only a few likely starting geometries have thus far been examined and further work, possibly using more capable optimisation algorithms, is required before any clear conclusions should be made.

Even given the preliminary nature of the results presented in this chapter some comparison between the screw dislocations described here and in chapter 6 is possible. Reconstruction of the $c[001]$ screw dislocation achieves a bulk like structure at the core in a similar way to the reconstruction of the $a<100>$ screw dislocation in MgO. This is in contrast to the nature of the reconstruction in zeolite A and preliminary calculations of the structure of the $a[100]$, $b[010]$ and $c[001]$ screw dislocations in MgSiO$_3$ perovskite$^{26}$ which have recently been detected in deformed crystals (Cordier et al., 2004). In these “framework” materials, the reconstruction is limited to a re-arrangement of easily identifiable framework building units (SiO$_4$ tetrahedra in the zeolite and SiO$_6$ octahedra in the perovskite). In these terms, the $a[100]$ screw dislocation is unusual and unlike the other screw dislocations studied so far.

---

$^{26}$These calculations also utilise the methodology described in chapter 6. It is hoped that future studies of edge and screw dislocations in MgSiO$_3$ perovskite will provide further insights into the rheology of the lower mantle as well as providing a useful platform for further developments of the methodology.
Chapter 8: Some conclusions and future directions

Although individual conclusions have been made throughout this thesis, there is much to be gained by bringing them together in one place. It is also worth considering the possible future directions of computational studies of defects in mantle minerals. This short chapter considers these issues and a brief list of the major results include:

1) The most common intrinsic defect in pure forsterite is likely to be the magnesium Frenkel defect consisting of a magnesium vacancy and a split interstitial.

2) The presence of iron permits a more diverse collection of point defects mediated by redox reactions; for example, there is the possibility of oxygen interstitial defects.

3) In general, calculations of point defects in forsterite using parameterised potentials are consistent with high quality embedded cluster calculations. The magnesium ion is more accurately described by the potentials than SiO$_4$ tetrahedra. This adds weight to studies of point defects in other silicates based only on parameterised potentials.

4) An atomic scale model of magnesium diffusion based on the movement of vacancies is consistent with experimental studies. Oxygen diffusion is possible by the movement of interstitials or vacancies. In pure forsterite, diffusion via the movement of vacancies is predicted to be favoured but iron redox makes it possible for oxygen interstitials to become mobile and dominate at high oxygen activity. Available experimental data is consistent with this process and suggests a crossover from one mechanism to the other within the Fo$_{90}$ stability field.

5) A method for the study of the cores of dislocations in materials with a complex structure has been developed.
6) Cores of dislocations in MgO and zeolite A have been studied; dislocations in zeolites are predicted to have unusual, potentially useful properties.

7) Models of the cores of some important dislocations in forsterite have been presented. If cross slip is an important process then the a[100] and c[001] dislocations are likely to behave rather differently; they are also likely to have different effects on diffusivity.

In this thesis, defects have been studied at the atomic scale. Currently computer modelling is the only available method that can be used to extract this information, a situation that is unlikely to change in the near future. However, many of the results of the calculations can be compared indirectly with experiment. In particular, the studies of point defects and diffusion presented in chapters 4 and 5 are consistent with experimental data for magnesium and oxygen diffusion in forsterite and iron bearing olivine. An important observation is that it is necessary to consider the oxidation and reduction of iron, in order to understand the meaning of the experimentally determined activation energy for diffusion in olivine. Further opportunities for research in this area include the possibility of calculation of absolute diffusivities by utilising Kinetic Monte-Carlo methods, and studies of silicon diffusion (although this will be hindered by the difficulty of examining the Si-O bond breaking process).

A further major contribution of the research presented in this thesis is the creation of a protocol for modelling dislocations in complex materials. The methodology, presented in chapter 6, can be used for a wide range of materials but further development would be useful. In particular, studies of perovskite structured materials would be of interest from both the scientific and developmental view point. Scientifically these materials are of interest as components in a wide range of solid state devices and as the major component of the Earth’s lower mantle. As a material to further develop the methodology the structure is ideal as it contains clearly defined channels which should contain the dislocation core and exists in a number of different crystal systems (including cubic, rhombohedral, orthorhombic and monoclinic). This provides an opportunity to develop the elastic part of the model to less symmetrical systems and examine the effect of this in a systematic way. Development of
flexible boundary conditions would also be useful as discussed in chapter 6; this is an essential prerequisite for studies of moving dislocations.

A possible alternative to the study of individual defects is to attempt to model very large polycrystalline systems using molecular dynamics, apply strain and observe the creation and movement of the various types of defect. In order to conduct such a study there are a number of requirements: the ability to model a large number of ions, the ability to model a long time scale and the ability to easily identify the various defects. The ability to identify the defects rests on the possession of idealised models of single defects as described in this thesis, the time and length scale requirements can increasingly be met by making use of new methods such as the various hyper dynamics methods to access long timescales (Voter, 1997b, 1997a, 1998; Sorensen and Voter, 2000) and special parallelisation algorithms to simulate large systems.

In any case, the results of the piece-by-piece approach can be used to analyse mantle dynamics in three distinct ways: Firstly, individual results enable a better understanding of experiment. Secondly, if enough individual pieces of data are gathered it should be possible to construct a synthetic deformation mechanism map following the procedures described by Frost and Ashby (1982) (although empirical information will be needed to select particular constitutive models). The third way that the data could be used is in combination with larger scale models, either to assist in the analysis of large scale atomistic models of a deformed polycrystalline sample, or to provide input parameters to more coarse grained models such as those provided by dislocation dynamics models or continuum models (Dawson, 2002).
Appendix 1: outline of new programmes

In the course of the research leading to this thesis a range of computer programs were created. This appendix briefly outlines the purpose, use and implementation of some of these.

A1.1 A program to support DL_Poly calculations

Generation of input files for molecular dynamics simulations using CCP5’s flagship code, DL_Poly (Smith and Forester, 1996), can be extremely difficult. This program, provisionally entitled GIDY, simplifies the process by generating DL_Poly FIELD and CONFIG files from a GULP output file (which is easily generated using a wide range of tools).

The code consists of over 2000 lines of Perl, takes a gulp output file on STDIN and produces a FIELD and CONFIG file in the working directory, information is written to STDOUT:

```bash
# ./GIDY [arguments] < gulp.out > output
```

If FIELD or CONFIG files already exist, they are renamed. Shells are assigned a fractional mass and the following non-exhaustive list of optional arguments can be used:

--help produces some helpful hints

--[no]verbose [integer] increase the verbosity to STDOUT

--[no]moveshells moves the shells from their location in the the gulp output file to the location of the core. This can help in the equilibration of the shell energy

--[no]testbonds forces a check of covalent radii between atoms in the same molecule before introduction of a bond.

--cutoffforshellsearch[real] distance based cut off for the core-shell matching routine, defaults to 0.6 Å and is defined in the same way as the corresponding option in GULP.

223
--[x|y|z]move [real] move all atoms in x y or z direction (distance in Ångström units).
--[no]rigid make all molecules ridged (ignore defined potentials)
--oldparser forces the use of an old parser for reading the GULP output files.
--molcheck forces checks of all molecules to see if they extend over the side of the cell (this is experimental).

### A1.2 Programs to enable simulations of dislocations

In order to allow modelling of dislocations a number of codes have been written, mostly in perl. The steps used to set up a dislocation simulation, from a bulk calculation of the system, are outlined in figure A1.1 and the scripts used for each step are outlined in the following sections.

```
Gulp input for bulk system

Rotate frame of reference and run gulp (full and nosym keywords)
  Make 1D bulk like cell
  Neutralise cell
  Introduce dislocation
  For edge dislocations remove extra atoms
```

**Figure A1.1:** General approach to generating simulation cells for dislocation simulations.

#### A1.2.1 Rotation of the Cartesian frame of reference

The first step is to take an optimised structure and rotate the frame of reference as described in section 6.2.1 and recalculate the elastic constants tensor. This is achieved by taking a gulp restart file and replacing the cell parameters (marked with the cell option) with rotated lattice vectors (marked with the vectors option) and performing a new calculation of the properties. The calculation of
the lattice vectors can be performed using an interactive FORTRAN program
"rotate frame":

```
# ./rotate_frame
```

### A1.2.2 Cell building

Once a gulp output file is calculated (on the non-standard reference frame) it is
used as the basis for the rest of the generation of the dislocation cell. The first
stage is to produce a circular cell, and a Perl script has been written for this
purpose:

```
# ./dislocate_make_cell radius height disp_x disp_y
    rotated_structure.got cell.gin >& outputinfo
```

The mandatory arguments define the radius and height of the simulation cell, its
origin, the name of the gulp output file containing the crystal structure and a file
name for the (stripped down) gulp input file of the cell. Basic information is
written to STDOUT. Note that there is an obscure bug in this script that results
in errors if the bulk structure does not have orthogonal cell vectors; a work
around is to build a super cell in the previous step.

### A1.2.3 Introducing dislocations

Once the simulation cell has been generated, the next step is to introduce the
dislocation by using the equations outlined in section 6.2.2. One of four scripts
are used for this process:

```
# ./dislocate_dislocate burg_vec cell.gin dislocell.gin >& outputdata
```

Introduces a screw dislocation assuming the material is elastically isotropic
(equation 6.1). Obligatory arguments are the Burgers vector length (it will be
aligned along the x-axis of the cell, which corresponds to the z direction in the
rotated structure) a file name for the structure and a file name for output
structure. Basic information is written to STDOUT. The other three scripts
operate in a similar manner, but are used for elastically anisotropic materials
and edge dislocations. The case of a screw dislocation in an anisotropic
material is handled by equation 6.5, used in "dislocate_dislocate_aniso":

```
# ./dislocate_dislocate_aniso rotated_structure.got burg_vec
    cell.gin dislocell.gin >& outputdata
```
where the additional argument is used to provide the elastic constants tensor. Two very similar scripts have been produced for the generating edge dislocations (using equation 6.2 and 6.7 to 6.10).

**A1.2.5 Cell neutralisation**

Three programs which neutralise simulation cells that are periodic in one dimension, following the approach of Braithwaite *et al.* (2002; 2003), which has been modified so that it functions for a system that is periodic in one dimension. The present implementation is dependent on the "connectivity" of the system, thus different codes are provided for siliceous zeolites, forsterite and wadsleyite:

```
# ./nutralise_zeolite radius del_r dislocell.gin
   nut_dislocell.gin >& outputdata

# ./nutralise_forsterite radius del_r dislocell.gin
   nut_dislocell.gin >& outputdata

# ./nutralise_wadslyite radius del_r dislocell.gin
   nut_dislocell.gin >& outputdata
```

In all cases, the parameters define the total radius of the cell, a cut of distance for the terminating procedure, an input file and an output file.
Appendix 2: QM/MM energies

Table A2.1: Total energies (a.u.) of the various embedded clusters using the smaller simulation cell.

<table>
<thead>
<tr>
<th></th>
<th>No defect</th>
<th>Mg vacancy</th>
<th>Mg split interstitial</th>
<th>Mg vacancy and 2 H</th>
<th>Mg vacancy and 1 H</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF/6-31G//HF/6-31G</td>
<td>-3087.4079</td>
<td>-2888.3557</td>
<td>-3286.0362^)</td>
<td>-2888.7405</td>
<td>-2888.5749</td>
</tr>
<tr>
<td>HF/6-31+G*/HF/6-31G</td>
<td>-3087.9985</td>
<td>-2888.97867</td>
<td>-3286.6028</td>
<td>-2889.3475</td>
<td>-2889.1899</td>
</tr>
<tr>
<td>HF/6-31+G*/HF/6-31+G*</td>
<td>-3088.0296</td>
<td>-2889.0085</td>
<td>-3286.6353</td>
<td>-2889.3794</td>
<td>-2889.2203</td>
</tr>
<tr>
<td>HF/6-31+G*/HF/6-31+G*</td>
<td>-3088.4386</td>
<td>-2889.4050</td>
<td>-3287.0588</td>
<td>-2889.7806</td>
<td>-2889.6201</td>
</tr>
<tr>
<td>PW91/6-31G//PW91/6-31G</td>
<td>-3095.5807</td>
<td>-2896.1043</td>
<td>-3294.6018</td>
<td>-2896.5258</td>
<td>-2896.3407</td>
</tr>
<tr>
<td>PW91/6-31G//PW91/6-31G</td>
<td>-3096.1272</td>
<td>-2896.7015</td>
<td>-3295.1303</td>
<td>-2897.0828</td>
<td>-2896.9170</td>
</tr>
<tr>
<td>PW91/6-31+G*/PW91/6-31+G*</td>
<td>-3096.1525</td>
<td>-2896.72534</td>
<td>-3295.1574</td>
<td>-2897.1083</td>
<td>-2896.9421</td>
</tr>
<tr>
<td>PW91/6-31+G*/PW91/6-31+G*</td>
<td>-3096.6843</td>
<td>-2897.2422</td>
<td>-3295.7051</td>
<td>-2897.6260</td>
<td>-2897.4331</td>
</tr>
<tr>
<td>PW91/6-31+G*/PW91/6-31+G*</td>
<td>-3096.6851</td>
<td>-2897.2432</td>
<td>-3295.7060</td>
<td>-2897.6271</td>
<td>-2897.4607</td>
</tr>
<tr>
<td>PW91/6-31G//PW91/6-31G</td>
<td>-8153.0073</td>
<td>-8153.0585</td>
<td>-8078.9173</td>
<td>-8078.9747</td>
<td>-8078.9422</td>
</tr>
<tr>
<td>PW91/6-31+G*/PW91/6-31G</td>
<td>-8154.2888</td>
<td>-8154.3382</td>
<td>-8080.5081</td>
<td>-8080.5351</td>
<td>-8080.5051</td>
</tr>
<tr>
<td>PW91/6-31+G*/PW91/6-31G</td>
<td>-8154.3382</td>
<td>-8154.3882</td>
<td>-8080.5351</td>
<td>-8080.5651</td>
<td>-8080.5351</td>
</tr>
<tr>
<td>PW91/6-31+G*/PW91/6-31G</td>
<td>-8154.3882</td>
<td>-8154.4382</td>
<td>-8080.5651</td>
<td>-8080.5951</td>
<td>-8080.5651</td>
</tr>
<tr>
<td>B3LYP/6-31G//B3LYP/6-31G</td>
<td>-8156.1969</td>
<td>-8156.2469</td>
<td>-8082.1969</td>
<td>-8082.2649</td>
<td>-8082.2329</td>
</tr>
<tr>
<td>B3LYP/6-31+G*/B3LYP/6-31G</td>
<td>-8156.2469</td>
<td>-8156.2969</td>
<td>-8082.2329</td>
<td>-8082.3009</td>
<td>-8082.2649</td>
</tr>
<tr>
<td>B3LYP/6-31+G*/B3LYP/6-31G</td>
<td>-8156.2969</td>
<td>-8156.3469</td>
<td>-8082.3009</td>
<td>-8082.3689</td>
<td>-8082.3329</td>
</tr>
<tr>
<td>B3LYP/6-31+G*/B3LYP/6-31G</td>
<td>-8156.3469</td>
<td>-8156.3969</td>
<td>-8082.3689</td>
<td>-8082.4369</td>
<td>-8082.3999</td>
</tr>
<tr>
<td>B3LYP/6-31+G*/B3LYP/6-31G</td>
<td>-8156.3969</td>
<td>-8156.4469</td>
<td>-8082.4369</td>
<td>-8082.5049</td>
<td>-8082.4689</td>
</tr>
</tbody>
</table>

(a) An identical relaxed cell with the Mg ion in octahedral co-ordination on the II site gives an energy of -3285.8743 a.u. ~ 4.4 eV higher in energy than the split interstitial.

Table A2.2: Total energies (a.u.) of the various embedded clusters using the larger simulation cell.

<table>
<thead>
<tr>
<th></th>
<th>No defect</th>
<th>Interstitial H</th>
<th>Si vacancy</th>
<th>O2 vacancy</th>
<th>O3 vacancy</th>
<th>Hydrogarnet defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF/6-31G//HF/6-31G</td>
<td>-8153.0073</td>
<td>-8153.0585</td>
<td>-7865.1782</td>
<td>-8078.9173</td>
<td>-8078.9747</td>
<td>-7866.2602</td>
</tr>
<tr>
<td>HF/6-31+G*/HF/6-31G</td>
<td>-8154.2888</td>
<td>-8154.3382</td>
<td>-7866.3379</td>
<td>-8080.5081</td>
<td>-8080.5351</td>
<td>-7867.4558</td>
</tr>
<tr>
<td>PW91/6-31G//PW91/6-31G</td>
<td>-8174.5280</td>
<td>-8174.5822</td>
<td>-7886.2586</td>
<td>-8100.1805</td>
<td>-8100.2354</td>
<td>-7887.3377</td>
</tr>
<tr>
<td>PW91/6-31+G*/PW91/6-31G</td>
<td>-8175.6998</td>
<td>-8175.7495</td>
<td>-7887.3680</td>
<td>-8101.4612</td>
<td>-8101.7016</td>
<td>-7888.4422</td>
</tr>
<tr>
<td>B3LYP/6-31G//B3LYP/6-31G</td>
<td>-8176.0274</td>
<td>-8176.0810</td>
<td>-7887.7312</td>
<td>-8101.6412</td>
<td>-8101.8192</td>
<td>-7889.8179</td>
</tr>
<tr>
<td>B3LYP/6-31+G*/B3LYP/6-31G</td>
<td>-8177.2270</td>
<td>-8177.2769</td>
<td>-7888.8486</td>
<td>-8103.1633</td>
<td>-8103.1892</td>
<td>-7889.9457</td>
</tr>
</tbody>
</table>
Table A2.3: Total energies (a.u.) of the isolated ions used in the calculation of defect energies for the QM/MM calculations in chapter 4.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mg$^{2+}$</th>
<th>Si$^{4+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF/6-31G</td>
<td>-198.8117</td>
<td>-285.1556</td>
</tr>
<tr>
<td>HF/6-31+G*</td>
<td>-198.8121</td>
<td>-285.1564</td>
</tr>
<tr>
<td>HF/6-311+G*</td>
<td>-198.8229</td>
<td>-</td>
</tr>
<tr>
<td>PW91/6-31G</td>
<td>-199.1883</td>
<td>-285.5195</td>
</tr>
<tr>
<td>PW91/6-31+G*</td>
<td>-199.1893</td>
<td>-285.5210</td>
</tr>
<tr>
<td>PW91/6-311+G*</td>
<td>-199.2036</td>
<td>-</td>
</tr>
<tr>
<td>B3LYP/6-31G</td>
<td>-3096.1291</td>
<td>-285.5552</td>
</tr>
<tr>
<td>B3LYP/6-31+G*</td>
<td>-3096.6870</td>
<td>-285.5566</td>
</tr>
<tr>
<td>B3LYP/6-311+G*</td>
<td>-3097.2322</td>
<td>-</td>
</tr>
</tbody>
</table>
References


Blanchard, M., Wright, K., and Gale, J.D. (In prep).


241


Tsai, T.-L., and Dieckmann, R. (2002) Variation of the oxygen content and point defects in olivines, \((\text{Fe},\text{Mg}_{1-x})_2\text{SiO}_4\). \(0.2 < x < 1.0\). *Physics and Chemistry of Minerals*, 29, 680 - 694.


