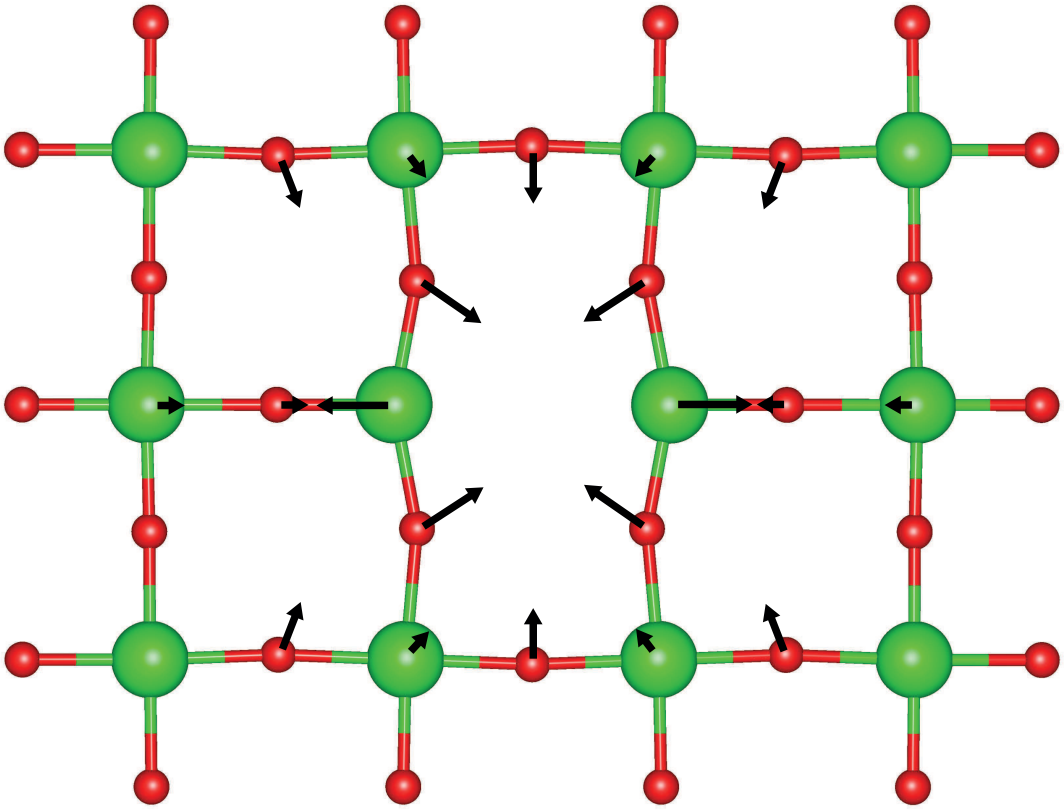
CHALMERS



Chemical Expansion in Barium Zirconate

Master's thesis in Computational Physics

ERIK JEDVIK

Department of Applied Physics
Division of Materials and Surface Theory
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2014
Master's thesis 2014:01

MASTER'S THESIS IN COMPUTATIONAL PHYSICS

Chemical Expansion in Barium Zirconate

ERIK JEDVIK

Department of Applied Physics

Division of Materials and Surface Theory

CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2014

Chemical Expansion in Barium Zirconate
ERIK JEDVIK

Master's thesis 2014:01
ISSN 1652-8557
Department of Applied Physics
Division of Materials and Surface Theory
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: +46 (0)31-772 1000

Cover:
Visualisation of the local lattice distortions upon introduction of a vacancy in an ideal lattice of $BaZrO_3$

Chalmers Reproservice
Göteborg, Sweden 2014

Chemical Expansion in Barium Zirconate
Master's thesis in Computational Physics
ERIK JEDVIK
Department of Applied Physics
Division of Materials and Surface Theory
Chalmers University of Technology

Abstract

The formation volume of several types of point defects in $BaZrO_3$ has been investigated using DFT and represented in terms of a defect strain tensor. The ionic radius for the vacancy and the hydroxide ion are both found to be smaller than the oxygen ionic radius. The radius for the substitutional yttrium ion is found to agree very well with experimental values, as is the formation volume of 0.28 $Å^3$ per unit cell upon hydration.

The strain tensor was found to be non-isotropic but diagonal for both the vacancy and the hydroxide ion while a combination of point defects caused shear strain. The strain tensor in a finite concentration is thus not a linear combination of the strain tensors for the individual defects in the dilute limit.

The cubic perovskite structure of $BaZrO_3$ was also found to be unstable in the LDA and the I4/mcm symmetry was found to exhibit a lower energy although the phonon spectrum still included some imaginary modes.

Keywords: Chemical Expansion, Barium Zirconate, BaZrO3, Perovskite, DFT

Acknowledgements

First and foremost, I would like to thank my supervisor, Göran Wahnström, for guiding me through this project, Anders Lindman for helping me sort out the implementation details and providing me with certain scripts and Adam Arvisson for fruitful discussions on matters of physics.

The research was conducted using computational resorces at Chalmers Centre for Computational Science and Engineering (C^3SE) and National Supercomputer Centre at Linköping University (NSC) through a grant from Swedish National Infrastructure for Computing (SNIC) which is gratefully acknowledged.

Contents

1 Introduction

One of the greatest challenges of the modern society, both from environmental and political considerations, is to reduce our dependence on fossil fuel which is used as an energy source in stationary power plants and as an energy carrier in cars. Among the benefits of fossil fuel are the high energy density and portability. Among the disadvantages are the low efficiency, the environmental issues and the political security issues that may arise. Hydrogen as an energy carrier is an appealing alternative. Ideally, hydrogen is produced through oxidation of water by means of environmentally sustainable solar power and powers the engine through electricity generating fuel cells with water as the only exhaust product. Other types of fuel, such as methane, methanol or even Diesel are also possible to use as energy carriers in fuel cell powered engines although some pre-processing is needed. Although these fuel types cause emission of CO_2 the increased efficiency of a fuel cell makes hydrocarbon energy carriers acceptable during a transition period.

1.1 Fuel Cells

In a combustion engine chemically stored energy is released into heat which causes the gas to expand. This mechanical energy is in turn used to power the engine, typically through the Otto or Diesel cycle. Like the Carnot cycle this process is limited by the second law of thermodynamics and for a petrol engine the Otto cycle limits the efficiency of a typical combustion engine to only about 60% even in the ideal case. In practice it is often well below.

A fuel cell on the other hand converts chemical energy directly into electricity through oxidation of the fuel. This direct conversion, which does not use heat as an intermediate step, has a significantly higher efficiency and efficiencies can reach up to 70% with an additional 20% increase as heat recovery[1].

There are several different types of fuel cells all with their own advantages and disadvantages, operating at different temperature and using different fuel. The following discussion is limited to hydrogen as a fuel, but the discussion is nevertheless quite general since it describes the general mechanism behind fuel cell operation.

Fuel cell operation

A fuel cell consists of three main parts; a positive cathode, an electrolyte and a negative anode (see figure 1.1). The overall chemical process is

$$\mathrm{H_2} + \tfrac{1}{2}\mathrm{O_2} \rightarrow \mathrm{H_2O} \tag{1.1}$$

In a fuel cell this process is separated into two subreactions. Hydrogen $\mathrm{H_2}$ is supplied to the anode at which it is split and oxidised into protons $\mathrm{H^+}$ and electrons as

$$\mathrm{H_2} \rightarrow 2\mathrm{H^+} + 2\mathrm{e^-} \tag{1.2}$$

The electron then flows through the electrical load to the cathode at which oxygen from the air is reduced according to

$$\tfrac{1}{2}\mathrm{O_2} + 2\mathrm{e^-} \rightarrow \mathrm{O^{2-}} \tag{1.3}$$

Finally water is formed according to the reaction

$$\mathrm{O^{2-}} + 2\mathrm{H^+} \rightarrow \mathrm{H_2O} \tag{1.4}$$

If this final process takes place at the anode or cathode depends on whether the electrolyte membrane is proton conducting or oxygen ion conducting. Figure 1.1 shows a proton conducting membrane fuel cell. An advantage of the proton conducting membrane is that fuel (hydrogen) and exhaust (water) are kept apart.

Among the limiting factors to the performance of a fuel cell are, first and foremost, that the electrolyte must be a good ionic conductor (protons or oxygen ions), a good electronic insulator to prevent electrons from

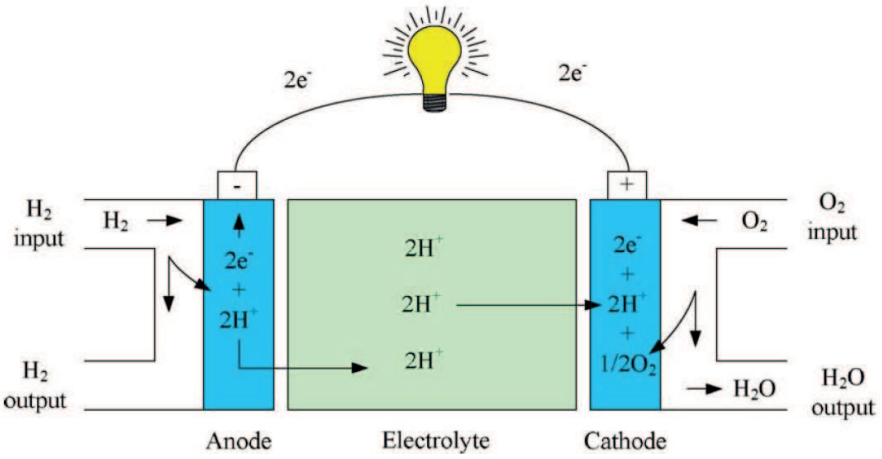


Figure 1.1: *Schematic of a proton conducting membrane fuel cell in operation (from [2] with permission)*

short-circuiting through the membrane, as well as impermeable to gas molecules (H_2 and O_2). Secondly the electrodes need to be a good catalysts for the splitting reaction to occur at a high enough rate and be both good electron conductors in order to transport the electrons to the circuit and good ionic conductors such that ions can pass the electrodes to the electrolyte.

Fuel cells are divided into two categories, low-temperature and high-temperature fuel cells, depending on their respective operating temperature. In common for most low-temperature fuel cells is that platinum is used as the catalyst for the splitting of the gas molecules. Platinum is both expensive and sensitive to impurities of carbon oxides which impedes the catalyst. Many low-temperature fuel cells are also sensitive to electrolyte leakage or water management in the electrolyte. However, their low operating temperature ($80 - 200°C$) make start-up times short.

For high-temperature fuel cells, such as oxygen ion conducting Solid Oxide Fuel Cells (SOFC), which operate at temperatures in the range $800 - 1000°C$ the start-up time is considerably longer. Among other challenges are thermal expansion of the membrane when heated, the strain induced between material of different thermal expansion coefficients and proper thermal insulation of the SOFC from the surroundings. The great benefit of SOFC:s is that they are chemically stable and that the high temperature makes platinum redundant which in turn makes the fuel cell resistant to carbon oxide impedance. Higher efficiencies are also possible.

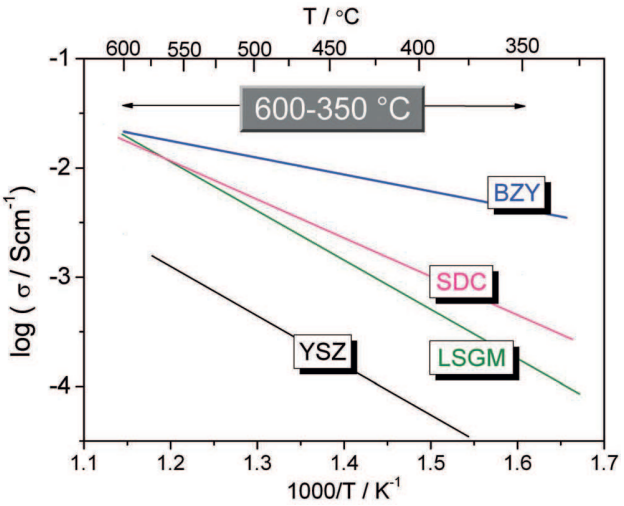


Figure 1.2: *Comparison of conductivity between proton conducting Y-doped barium zirconate (BZY) and some of the best performing oxygen-ion conducting electrolytes, Sm-doped ceria (SDC), doped lanthanum gallate (LSGM) and yttria-stabilized zirconia (YSZ), measured below 600 °C (from [3]) with permission*

The wide gap in operating temperatures ($200 - 800°C$) between low-temperature and high-temperature fuel cells has opened up the possibility of finding a fuel cell which combines the advantages of both low and high temperature fuel cells, which is both resistant to carbon oxide impedance and electrolyte leakage and has a shorter start-up time. Such materials have been found in proton conducting perovskite oxides, such as strontium and barium cerates and later calcium, strontium and barium zirconate [4, 5, 6, 7, 8].

A material that has stood out and combines good proton conductivity at lower temperatures (below $700°C$) and resistance to carbon oxides is yttrium doped barium zirconate [9]. Barium Zirconate, $BaZrO_3$, is an electrically insulating metal oxide with a fairly large band gap of 5.3 eV [10] and high ionic conductivity [9, 3]. Figure 1.2

shows the superior conductivity for Y-doped barium zirconate at intermediate temperatures in comparison to some of the best SOFC electrolytes. Consequently, $BaZrO_3$ has been proposed as a competitive electrolyte at intermediate temperatures in fuel cells.

Although the proton conducting SOFC are promising there are still some unresolved questions. The high proton conductivity is intimately associated with the hydration of the electrolyte. This hydration, which is described in some detail in section 3, causes a volume expansion of the lattice. This volume change associated with increasing vacancy concentration has been defined as *chemical expansion* by Adler [11]. Chemical expansion causes a mechanical strain in the material which in turn can cause deformation or micro-cracking of the electrolyte membrane especially when exposed to a gradient in partial pressure of oxygen [12, 13].

1.2 Aim of the thesis

The aim of this thesis is to investigate the chemical expansion induced in $BaZrO_3$ due to the presence of point defects through theoretical and computational means. Specifically, the size and shape of the vacancy as well as the volume increase due to hydration of the material is investigated and expressed in terms of the defect strain tensor associated with the defect. The local strain behaviour will be investigated and an attempt to explain the origin of the strain will be made. The strain tensor will also be related to experimentally obtained values. The aim is also to investigate the existence of a dilute limit value of the defect strain tensor and study if these values can be added linearly to account for the effect of several defects. Finally, a realistically doped material will be investigated and the result will be compared with available measurements.

2 Perovskite Structure

The crystal structure of $BaZrO_3$ is the perovskite structure, a structure that is rather common among SOFC electrolytes. The perovskite structure is type ABO_3 compound with two interlaced simple cubic lattices of A and B atoms with oxygen atoms on the faces of the A-cube (se figure 2.1).

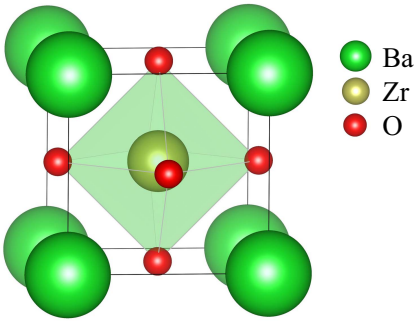


Figure 2.1: *A unit cell of the ideal perovskite crystal structure.*

The dominating bond type in Perovskite oxides is ionic. The crystal can effectively be viewed as built up of charged spheres packed close together with a well-defined radius. The lattice parameter can be defined along the B-O-B axis as $a = 2(r_B + r_O)$ or along the face diagonal as $\sqrt{2}a = 2(r_A + r_O)$. This means that the lattice parameter a can be defined in two, non-necessarily compatible ways. The ratio between the two is the *Goldschmidt tolerance factor* [14]

$$t = \frac{r_A + r_O}{\sqrt{2}(r_B + r_O)} \tag{2.1}$$

If the factor is close to one the lattice parameter computed through the A-atom and the B-atom radii are similar and the structure is cubic with high probability. However, if the ratio is not close one then the structure becomes unstable and reorientation of the atoms can make the structure attain a lower energy. This reorientation is, in ABO_3 perovskites often expressed in terms of the *Glazer notation* [15].

2.1 Tilting of oxygen octahedra

Figure 2.1 shows the highest symmetry phase of the ideal perovskite, henceforth denoted the *aristotype*. The symmetry of this structure is $Pm\bar{3}m$ and it belongs to spacegroup 221. The three oxygen positions are equivalent. This means that a defect formed at one of the three oxygen sites can be moved to any of the other positions by symmetry operations, such as rotation or relabeling of the axes. When the Goldschmidt tolerance factor is not close to one, the oxygen octahedra may be tilted (or rotated), and several lower symmetry phases are energetically more favourable. These structures are denoted *hettotypes*.

The notation for these tilting of octahedra can be described in the *Glazer notation* [15]. By denoting the angles of tilt about the pseudocubic axes [100], [010] and [001] by α, β and γ the new lattice parameters are given by

$$\begin{aligned} a &= a_0 \cos\beta \cos\gamma \\ b &= a_0 \cos\alpha \cos\gamma \\ c &= a_0 \cos\alpha \cos\beta \end{aligned} \tag{2.2}$$

It follows from the definition that e.g. *aac* implies equal tilts about the axes [100] and [010], that *aaa* implies equal tilts around all three axes and that *abc* implies unequal tilts about all three axes.

In addition to the length of the axes the tilts can be *in-phase* about the axis, denoted + or *anti-phase*, denoted −. If there is no tilt about a certain axis this is denoted 0. If, for example, the octahedra are tilted along the [001] axis there are two possibilities, $a^0a^0c^+$ or $a^0a^0c^-$, where the first denotes two octahedra separated along the [001]-axis are rotated in the same direction and the second where they are rotated in different directions. Obviously, two octahedra separated by some other axis, e.g. the [100]-axis, must be rotated out of phase similar to how cogwheels are connected.

In total there are 23 different tilts, including the $a^0a^0a^0$ aristotype, belonging to 15 different symmetry groups, with the aristotype $Pm\bar{3}m$ (group 221) having the highest symmetry.

3 Point defects

In an ideal crystalline material the atoms will arrange themselves in the energetically most favourable state, the perfectly ordered lattice. However, in a real material there are many different types of defects present, including line defects such as dislocations and plane defects such as grain boundaries. This work will focus on point defects. There are three different point defects considered in this work; *substitutional defects*, *vacancies* and *interstitials*. A point defect is a non-stoichiomertic perturbation of the ideal lattice and may carry a charge with it. Defects can be caused by the entropic contribution to the free energy at non-zero temperatures but material can also be prepaired in such a way as to increase the number of defects. This is the purpose with doping of a material.

A zirconium ion in $BaZrO_3$ has the charge +4. By *substituting* a zirconium ion with an yttrium ion of charge +3 a defect with charge −1 has been formed. In the Kröger-Vink (KV) notation, which is often used for defects, this can be written as Y'_{Zr}, where Y signifies yttrium, the subscript Zr indicates that it occupies a zirconium site and the apostrophe indicates the charge −1. Note that the yttrium ion is charged +3 but the charge in the KV notation is relative to the ideal site occupation.

Similarly, a *vacancy* can be formed by removing an oxygen ion. This would be denoted $V_O^{\bullet\bullet}$ in the KV notation, where V indicates a vacancy, the subscript O indicates that it occupies an oxygen site and the +2 charge is denoted by the two dots.

The last kind of point defect considered in this work is the *interstitial* defect. An interstitial is an ion situated between the sites of the ideal crystal. In principle an interstitial atom can be any atom, including the atoms already in the lattice, but in this work the interstitials will be limited to protons. A proton on an interstitial site can be denoted $H_i^{\bullet}$. In $BaZrO_3$ the protons occupy a size so close to an oxygen ion that the denomination $OH_O^{\bullet}$ is better suited.

For future reference the notation for a non-defect site is introduced as $O_O^{\times}$, which is an oxygen on an oxygen site with neutral charge.

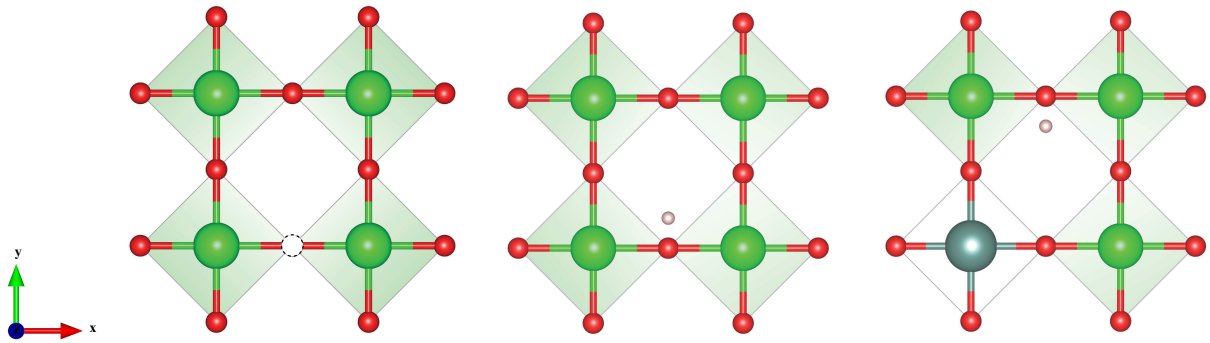
Properties of defects

Although defects will be induced only because of entropy considerations the crystal must be charge neutral on a macroscopic scale. For every positively charged defect created, negatively charged defects must be created to keep the system charge neutral. For example, the formation of the substitutional defect Y'_{Zr} in $BaZrO_3$ may cause the creation of a hole state in the valence band. The yttrium ion thus behaves like a shallow acceptor and consequently the addition of yttrium or other trivalent cations is called *doping* and the ion itself is called *dopant*. Similarly, the formation of an oxygen vacancy of charge +2, $V_O^{\bullet\bullet}$, may cause the creation of two electron states in the conduction band and the vacancy behaves like a shallow donor. These two types of defects can be combined in the formation of two substitutional defects and one vacancy. This is assumed to be the case for $BaZrO_3$ at elevated temperatures and low partial pressure of oxygen [16]. The purpose of doping is thus to increase the number of vacancies in the material.

The number of vacancies in the material is essential for the conduction of protons since there is a balance between oxygen ions and vacancies in the material and water in the gas phase according to

$$H_2O_{(g)} + V_O^{\bullet\bullet} + O_O^{\times} = 2(OH_O^{\bullet}) \tag{3.1}$$

The proton interstitial formed on the site of the vacancy can then move from one oxygen site to another with some activation energy. Far from a dopant these activation energies are rather independent of the species of the dopant, but close to the dopant the activation barrier can be quite significant [17]. Furthermore, the dopant can both repel and attract the interstitial, in the latter case forming a trap site of a lower energy. At higher doping concentrations, in practise these can be as high as 20%, the likelihood of two dopants occupying nearest neighbour sites is significant and can then form a double trap. While dopants are necessary to form sites for the interstitial protons, the dopants at the same time form trap sites which can impede the proton conduction.



(a) (b) (c)

Figure 3.1: *Illustration of the defect location notation (a) an oxygen vacancy* $V_O^{\bullet\bullet}|_x$ *(b) one proton interstitial* $OH_O^{\bullet}|_x{}^y$ *and (c) a combination of an yttrium dopant and a proton interstitial as next nearest neighbour* Y'_{Zr} *and* $OH_O^{\bullet}|_x{}^{-y}$

Extension of the Kröger-Vink notation

When it comes to strain tensor calculations the position of the vacancy is important. Specifically, it is important to be able to distinguish between the three coordinate axes. The three strain principal values will in general not be equal and in order to understand the origin of the strain the orientation of the defect dipole must be specified. An extension of the Kröger-Vink (KV) notation is then necessary.

In an ideal crystal the three coordinate axes are equivalent and the (pseudo-)cubic lattice parameter is the same in all directions, i.e. they can be shifted into each other by symmetry operations such as rotations. Also for a substitutional defect on a B site the orientation of the defect is of no importance, since a substitutional defect is isotropic and doesn't break the four fold rotational symmetry on any off the three axes. The lattice parameters remain equal although not the same as in the ideal crystal. The KV notation is then sufficient for an yttrium defect.

The vacancy on the other hand, has a definite orientation. A zirconium atom has three nearest neighbour oxygen ions, which are all equivalent in an ideal lattice. If a vacancy is formed by removing the nearest neighbour oxygen ion in the (positive) x-direction (see figure 3.1a) this is denoted $V_O^{\bullet\bullet}|_x$. The x-axis remains an axis of four fold rotational symmetry and the effect of the vacancy must be such that the axes perpendicular to the axis of symmetry are equivalent.

A proton interstitial breaks the symmetry even more and requires two specified directions. In addition to specifying on which atom the interstitial has formed, e.g. the $O_O^{\times}|_x$ oxygen, it is necessary to specify in which direction the interstitial is located, e.g. for a hydrogen (proton) interstitial situated on the nearest neighbour oxygen in the x-direction situated slightly in the y-direction (see figure 3.1b) the notation will be $OH_O^{\bullet}|_x{}^y$.

Figure 3.1c also shows the combination of an Y'_{Zr} substitutional defect and a $OH_O^{\bullet}|_x{}^{-y}$ interstitial. The inclusion of the minus sign in the superscript is more of a convenience since the strain tensor is invariant to a 180° rotation of the configuration. The notation is thus not sufficient for non-nearest neighbour defects. However, it is quite sufficient to distinguish different principal directions for strain tensor calculations.

3.1 The concept of an elastic dipole

A point defect produces local distortions of the surrounding lattice in a manner specific for each defect. Often the size of the defect is assessed through the change in lattice parameter of the sample thus averaging the possibly anisotrope strain induced by the defect to a single scalar defect volume. The aim of the following section is to derive a tensor quantity for the strain induced by the defect which can be related to the effective defect radii from experiments. This section follows closely the rendering by Nowic and Berry [18]. The derivation of the strain tensor, ε_{ij}, is to a large extent basic continuum mechanics and can be found in the appendix.

The elastic dipole strain tensor

There is an interaction between the distortions induced by a point defect and a surrounding stress field induced externally. In some ways these interactions are analogous to the interaction between an electrical dipole and an applied electrical field. In analogy, a defect that produces such a distortion has been called an "elastic dipole". A major difference is that, whereas the electrical dipole is a vector quantity, the elastic dipole is characterised by a second rank tensor. By diagonalising this, by necessity symmetric tensor, the principal axes and corresponding eigenvalues can be obtained. By plotting this as an ellipsoid a visualisation of the strain tensor can be obtained.

Since different crystal sites will cause different stress fields the orientation p of the elastic dipole is important. Denoting the number of defects per unit volume N_p in orientation p and the molecular volume of the defect v_0, the molar concentration of defects in orientation p can be written as

$$C_p = v_0 N_p \tag{3.2}$$

The elastic dipole strain tensor associated with the orientation p is denoted $\lambda_{ij}^{(p)}$ and the total elastic dipole tensor is the sum over all n_p non-equivalent orientations p of the defect

$$\varepsilon_{ij}^{rel} = \varepsilon_{ij}^{(d)} - \varepsilon_{ij}^{(0)} = \sum_{p=1}^{n_p} \lambda_{ij}^{(p)} C_p = \sum_{p=1}^{n_p} \lambda_{ij}^{(p)} v_0 N_p \tag{3.3}$$

where $\varepsilon_{ij}^{(d)}$ and $\varepsilon_{ij}^{(0)}$ are the strain tensors with and without defects respectively. The elastic dipole tensor $\lambda_{ij}^{(p)}$ can be expressed as the derivative of the concentration

$$\lambda_{ij}^{(p)} = \frac{d\varepsilon_{ij}^{rel}}{dC_p} = \frac{d\varepsilon_{ij}^{rel}}{dN_p}\frac{dN_p}{dC_p} = \frac{1}{v_0}\frac{d\varepsilon_{ij}^{rel}}{dN_p} \tag{3.4}$$

or

$$\frac{d\varepsilon_{ij}^{rel}}{dN_p} = v_0 \lambda_{ij}^{(p)} = \Lambda_{ij}^{(p)} \tag{3.5}$$

Since the molecular volume is difficult to determine the *defect strain tensor* $\Lambda_{ij}^{(p)}$ is the preferred measure. The derivative requires the concentration N_p, the number of defects per unit volume, to tend to zero. However, this is difficult to obtain when defects come in integer numbers. On the other hand, letting the volume of the supercell tend to infinity achieves the same result.

The defect strain tensor $\Lambda_{ij}^{(p)}$ is a tensor measure of the volume increase and is not a dimensionless quantity but has dimensions of length cube. If the number of defects is kept constant $N_p = 1$ while the dimensions of the supercell increases, the relation between the defect strain tensor and the continuum mechanical strain tensor is

$$\Lambda_{ij}^{(p)} = V_0 \varepsilon_{ij} \tag{3.6}$$

where $V_0 = \det(\mathsf{L_0})$ is the volume of the supercell.

Using the *Lagrangian strain tensor*, defined in equations (A.11) and (A.22) in the appendix the defect strain tensor can be written as

$$\Lambda = \frac{1}{2}\det(\mathsf{L_0})\left(\left(\mathsf{L}\mathsf{L}_0^{-1}\right)^T\left(\mathsf{L}\mathsf{L}_0^{-1}\right) - 1\right) = \frac{1}{2}\det(\mathsf{L_0})\left(\mathsf{L}_0^{-1}\right)^T\left(\mathsf{L}^T\mathsf{L} - \mathsf{L}_0^T\mathsf{L}_0\right)\mathsf{L}_0^{-1} \tag{3.7}$$

This expression is manifest symmetric and hence free of any rotations, which can have no thermodynamic significance since rotations do not alter the internal relation between the atoms. Other expressions, which are not symmetric, have been suggested [19].

A scalar quantity for the total increase in volume can be obtained from

$$\begin{aligned} v^{rel} &= v^{(d)} - v^{(0)} = \det(\mathsf{L}) - \det(\mathsf{L_0}) = \det(\mathsf{L_0} + V_0\varepsilon) - \det(\mathsf{L_0}) = V_0\det(\mathbf{1} + \varepsilon) - V_0 \\ &\approx V_0\left(1 + \varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33} + \mathcal{O}(\varepsilon^2) - 1\right) = V_0\,\mathrm{tr}(\varepsilon) = \mathrm{tr}(\Lambda) \end{aligned} \tag{3.8}$$

and is thus a measure of the defect volume. The advantage of the tensor quantity Λ_{ij} is that, contrary to the scalar volume change $v^{(rel)}$, the tensor quantity also contains information about in which directions the increase and decrease occurs respectively as well as any shear that may take place.

The defect strain tensor in the dilute limit

The defect strain tensor is defined in equation (3.5) as the volume increase contributions in different direction due to the formation of a defect in a crystal of infinite size. In practical computer simulations the crystal is infinite in the sense that periodic boundary conditions are imposed on the supercell, but the supercell itself must be of finite size. For example, a supercell may be formed by joining two unit cells next to each other in all three dimensions forming a 2x2x2 supercell consisting of 8 original unit cells.

Assuming there is only one atom A per unit cell, such as is the case for ABO_3 perovskites, the defect concentrations possible are 1/8, 1/27, 1/64 etc. By increasing the size of the supercell the assumption is that the defect strain tensor will converge to a dilute limit value. The task is complicated by the increasing sensitivity to the defect strain tensor calculation for large supercells, the finite convergence criterion for relaxation in DFT and the computational cost associated with large systems.

Linearity for several defects

In real material, the defect concentration is often far from the dilute limit. For example, the doping concentration in yttrium doped $BaZrO_3$ if often in the range 0.05 to 0.2. At such concentrations the defects cannot necessarily be assumed not to interact with each other and there is no guarantee that the strain tensors computed in the dilute limit can be added linearly as in equation (3.3) to correctly reproduce the strain tensor computed in a supercell containing two defects.

4 Computational Methods

In this work two computational methods have been used; the *ab initio* Density Functional Theory and the model potential for a Rigid Ion Model. Since the fully fledged *ab initio* calculations usually require massive computational resorces and are unfeasible for larger systems, simpler models are often useful. In this section the Rigid Ion Model is described followed by the Density Functional Theory. At the end the computational details nessecary to reproduce the results here obtained are presented. The treatment is rather compact and more detailed treatments can be found in [20, 21].

4.1 Rigid Ion Model

A simple way of modelling the interaction between ions is the Rigid Ion Model (RIM), in which the interaction between ions is a pair potential, i.e. is assumed to depend only on the distance between the ions. Among famous pair potentials is the Lennard-Jones potential and the Buckingham potential which describe the interaction in a repulsive Pauli part and an attractive van der Waals part. For ionic crystals an additional Coulomb term is added. The Buckingham potential used in this work is defined as

$$V(r) = Ae^{-r/\rho} - \frac{B}{r^6} + \frac{Cq_iq_j}{r} \tag{4.1}$$

where A, B, C and ρ are fitted parameters which depend on the particular atoms involved, and q is the charge of the ionic atom.

Sometimes an additional polarisation term is added in the shape of a shell connected to the core by harmonic springs of the form

$$V_P = k|\Delta r|^2 \tag{4.2}$$

where k is the ion-specific spring constant and $|\Delta r|^2$ the core-shell separation. In this latter case the charge of the ion has to be divided between the core and the shell and is hence called Shell Model (SM).

The first two terms in the Buckingham potential are short ranged, i.e. are so small that they can be neglected outside a certain cutoff radius, r_c, in order to speed up calculations. The Coulomb term however, is long ranged and cannot be neglected. Inside the cutoff radius the Coulomb term is computed directly while an Ewald summation in the k-space accounts for the Coulombic interaction outside the cutoff radius. In principle the cutoff radii of the Buckingham potential and the Coulomb potential need not be the same.

Computational details

The Buckinham pair potential calculations in the RIM model were performed using the LAMMPS software [22] with a cutoff radius of 6 Å and an Ewaldsummation with a convergence criterion of 10^{-10} eV/Å to account for the long range Coulomb interaction outside the cutoff radius. The ionic convergence criterion was 10^{-25} eV/Å. The parameters in equation (4.1) were taken from Stokes and Islam [23].

4.2 Density Functional Theory

Density Functional Theory (DFT) is among the methods called *ab initio* methods or *first-principle* methods. This means that the methods are based on solving the most first-principle equation known, the *Schrödinger equation*.

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{\mathcal{H}}\psi \tag{4.3}$$

where $\hat{H}$ is the *Hamiltonian*. In principle, the solution to the Schrödinger equation contains all the wanted properties of a system and is thus essential in quantum mechanics. The solution to the Schrödinger equation requires the solutions to the eigenvalue problem

$$\hat{\mathcal{H}}\psi_n = \varepsilon_n \psi_n \tag{4.4}$$

also known as the *time independent* Schrödinger equation. For a stationary state only the latter equation needs to be solved.

Solving the Schrödinger equation

For one electron in a Coulomb potential the Hamiltonian in the Schrödinger equation can be written as[1]

$$\mathcal{H} = -\frac{1}{2}\nabla^2 - \frac{1}{r} \tag{4.5}$$

With this Hamiltonian the Schrödinger equation can be solved by analytical means, but already for the Helium atom

$$\mathcal{H} = -\frac{1}{2}\nabla_1^2 - \frac{1}{2}\nabla_2^2 + \frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|} - \frac{1}{r_1} - \frac{1}{r_2} \tag{4.6}$$

one must resort to numerical methods. For a multi-particle system of many atoms the Hamiltonian reads

$$\begin{aligned}\mathcal{H} = \sum_{i=1}^{N} \frac{p_i^2}{2} + \sum_{k=1}^{K} \frac{P_k^2}{2M_k} + \frac{1}{2}\sum_{i\neq j} \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|} \\ - \sum_{k=1}^{K}\sum_{i=1}^{N} \frac{Z_k}{|\mathbf{x}_i - \mathbf{X}_k|} + \frac{1}{2}\sum_{k,k'=1;k\neq k'}^{K} \frac{Z_k Z_{k'}}{|\mathbf{X}_k - \mathbf{X}_{k'}|}\end{aligned} \tag{4.7}$$

where the terms represent, the kinetic energy for the electrons and nuclei (with mass M_k) respectively, the interaction between electrons, the interaction between electrons and nuclei and the interaction between the nuclei.

It is customary to apply the adiabatic or Born-Oppenheimer approximation in which the second term is neglected (since $M_k >> 1$ the heavy ions are practically stationary compared to the much lighter electrons) and the last term is treated separately (since as far as electron degrees of freedom is concerned this is only a constant). The Hamiltonian in the Born-Oppenheimer approximation reads

$$\mathcal{H}_{BO} = \sum_i \left[-\frac{1}{2}\nabla_i^2 - \sum_k \frac{Z_k}{\mathbf{x} - \mathbf{X}_k} \right] + \frac{1}{2}\sum_{i\neq j} \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|} \tag{4.8}$$

Even when applying the Born-Oppenheimer approximation the multi particle Schrödinger equation is quite intractable.

The Hartree-Fock approximation

The problem can be simplified by neglecting the correlation between electrons in the full wave function $\Psi(\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N)$. In probability theory the simultaneous probability of $\mathbf{x}_1$ and $\mathbf{x}_2$ can be written $P(\mathbf{x}_1, \mathbf{x}_2) = P(\mathbf{x}_1)P(\mathbf{x}_2)$ if the two probabilities are uncorrelated. The quantum mechanical analogue is $\Psi(\mathbf{x}_1, \mathbf{x}_2) = \psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)$, where ψ_i are one particle wave functions. However, this does not respect the anti-symmetry of the wave function required by permutation symmetry for fermions. When two particles exchange positions the

[1] Using atomic units in which $\hbar = e = m_e = 1/(4\pi\varepsilon_0) = 1$ and dropping the hat on $\mathcal{H}$ for simplicity.

wave function must change its sign. This partial inclusion of correlation, called *exchange*, is due to the Pauli exclusion principle and can, for N fermions, be included in the shape of a *Slater determinant*

$$\Psi(\mathbf{x}_1\mathbf{x}_2...\mathbf{x}_N) = \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix} \tag{4.9}$$

where the $\psi_i(\mathbf{x}_j)$ is the atomic orbital of atom i at position $\mathbf{x}_j$. Using this approximation of the wave function Ψ we can write the Schrödinger equation for the single particle state ψ_i as

$$\mathcal{F}\psi_i = \varepsilon_i\psi_i \tag{4.10}$$

with

$$\begin{aligned} \mathcal{F}\psi_i = \left[-\frac{1}{2}\nabla^2 - \sum_k \frac{Z_k}{\mathbf{x}-\mathbf{X}_k}\right]\psi_i(\mathbf{x}) + \sum_{j=1}^{N}\int \mathrm{d}\mathbf{x}'|\psi_j(\mathbf{x}')|^2\frac{1}{|\mathbf{x}-\mathbf{x}'|}\psi_i(\mathbf{x}) \\ -\sum_{j=1}^{N}\int \mathrm{d}\mathbf{x}'\psi_j^*(\mathbf{x}')\frac{1}{|\mathbf{x}-\mathbf{x}'|}\psi_i(\mathbf{x}')\psi_j(\mathbf{x}) \end{aligned} \tag{4.11}$$

where the first term is the kinetic energy, the second the external potential from by the nuclei, the third called the Hartree potential is caused by the interaction from the charge distribution caused by all the electrons and the fourth term called the exchange term results from the anti-symmetry of the wave function and is non-local.[2] This equation, which is known as the Hartree-Fock approximation, incorporates some of the correlation but not all, and does so in a mean-field way. By definition, what is included is called *exchange* and what is not included is called *dynamical correlation*. Note that the solution appears on both sides of the equation. The solution thus depends on the Hamiltonian which in turn is determined by the solution. The equation is thus non-linear. This implies that equation (4.11) has to be solved iteratively in a self consistent manner.

The Hohenberg-Kohn Theorems

The great advantage of the Hartree-Fock equation (4.11) is that the N-multiparticle Schrödinger equation has been reduced to N non-interacting single state equations. The great disadvantages are that dynamical correlation has been left out, that the last term is non-local and that information about all single particle wavefunctions has to be stored. However, with the electron density defined as $n(\mathbf{x}) = \sum_{i=1}^{N}|\psi_i(\mathbf{x})|^2$ the Hartree-Fock equation can be written in a more illuminating way as

$$\mathcal{F}\psi_i(\mathbf{x}) = \left[-\frac{1}{2}\nabla^2 + V_{ext} + \int\frac{n(\mathbf{x})}{|\mathbf{x}-\mathbf{x}'|}\mathrm{d}\mathbf{x}' + V_x[n](\mathbf{x})\right]\psi_i(\mathbf{x}) \tag{4.12}$$

where V_{ext} is the external potential from te nucleii and $V_x[n](\mathbf{x})$ is the exchange term. The Hamiltonian on the right hand side has been written in terms of the electron density only. This possibility of using the density as a fundamental variable was first noted by Hohenberg and Kohn [24] who proved in two famous theorems that

1. the external potential V_{ext} and thus the full Hamiltonian is uniquely determined by the ground state density $n_0(\mathbf{x})$
2. there exists a functional $E[n]$ for any external potential such that the electron density $n(\mathbf{x})$ that minimises this functional will be the exact ground state density.

In short this means that there exists a functional E[n] of the electron density only which solves the problem exactly, not only including exchange but also dynamical correlation which was previously left out in the Hartree-Fock approximation. Unfortunately, this functional is unknown.

[2]Note that $\mathbf{x} = (\mathbf{r}, \sigma)$ also includes the spin and that effectively the last term only appears if the spin are parallel.

The Kohn-Sham Approach

Despite the theoretical beauty of the Hohenberg-Kohn theorems they give no clue how to make use of these theorems. An ansatz was proposed by Kohn and Sham [25] in which the full interacting many-body system is replaced by an auxiliary system of independent particles. The ansatz relies heavily on the first Hohenberg-Kohn theorem which can be regarded as the inverse relation to the Schrödinger equation. While the Schrödinger equation uniquely determines the wave functions and thus the electron density for a given external potential the first Hohenberg-Kohn theorem proves that the inverse relation also holds.

The great benefit of the auxiliary system is twofold; firstly, it reduces the full problem of N interacting particles to N single particle systems, secondly, it separates out the kinetic energy term and the long-range Hartree term in such a way that the remaining exchange-correlation term can be reasonably approximated by local or nearly local functionals of the density.

It should be stressed that the Kohn-Sham approach is *not* an approximation. In theory the Kohn-Sham approach would be exact if only the exchange-correlation functional were known. The approximation is only made when implementing the exchange-correlation functional.

The Kohn-Sham equations

The Kohn-Sham approach makes use of the variational principle, the expectation value of the Hamiltonian is always higher than the true ground state energy for any wave function ψ

$$E_0 \leq \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{4.13}$$

with equality for the true ground state $|\Psi_0\rangle$. The true ground state is therefore the minimisation of $E[n]$ with respect to the wavefunction Ψ subject to the constraint

$$\int \mathrm{d}\mathbf{r}\, n(\mathbf{r}) = N \tag{4.14}$$

where N is the number of electrons. Assuming $|\Psi\rangle$ to be normalised this leads to the functional equation for the Born-Oppenheimer Hamiltonian (4.8)

$$E_{HK}[n] = \langle \psi | \mathcal{H} | \psi \rangle = T[n(\mathbf{r})] + E_{ee}[n(\mathbf{r})] + \int \mathrm{d}\mathbf{r}\; V_{ext}(\mathbf{r}) n(\mathbf{r}) \tag{4.15}$$

The last term is easily computed, but the first two terms, containing the kinetic energy and the electron-electron interaction of the full interacting system, are unknown.

Kohn-Sham proposed a separation of the two first terms into tree terms; a kinetic energy term of a *non*-interacting single particle $T_s[n(\mathbf{r})]$, a Hartree term $E_H[n(\mathbf{r})] = \frac{1}{2}\int \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}'\, \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$ and an exchange-correlation term E_{xc}. The Hohenberg-Kohn equation (4.15) then becomes in the Kohn-Sham ansatz

$$E_{KS}[n] = T_s[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] + \int \mathrm{d}\mathbf{r}\; V_{ext}(\mathbf{r}) n(\mathbf{r}) \tag{4.16}$$

A variation of this expression with respect to the single particle wave functions $\psi_i(\mathbf{r})$ subject to the constraint in equation (4.14) leads to

$$\left[-\frac{1}{2}\nabla^2 + \int \mathrm{d}\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + V_{xc}(\mathbf{r}) + V_{ext}(\mathbf{r}) \right] \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}) \tag{4.17}$$

an equation strikingly similar to the Hartree-Fock equation (4.11). It must be stressed that, while very similar, the two equations are fundamentally different. While the wavefunctions in the Hartree-Fock equation (4.11) are atomic orbitals from which e.g. excitation energies can be computed (to within some approximation), the wave

functions in the Kohn-Sham equation (4.17) are nothing but auxiliary functions consistent with the density $n(\mathbf{r})$ and the total energy of the system is given by

$$E_{KS} = \sum_i \varepsilon_i - E_H[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] - \int \mathrm{d}\mathbf{r} V_{xc} n(\mathbf{r}) + E_{II} \tag{4.18}$$

where the ion-ion interaction energy E_{II} is computed separately.

The Exchange-Correlation Functional

Up to this point the Density Functional Theory has been exact, save for the Born-Oppeheimer approximation, but since the Hohenberg-Kohn functional

$$F_{HK} = T[n(\mathbf{r})] + E_{ee}[n(\mathbf{r})] = T_s[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] \tag{4.19}$$

is unknown also the exchange-correlation (XC) functional is unknown. The XC-functional may be written exactly as

$$E_{xc}[n(\mathbf{r})] = T[n(\mathbf{r})] - T_s[n(\mathbf{r})] + E_{ee}[n(\mathbf{r})] - E_H[n(\mathbf{r})] \tag{4.20}$$

and the XC-potential V_{xc} is then

$$V_{xc} = \frac{\delta E_{xc}}{\delta n(\mathbf{r})} \tag{4.21}$$

The Kohn-Sham approach of separating the Hohenberg-Kohn functional F_{HK}in kinetic energy, a long range Hartree energy and an exchange-correlation functional E_{xc} has the advantage that the exchange-correlation functional is rather small compared to the kinetic energy and the Hartree energy and that it may be reasonably approximated as a local or nearly local functional of the particle density at the point $\mathbf{r}$ and thus be written as

$$E_{xc}[n(\mathbf{r})] = \int \mathrm{d}\mathbf{r}\, n(\mathbf{r})\, \varepsilon_{xc}([n(\mathbf{r})], \mathbf{r}) \tag{4.22}$$

where $\varepsilon_{xc}([n(\mathbf{r})], \mathbf{r})$ is the energy density per electron at a point $\mathbf{r}$. The accuracy of any calculation any DFT-calculation relies on the approximations of the XC-functional. Over the years an abundance of different XC-functional approximations have been proposed [26], but still today the original functional proposed by Kohn-Sham, the Local Density Approximation, is competitive in certain areas.

Local Density Approximation

The local density approximation (LDA) is perhaps the least sophisticated XC-functional and was suggested already by Kohn and Sham. As Kohn and Sham pointed out, the natural starting point for many solids is the homogeneous electron gas. In this limit the exchange and the correlation are local, i.e. the value at position $\mathbf{r}$ is determined by its value at $\mathbf{r}$ only, and the exchange can even be computed exactly. The correlation has been computed using Monte Carlo methods to great accuracy.

This approximation is remarkably accurate for many solids and was a contributing factor to the great success of the density functional theory. In particular the LDA predicts e.g. bond lengths in solids with close to homogeneous electron density to within a few percent. However, the LDA fails in highly correlated systems and in molecules where the density varies rapidly.

Generalised Gradient Approximation

The general success of the LDA has inspired the development of various Generalised Gradient Approximation (GGA) with the explicit aim of accounting for the inhomogeneous electron density found in real material. The natural extension of the LDA is to include not only the density at a point $\mathbf{r}$ but also gradients of the density. Several various forms have been proposed and although derived in different manners the GGA:s give similar improvements over LDA. In the present text the GGA proposed by Perdew, Burke and Ernzerhof [27] (PBE) has been used.

Implementation in periodic solids

As mentioned above the Kohn-Sham equations (4.17) need to be solved in a self-consistent manner, i.e. an initial electron density is assumed (often the density of a non-interacting electron gas) and the new density is computed. This procedure is repeated until the new and the old density do not differ appreciably any more. In practise the energy of the system is used as convergence criterion; when the energy difference between two consecutive iterations is less than a certain threshold value the calculation is said to have converged.

Is is also necessary to construct a mathematical representation of the one electron wavefunctions ψ_i. According to Fourier's Theorem, and its reformulation for periodic structures made by F. Bloch, any function can be expressed as an infinite sum of plane waves.

$$\psi(\mathbf{r}) = \sum_{\mathbf{q}} u_{\mathbf{q}}(\mathbf{r}) \mathrm{e}^{i\mathbf{q}\cdot\mathbf{r}} \tag{4.23}$$

Bloch also showed that $u_{\mathbf{q}}(\mathbf{r})$ must be periodic with the period of the lattice, i.e. $u_{\mathbf{q}}(\mathbf{r}) = u_{\mathbf{q}}(\mathbf{r}+\mathbf{T})$, where $\mathbf{T}$ is a lattice vector. As a consequence, just as $\mathbf{r}$ is restricted to the unit cell, $\mathbf{q}$ is restricted to the reciprocal unit cell, also named the first Brillouin zone, and can be written as $\mathbf{q} = \mathbf{k} + \mathbf{K}$, for each wave vector $\mathbf{k}$ in the first Brillouin zone and reciprocal lattice vector $\mathbf{K}$.

The Schrödinger equation is now reduced to a linear eigenvalue problem in Fourier or reciprocal space as

$$\left[\frac{1}{2}(\mathbf{k}+\mathbf{K})^2 - \varepsilon\right] C_{\mathbf{k}+\mathbf{K}} + \sum_{\mathbf{K}'} V_{\mathbf{K}+\mathbf{K}'} C_{\mathbf{K}+\mathbf{K}'} = 0 \tag{4.24}$$

and the solutions are

$$\psi_{\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}} \left(\sum_{\mathbf{K}} \mathrm{C}_{\mathbf{k}+\mathbf{K}} \mathrm{e}^{i(\mathbf{k}+\mathbf{K})\cdot\mathbf{r}} \right) \tag{4.25}$$

In principle this sum over $\mathbf{K}$ is infinite but in practice the sum is truncated at an energy cut-off $\frac{1}{2}|\mathbf{k}+\mathbf{K}|^2 < E_{\mathrm{cut}}$. This introduces a small error and E_{cut} has to be chosen judiciously taking both accuracy and computational cost into account. Usually, the energy cut-off is taken as a value beyond which accuracy increases only marginally when increasing the energy cut-off.

Pseudopotentials and PAW

A problem with plane wave basis sets is that rapidly varying functions, such as the wave functions close to heavy nuclei, require a very high energy cut-off in order to be well represented. One solution to this is to use *pseudopotentials* in which the potential from the core, i.e. the nucleus and the innermost electrons, are replaced by a different potential. Among the requirements on psudopotentials are that they should reproduce the correct potential outside the core region such that the correct result is produced, and be smooth enough that a low E_{cut} is possible.

Another possibility is the *Projector Augmented Wave (PAW)* method in which the wave functions in the core region is expanded in atomic orbitals instead of plane waves. Since the core region is rather unperturbed by the neighbouring atoms the atomic orbitals give a much better representation of the wavefunction close to the nuclei and a rather small basis set is possible.

Computational details

The DFT calculations were performed using the VASP software (version 5.3.3) [28, 29, 30, 31] which uses a plane-wave basis set with periodic boundary conditions in all directions. The k-point sampling was performed using a Monkhorst-Pack grid equivalent of 6x6x6 and energy cutoff of 520 eV. The projector augmented wave method (PAW) [32, 33] was used to describe the ion-ion interaction The electronic convergence was within 10^{-5} eV and the forces were less than 0.01 eV/Å for the relaxation of the super cells. The calculations of the phonon diagrams were conducted with an increased precision with electronic convergence of 10^{-8} eV and forces less than 0.25 meV. The phonon diagrams were calculated using the PHONOPY software [34].

5 Results

The principal aim of the present work is to compute the defect volume and the chemical expansion of $BaZrO_3$. However, it is well known that other cubic perovskites, such as $BaTiO_3$ and $BaCeO_3$, do not exhibit a perfect cubic perovskite structure $Pm\bar{3}m$. There have also been reports that $BaZrO_3$ possess the same kind of instability in simulations [35, 36, 37, 38, 39], although experimental data contradicts this [40]. Measurements of ionic radii performed by Shannon [41] also leads to a Goldschmidt tolerance factor of $t = 0.96$ for $BaZrO_3$ which is too close to one to imply an instability with respect to tilting the oxygen octahedra. Before addressing the chemical expansion the computational methods and the stability of the ground state will have to be discussed.

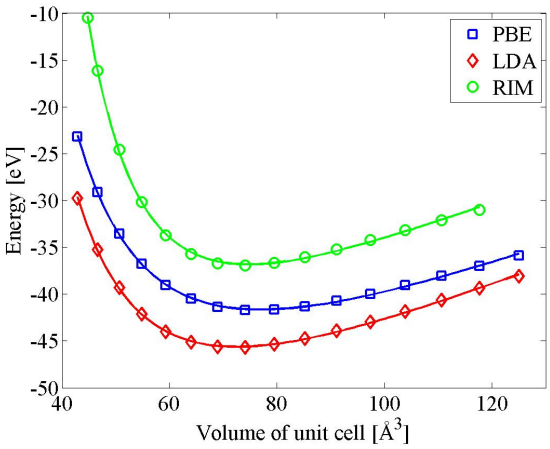


Figure 5.1: *Energy versus volume for $BaZrO_3$ computed with RIM and using DFT with LDA and PBE xc-functionals respectively.*

5.1 The ground state structure

Figure 5.1 shows the energy as a function of volume for the three computational methods used in this work. The energy scales have been shifted to fit in the same figure and only relative values within each computational method are to relevant. The data were fitted to a Murnaghan equation of state [42]

$$E(V) = E_0 + K_0 V_0 \left[\frac{1}{K_0'(K_0' - 1)} \left(\frac{V}{V_0} \right)^{1-K_0'} + \frac{1}{K_0} \frac{V}{V_0} - \frac{1}{K_0' - 1} \right] \tag{5.1}$$

where E_0 is the equilibrium energy, V_0 the equilibrium volume and $K = K_0 + PK_0'$ is the bulk modulus as a function of pressure $P = -dE/dV$, which is zero at equilibrium. Thus the bulk modulus is effectively K_0 at equilibrium.

Table 5.1 lists the computed lattice parameter a and the bulk modulus B together with experimental data as well as data obtained in other computational references. It is immediately clear that the LDA functional underestimates the lattice parameter and that the PBE functional overestimates the same with the same order of magnitude. The pair potential parameters in the rigid ion model have been fitted to crystal data and it is no surprise that the error is small for the RIM. The difference between the values obtained in this work and the cited values are systematic and probably depend on the different implementation of the DFT used. Both cited works use the Wien2k software so the deviations are within the acceptable.

5.2 Stability of the ground state

The stability of a ground state can be investigated through its phonon spectrum. If a configuration exhibits imaginary modes in a phonon spectrum, shown as negative values in the figures, the perturbation associated

		LDA	PBE	RIM	Exp.[43, 44]	LDA[45]	PBE[46]
Lattice parameter	a ($V_0^{1/3}$)	4.159	4.238	4.188	4.19	4.148	4.227
Bulk modulus	B (K_0)	154.3	134.5	190.7	127.2	174.7	155.9

Table 5.1: The calculated lattice parameter and bulk modulus for $BaZrO_3$ in comparison with experimental data and data obtained in other works.

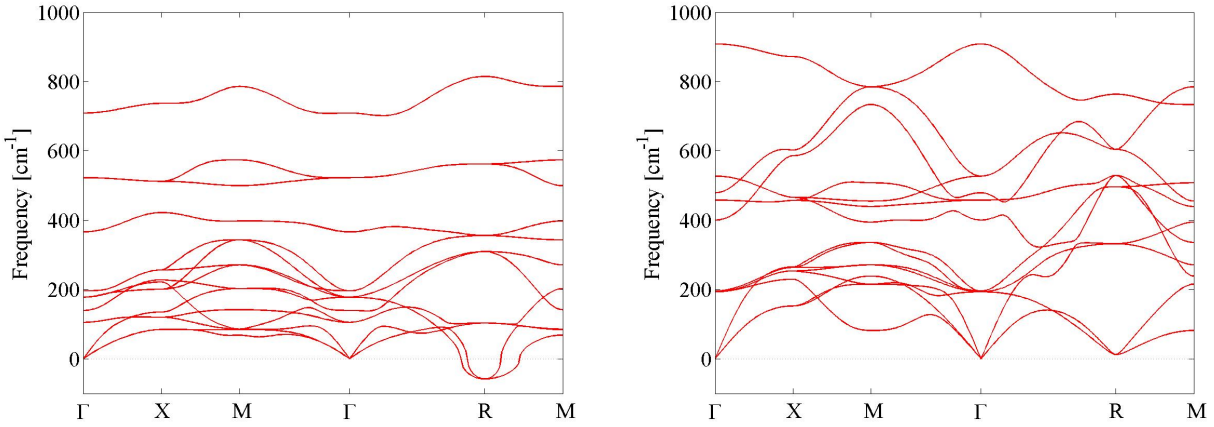


Figure 5.2: *Phonon spectrum for $BaZrO_3$ computed using (a) LDA and (b) PBE xc-functionals respectively in a 2x2x2 super cell of aristotype $Pm\bar{3}m$.*

with this mode is not positive definite and a small perturbation from the saddle point will push the structure into a more stable configuration.

Figure 5.2 shows the phonon spectrum for $BaZrO_3$ in the $Pm\bar{3}m$ symmetry computed from a 2x2x2 supercell of 40 atoms using (a) LDA and (b) PBE. The PBE spectrum shows no imaginary modes and is thus positive definite. The structure has attained a local minimum and is stable. However, for LDA the phonon spectrum does show imaginary modes at the R points consistent with a saddle point.

Three different trial structures were constructed, each of them with a perturbation consistent with the Glazer rotations $a^0a^0c^-$, $a^0b^-b^-$ and $a^-a^-a^-$ which results in the space groups I4/mcm, Imma and $R\bar{3}m$. The structures were then relaxed until the forces were less than 0.25 meV/Å. In the new phonon spectra computed for each trial structure, two of the structures, space groups Imma and $R\bar{3}m$, still exhibited imaginary modes. Also the $a^0a^0c^-$ rotation in figure 5.4a showed imaginary modes in the conventional unit cell setup but did so to a much smaller extent and is thus a candidate for the "true" ground state structure for $BaZrO_3$ if the residual imaginary modes could be resolved. This structure also had the lowest energy of the trial cells, 18.6 meV lower than the aristotype. The energies for the other space groups can be found in table 5.2. It can be noted that all hettotype structures attain configurations with energies lower than the aristotype $Pm\bar{3}m$, however, the difference in energy between the different configurations is very small, in the meV for a unit cell of 40 atoms.

The positions for the structure with the lowest energy after relaxation, written in tetragonal coordinates, is found in table 5.3 and shown graphically in figure 5.4b. This cell is the primitive unit cell containing 20 atoms for the I4/mcm structure.

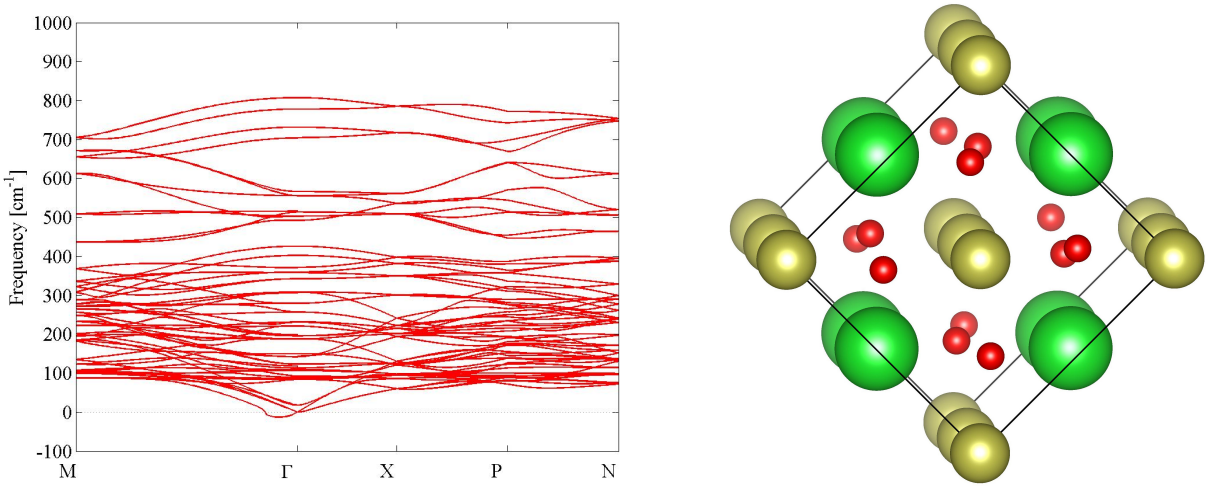
Both the phonon spectrum in figure 5.2a and the positions in table 5.3 agree well with Lebedev *et al.* [37] with the difference that the length of the c-axis and the rotation is somewhat smaller in this work. An attempt to start at the positions given by Lebedev *et al.* resulted, after relaxation, in the same positions given in table 5.3. Also the energies listed in table 5.2 agree well with Lebedev *et al.* but are too large by a factor of 2. This probably depends on different cell sizes used, 40 versus 20 atoms.

Glazer rotation	Space group	Energy meV	Functional
$a^0a^0c^-$	I4/mcm (140)	−18.6	LDA
$a^0b^-b^-$	Imma (74)	−17.0	LDA
$a^-a^-a^-$	$R\bar{3}c$ (167)	−16.5	LDA

Table 5.2: Computed energies for various tilted $BaZrO_3$ structures relative to the 2x2x2 aristotype $Pm\bar{3}m$ supercell of 40 atoms using LDA

Phase	Lattice parameter, Å	Atom	x	y	z
$a^0a^0c^-$	a = 5.8730	Ba	0	0.5000	0.2500
	c = 8.3313	Zr	0	0	0
		$O^{(1)}$	0	0	0.2500
		$O^{(2)}$	0.2315	0.7315	0

Table 5.3: The new equilibrium positions for a relaxed I4/mcm structure computed using LDA corresponding to a Glazer rotation $a^0a^0c^-$ of 4.2°.



(a) (b)

Figure 5.4: *(a) The phonon spectrum for I4/mcm ($a^0a^0c^-$) computed using PBE. The data were obtained using phonopy with non-analytical term correction. (b) A visualization of the crystal structure of symmetry I4/mcm.*

In order to investigate whether there might be a local minimum, albeit at a higher energy, also for the PBE functional the energy landscape around the aristotype was investigated. Attempts to start a relaxation in a hettotype structure failed and converged back to the aristotype. Figure 5.3 shows the energy as a function of tilting for the I4/mcm symmetry relative to the aristotype, where 1 (one) on the abscissa denotes the rotation of 4.2° equivalent to the fully relaxed configuration. The figure shows clearly how a minimum is attained for LDA (red diamonds). For PBE (blue squares) no such minimum can be found and the energy is increasing monotonously in the studied region. Furthermore, the energy scale for PBE is about two orders of magnitude larger than the scale for LDA.

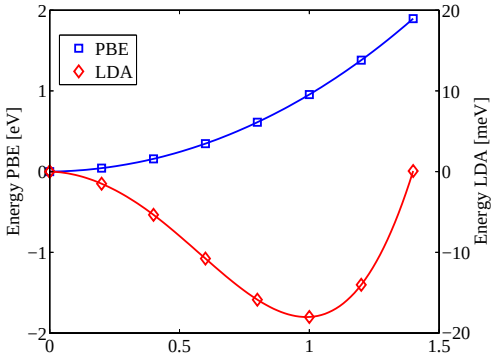


Figure 5.3: *The energy per unit cell (40 atoms) relative to the aristotype as a function of rotation along the $a^0a^0c^-$ direction for PBE (blue squares) and LDA (red diamonds) where 1 (one) on the abscissa denotes the rotation of 4.2° equivalent to the fully relaxed configuration.*

5.3 Chemical expansion

The chemical expansion has been computed in terms of the defect strain tensor for three types of defects; oxygen vacancy, proton interstitial and B-atom substitutional defects. The oxygen vacancy has been used for benchmarking the three different computational methods used and is treated in some more detail.

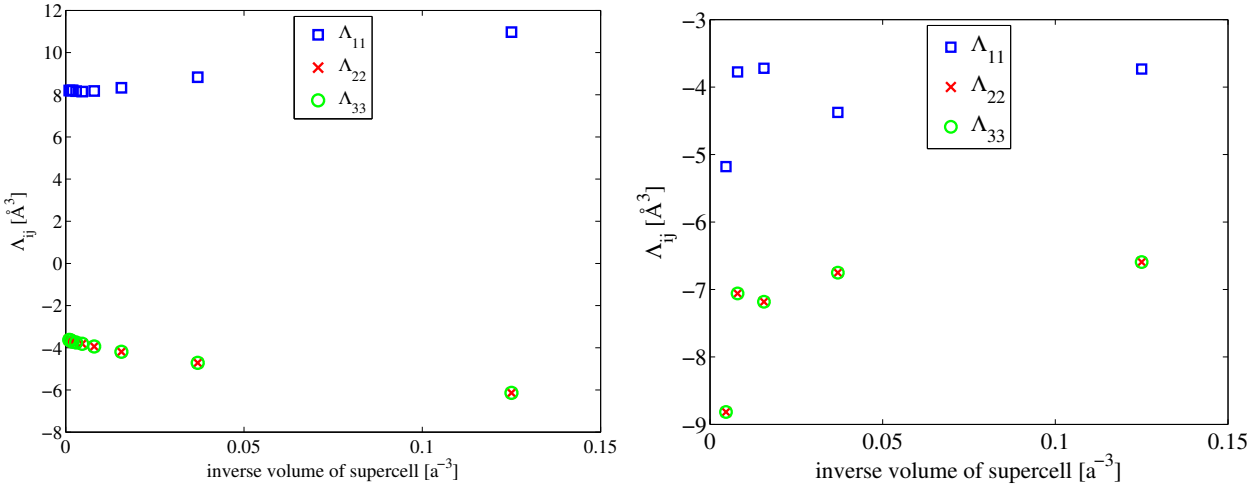


Figure 5.5: *Strain tensor for vacancy formation in $BaZrO_3$ computed using (a) RIM and (b) DFT with LDA xc-functional.*

5.3.1 RIM

In the Rigid Ion Model (RIM) the ions are approximated with hard spheres and the interaction between the ions is computed with relative ease making both large systems and high precision possible. Since the Shell Model (SM) in equation (4.2) is not supported by the LAMMPS software, calculations to assess the effect of an inclusion of the shell model term were performed in DL POLY, a software that, on the other hand, does not include a module for relaxation of the supercell. By mapping out the energy landscape the effect of adding the shell model term was found to be on the order of one unit in the third decimal for a 4x4x4 supercell. This contribution was considered too small to be of importance and the following calculations were conducted in the RIM using the Buckingham potential in equation (4.1) without shell.

The defect strain tensor computed in the RIM model for a vacancy formed along the x-direction, $V_O^{\bullet\bullet}|_x$, can be seen graphically in figure 5.5a in which the three principal components of the defect strain tensor are plotted versus the inverse volume of the supercell. The strain tensor is diagonal due to the high symmetry of a vacancy and the principal directions coincide with the basis vectors of the supercell. The symmetry also requires the second and third strain tensor components, Λ_{22} and Λ_{33}, to be equal.

The linear extrapolation to the dilute limit of infinite volume leads to strain tensor values of 8.2, -3.6, -3.6 and a defect volume of 1.0 Å^3. However, the scaling is not linear in the plot and the residuals upon linear fitting are much larger for the strain as a function of inverse volume that as a function of inverse length of the supercell. This is not a behaviour typical for an elastic dipole [47, 48] but rather like the behaviour of a monopole.

5.3.2 GGA

Figure 5.6a shows the strain tensor computed with DFT using the generalised gradient approximation. Due to the computational cost for DFT calculations the convergence criterion had to be set less strictly and the largest supercell was only a 6x6x6 supercell containing 1080 atoms. The strain tensor is still diagonal but all principal components of the strain tensor are negative leading to negative formation volumes. The extrapolation to infinite size is not as easily performed from the DFT calculation since the data points do not converge conclusively. This depends on the less strict convergence criterion for the relaxation which in turn is amplified by the sensitivity of the strain tensor calculation, a sensitivity which increases with increasing system size.

Figure **??** shows the magnitude of the diagonal components of the strain tensor for various supercell sizes at an increase of the supercell side length by 0.001 Å for a supercell consisting of primitive unit cells of lattice

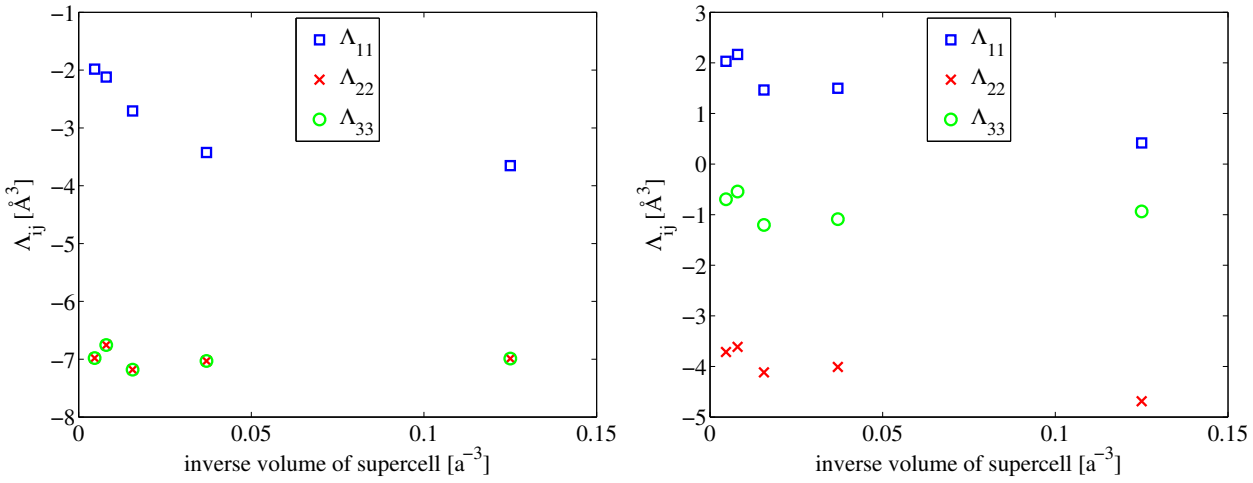


Figure 5.6: *Strain tensor for $BaZrO_3$ with (a) an oxygen vacancy and (b) one proton interstitial using the PBE xc-functional.*

parameter 4.188 Å. Although the relative change in side length is much smaller for larger systems the strain tensor is increasingly sensitive to inaccuracies in the relaxation of the supercell. Figure **??** also shows the strain tensor for a fractional increase of the lattice parameter by $1.5 \cdot 10^{-4}$. This error is within the ionic convergence criterion for relaxation. As can be seen from the figure the error grows rapidly with increasing supercell size and a trade-off has to be made between computational cost and accuracy on one hand and defect-defect interactions on the other hand. The former both suggest smaller systems while the latter requires a larger system in order to be an valid approximation for the dilute limit.

5.3.3 LDA

The strain tensor calculation in the local density approximation (LDA) is somewhat more complicated due to the possibility of tilting of the oxygen octahedra mentioned above. For an odd dimension of the supercell, i.e. 3x3x3 and 5x5x5, the tilting is constrained by the simulation cell and cannot occur. For even dimensions however, the ground state is a tilted system. By creating a vacancy in an aristotype symmetry the configuration will correspond to a saddle point in the energy landscape. A perturbation off the saddle point is required if the symmetry is to be broken.

Figure 5.5b is computed without any perturbations off the saddlepoint, hence no tilts are found in the relaxed configurations. The strain tensor values are of the same order for the LDA as for the GGA. For the two equivalent axes, Λ_{22} and Λ_{33} the values are in agreement between the two functional approximations, whereas for the dipole axis it differs somewhat. The only major deviation is for the largest cell size. This is attributed to the large sensitivity of the strain tensor calculation for large systems. In fact, an increase of the super cell by one part in 10000, or 0.003 Å, would bring the strain tensor for the largest super cell in accordance with the other values. This value is within the convergence criterion of the DFT calculation and high-lights the difficulty of determining a dilute limit value for the strain tensor using DFT calculations.

Since it has been established above that the $Pm\bar{3}m$ symmetry is not the ground state of $BaZrO_3$ in the LDA, attempts were made with vacancies introduced in the I4/mcm symmetry. In this symmetry there are two non-equivalent oxygen positions. These sites are denoted $O^{(1)}$ and $O^{(2)}$ in table 5.3 and the corresponding vacancies are denoted $V^{(1)}$ and $V^{(2)}$. Although the energy of a supercell without defects is 18.6 meV lower for a system of 40 atoms in the hettotype structure compared to the aristotype a vacancy formed in the hettotype structure, be it $V^{(1)}$ or $V^{(2)}$ which have more or less the same energy, is 64 meV higher for a system of 320 atoms, counting the vacancy. This is equivalent to 8 meV in a system of 40 atom. This indicates that a system with a vacancy, started in a I4/mcm symmetry, will converge to the aristotype structure after enough many iterations if the convergence criterion is set more strict. Similar simulations with proton interstitials behave in

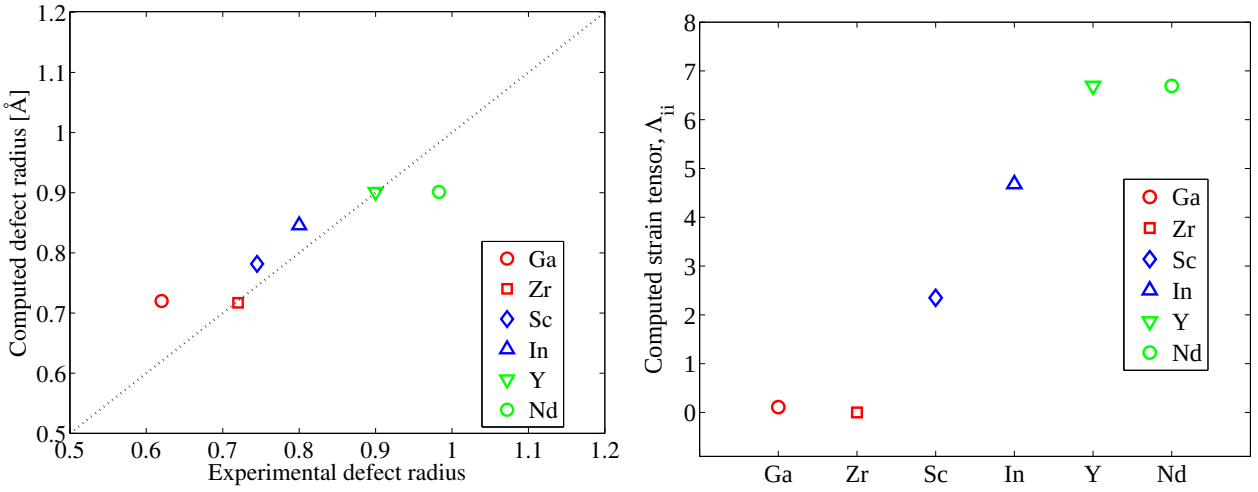


Figure 5.7: *(a) Ionic radii for various substitutional defects and (b) strain tensor principal component in $BaZrO_3$ computed using PBE in a 4x4x4 super cell.*

an opposite manner. A system started in a $Pm\bar{3}m$ symmetry was found to converge in a Imma symmetry for the 2x2x2 and 4x4x4 systems using the LDA.

5.3.4 Proton interstitial

Due to the unresolved question of the ground state symmetry of $BaZrO_3$ in the LDA only the GGA calculations of the defect strain tensor for a proton interstitial $OH^{\bullet}_{O}|_{x}{}^{y}$ are shown in figure 5.6b. The lattice expands in the x-direction and contracts noticeably in the y-direction. The lattice contracts slightly also in the third direction causing a defect volume to be negative. This implies that the lattice actually contracts upon formation of a proton interstitial.

5.3.5 Substitutional defects

The substitutional defect M'_{Zr} is isotropic with diagonal strain tensor with all the three strain tensor components equal. The strain can then be described by a scalar quantity. An often used quantity is the ionic effective radius, which is used to produce the figure 5.7a. This quantity, which is often used to describe the chemical expansion also for non-isotropic defects, is introduced in the following section.

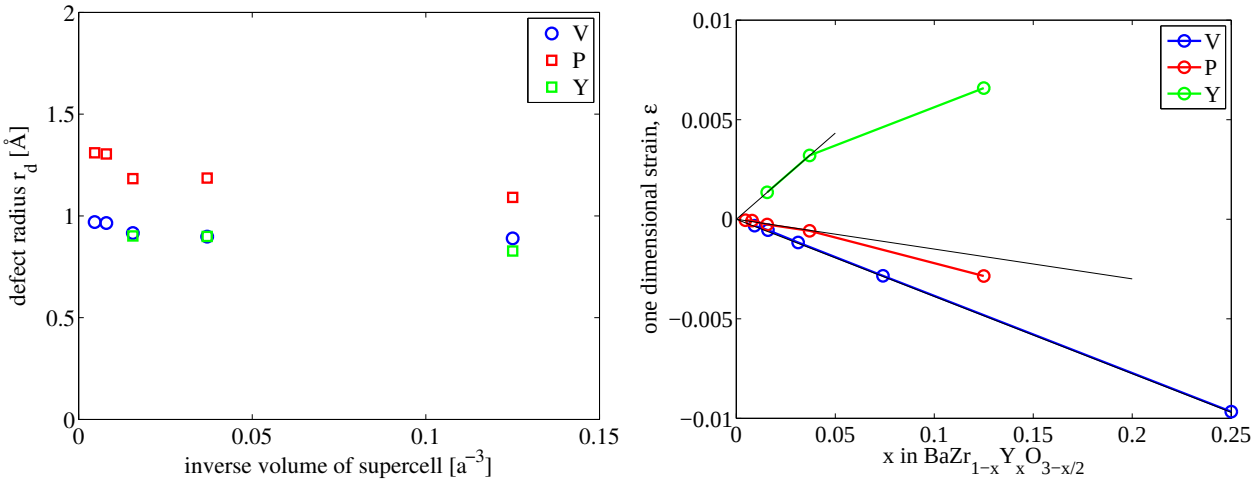
5.4 Effective radius representation

In order to compare the defect strain tensors obtained in this work with experimental values, given as the average of many defects oriented randomly, the ionic radii can be defined from the volume change of the supercell [49, 50]. Assuming the ions can be approximated as close packed hard spheres and the doped and hydrated material can be expressed in the chemical formula

$$BaZr_{1-x}Y_xO_{3-x/2-y/2}(OH)_y \tag{5.2}$$

where x is the atomic fraction yttrium and y the hydration, the pseudo cubic lattice parameter a can be defined as

$$a = 2(\overline{r_B} + \overline{r_O}) \tag{5.3}$$



(a) (b)

Figure 5.8: *Defect radii for $BaZrO_3$ computed (a) using equation* (5.6) *with data obtained from a 4x4x4 supercell and (b) using equation* (5.9) *with data from supercells of various cell sizes containing one defect. The data were obtained using DFT-calculations with the GGA-PBE exchange-correlation functional.*

where

$$\overline{r_B} = (1-x)r_{B_B^\times} + x r_{Y_B'} \tag{5.4}$$

is the effective radius of the B ion and

$$\overline{r_O} = \tfrac{1}{3}\left[\left(\left(3-\tfrac{x}{2}\right)-\tfrac{y}{2}\right) r_{O_O^\times} + \left(\tfrac{x}{2}-\tfrac{y}{2}\right) r_{V_O^{\bullet\bullet}} + y r_{OH_O^\bullet}\right] \tag{5.5}$$

is the effective radius of the oxygen ion. The hydration y is limited by the dopant concentration x and cannot exceed this value. It is important to note that the value of x is based on the fraction of doping rather than the oxygen non-stoichiometry (δ) often found in the literature. The two are related as $x = 2\delta$.

In a real material two defects cannot be isolated from each other, but in a simulation the effect of a single defect can be studied. Thus an isolated proton interstitial represents setting x to zero in equation (5.4) while setting $x = y$ to a finite value in equation (5.5). Similarly, a vacancy represents setting $y = 0$ in the last equation. A substitutional atom can be expressed by the reversed relation.

By computing the cubic lattice parameter as $a = V^{1/3}$ the vacancy ionic radius can be computed as[1]

$$r_V^{(i)} = \tfrac{1}{x}\left(3a - 6r_B - (6-x)r_O\right) \tag{5.6}$$

or preferably, since it does not require knowledge of the ionic radii for r_B and r_O, the difference between a vacancy and an oxygen ionic radius, as

$$\Delta r_V^{(i)} = r_V^{(i)} - r_O = \tfrac{3}{x}(a - a_0) \tag{5.7}$$

where a_0 is the lattice parameter in the ideal lattice and a the parameter in the system with defects. The defect radii computed in this way from DFT calculations using a GGA functional are shown graphically in figure 5.8a for the three types of defects.

Marrocchelli *et al.* [50] extended this further by defining a *chemical expansion coefficient*, α_C, in terms of the one dimensional strain as

$$\begin{aligned}\varepsilon = \frac{a-a_0}{a_0} &= \frac{(\overline{r_B}+\overline{r_O}) - (r_B + r_O)}{(r_B + r_O)} \\ &= \left[\frac{r_Y - r_B}{r_B + r_O} + \frac{1}{6}\frac{r_V - r_O}{r_B + r_O}\right] x = (\alpha_B + \alpha_O)x = \alpha_C x\end{aligned} \tag{5.8}$$

[1]The superscript (i) refers to the defect radius defined from *ionic radii* (equation (5.6)) and (c) refers to the defect radius computed using the *chemical expansion coefficient* (equation (5.9))

This separates the chemical expansion coefficient α_C into a B-ion effective radius change α_B and an oxygen effective radius change α_O. It also shows that the relation between the strain and the concentration must be linear. By rearranging equation (5.8) we can get an expression for the difference between the vacancy and the oxygen ionic radii as[1]

$$\Delta r_V^{(c)} = r_V^{(c)} - r_O = 6(r_B + r_O)(\alpha_C - \alpha_B) = 3a_0(\alpha_C - \alpha_B) \tag{5.9}$$

By defining the one-dimensional strain tensor ε as the average of the principal components of the second rank strain tensor ε_{ij} the ionic radius of the vacancy can be computed from the strain tensor. By only performing calculations on one single defect at a time the calculations of the defect radii can be separated from the radii of other defects, e.g. the effective vacancy radius in equation (5.9) can be separated from α_B.

Figure 5.8b shows the strain tensor as a function of the fraction of doping together with linear fits of the lowest concentration configurations. The figure is obtained from calculations on single defects in supercells of various sizes. It shows that there is indeed a linear dependence assumed in equation (5.8) and that a chemical expansion coefficient α_C and thereby an effective ionic radius can be defined, also for the vacancy. It also indicates that the highest concentration, or equivalently the smallest supercell, is not within this linear approximation and that these values cannot be taken as dilute limited values.

Table 5.4 lists the defect radius differences computed using equations (5.6) and (5.9) respectively. The two methods are in agreement but both overestimate the defect radius difference for the vacancy as well as the proton interstitial. However, the results are qualitatively in agreement with the experimental results. The ionic radii for both the vacancy and the proton interstitial $(\mathrm{OH})_\mathrm{O}^\bullet$ -ion are smaller than for the oxygen ion and the crystal contracts in both cases. The $(\mathrm{OH})_\mathrm{O}^\bullet$ hydroxide ion is also larger than the vacancy causing an expansion during hydration of the lattice. The agreement between the two methods at a supercell consisting of 4x4x4 primitive aristotype unitcells also indicate that the size is large enough for a dilute limit while it is apparent from figure 5.8b that especially a 2x2x2 supercell is too small.

The ionic radius for the yttrium substitutional defect in table 5.4 is in very good agreement with experimental values. The ionic radii for several other substitutional defects are shown in figure 5.7a. The radii have been computed relative to the ionic radius of zirconium $r_{\mathrm{Zr}^{+4}} = 0.72$ Å and are in reasonable agreement with experiments. It seems that the larger the radius the more likely an underestimation becomes.

5.5 Dopant association

All calculations so far have been performed at the dilute limit of infinite cell size or infinitesimal concentration. Defect-defect interaction has been assumed to be minimal or non-existent. This is not the case in a real material.

In order to study the effect of defect-defect interaction and investigate if the combined defects could be described as the sum of two individual defects in the dilute limit systems were prepared with two neighbouring defects of different type. Three different systems were prepared, all containing an yttrium dopant. One system, denoted YV, contained one yttrium dopant Y'_Zr and a nearest neighbour oxygen vacancy $\mathrm{V}_\mathrm{O}^{\bullet\bullet}|_x$ and two systems were combinations of yttrium dopants and proton interstitials in the configurations Y'_Zr and $(\mathrm{OH})_\mathrm{O}^\bullet|_x{}^y$, one as nearest neighbours (YPn) and the second as next nearest neighbours (YPnn) as shown in figure 3.1c.

type	$\Delta r_D^{(i)}$	$\Delta r_D^{(c)}$	Experimental
$\mathrm{V}_\mathrm{O}^{\bullet\bullet}$	−0.47	−0.49	−0.20 [2]
$(\mathrm{OH})_\mathrm{O}^\bullet$	−0.19	−0.18	−0.03 [3]
Y'_Zr	+0.18	+0.18	+0.18 [b]

Table 5.4: Ionic radius increase for studied defects in $BaZrO_3$. The ionic radius increases $\Delta r_D^{(i)}$ were obtained from a 4x4x4 supercell using equation (5.7) and analogous equations. The values for $\Delta r_D^{(c)}$ were obtained using the chemical expansion coefficients defined in equations (5.8) and (5.9) and fitted to the data in figure 5.8b.

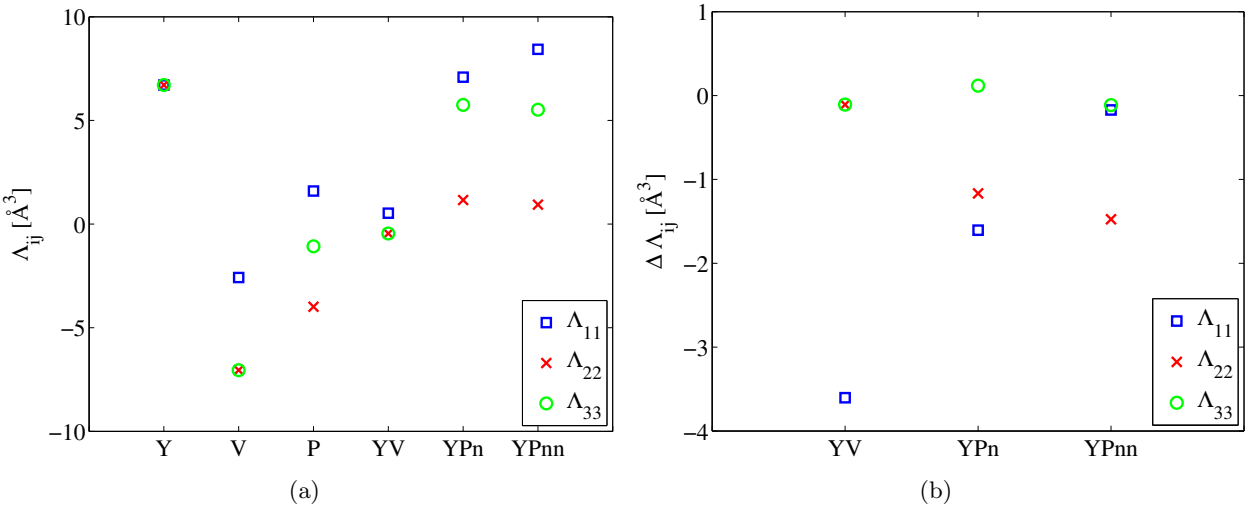


Figure 5.9: *(a) The strain tensor principal components for three individual defects and three combined defects. (b) The difference between the strain tensor for a system of two neighbouring defects and the strain tensor of the two individual defects at the dilute limit.*

The strain tensor values for the individual defects can be seen in figure 5.9a and the difference between the strain tensor computed for the combined system and the two individual defects can be seen in figure 5.9b. As expected the effect is not linear. The chemical expansion computed for the combined system is in general smaller than chemical strain for the sum of individual defects, hence the negative values in figure 5.9b. It is noteworthy that the strain from a single vacancy is highly *an*isotropic but, in combination with the isotropic yttrium defect, the anisotropic effect is almost nullified. Furthermore, the strain is almost zero and the strain tensor is diagonal.

When the material is hydrated the oxygen vacancy is filled with an $(OH)^-$ ion and the material expands in all directions. The defect effective radius increase is 0.40 Å for the nearest neighbour configuration and 0.42 Å for the next nearest neighbour configuration. This is somewhat greater than the linear addition of effects from individual defects as calculated from table 5.4, although in reasonable agreement.

The diagonal components of the strain tensor in figure 5.9b give a more detailed description of the expansion/contraction of the material under combined defects relative to the ideal lattice. Relative to the yttrium-vacancy configuration the yttrium-proton configurations seem to be less non-linear. The difference between the diagonal components of the strain tensor for the yttrium-proton configurations and those for the individual strain tensors are smaller and the trace of the difference is smaller. However, the strain tensors for the yttrium-proton configurations are *non-diagonal*. There are shear strain components, ε_{12}, in the strain tensor for the yttrium-proton configurations of approximately the same magnitude as the smallest diagonal strain component. The corresponding Tait-Bryan angles (see e.g. Goldstein [51] pp.154 & 663) describing this rotation of the principal axes are $\phi = 15°$ and $\phi = 11°$ respectively while $\theta = \psi = 0$. This certainly is a non-linear feature which could not have been obtained from the diagonal single defect strain tensors.

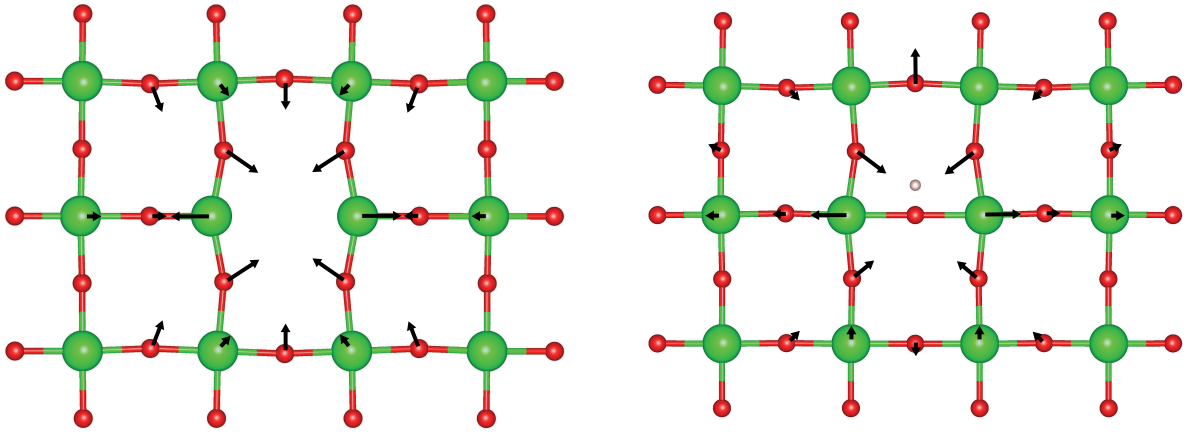
type	$\Delta r_D^{(i)}$				ΔV (Å^3)		
	(1)	(2)	dil.lim.	exp.	(1)	(2)	exp.[4]
Y'_{Zr}	0.18	0.18	0.18	0.18	1.87	1.83	–
$Y'_{Zr} + V_O^{\bullet\bullet}$	−0.51	−0.46	−0.47	−0.20	−0.86	−0.78	–
$Y'_{Zr} + (OH)_O^{\bullet}$	−0.35	−0.29	−0.19	−0.03	−0.59	−0.50	–
$V_O^{\bullet\bullet} \rightarrow (OH)_O^{\bullet}$	0.16	0.16	0.28	0.17	0.279	0.276	0.281

Table 5.5: Ionic radii for defects as computed from a realistically doped material, $BaZr_{0.9}Y_{0.1}O_3$. The ionic radii for $r_D^{(i)}$ were obtained from a 4x4x4 supercell using the ionic radius defined in equations (5.3) to (5.6) and the difference $\Delta r_D^{(i)}$ from equation (5.7) and analogous equations. The dilute limit and the experimental values are restated to facilitate comparison. The volume difference ΔV is given for a 5 atoms unit cell.

5.6 Finite concentration

The defect radii in table 5.5 are computed from a two random configurations of $BaZrO_3$ with $6/64 \approx 10\%$ yttrium doping on the B-sites. The defect radii computed from these configurations are in all cases in qualitative agreement with both the dilute limit values and the experimental values. The yttrium ionic radius is computed to the very same value tabulated by Shannon, both in the dilute limit and in the finite concentration. However, due to the defect-defect interaction the strain tensor is non-diagonal with roughly equal principal components of about 40, which is rather close to six times the defect strain tensor for yttrium, but with Tait-Bryan angles up to about 25° of positive and negative signs, which is a non-linear feature.

Also the defect radius for the vacancy is computed to about the same value as in the dilute limit, which however, is smaller than the experimental value determined by Eriksson Andersson *et al.* [49] The ionic radius of the OH-ion on the other hand is smaller than both the dilute limit value and the experimental value given by Shannon. It seems that there is a longer range defect-defect interaction between protons than between vacancies. What is important however, is that the difference in ionic radius between the OH-ion and the vacancy (0.16 Å) is the about same as the value found by Eriksson Andersson *et al.* as is the volume difference per unit cell (ΔV) upon hydration of the material on the last row in table 5.5. The values are even within 2% of the experimental value in both simulations and within 1% in the first.



(a) (b)

Figure 5.10: *Visualisation of the lattice distortions for (a) a vacancy and (b) a proton interstitial in a plane of oxygen and zirconium atoms. The figure is from an actual configuration and longer arrows signify larger displacements relative to the defect lattice point.*

6 Discussion

The RIM predicts an expansion of the lattice in the x-direction upon formation of a $V_O^{\bullet\bullet}|_x$ vacancy. This can be understood from pure electrostatic considerations in the form of ion-ion interaction in the RIM. Upon removing the doubly negative oxygen ion, a vacancy of charge +2 is formed which repels the positive B-ions and attracts the negative oxygen ions. The second nearest neighbour oxygen however, is repelled by the relocated B-ion and moves in the same direction, albeit to a lesser extent due to the presence of the vacancy. This is in agreement with the results of Freedman *et al.* [52] computed for $SrTiO_3$.

The defect strain tensor as computed using DFT however, is negative in all three directions, and leads to a negative formation volume, regardless of which functional is used. This is in contrast to the RIM model values but in agreement with experimental values [49]. The visualisation of the lattice distortions in figure 5.10a show that the nearest neighbour ions, as computed using DFT, move in the same direction as the in the RIM, but the second nearest oxygen neighbour actually moves towards the vacancy causing the lattice to contract. This can only be understood as an electronic structure feature of redistribution of the electrons of the ions. The assumption that an inclusion of the shell model term to the RIM whould take this into account was proved to be false and the conclusion must be that the RIM is insufficient. It has not been investigated whether other values of the parameters in the Buckingham potential can be found that agree better with the results from the DFT calculations.

The calculations using the LDA were difficult to assess since the question of the ground state structure of $BaZrO_3$ is still not completely resolved. All trial configurations were found to exhibit lower energies than the aristotype $Pm\bar{3}m$-symmetry with the $a^0a^0c^-$ Glazer rotation corresponding to the symmetry I4/mcm having the lowest energy. An attempt by Lebedev *et al.* [37] to start from a P1 symmetry was found to relax to the I4/mcm symmetry, but only after more than 400 iterations. However, the two possible non-equivalent oxygen positions in the I4/mcm have a higher energy than the corresponding vacancy in the aristotype. This would imply that vacancy formation in a hettotype I4/mcm symmetry would fall back into the aristotype upon relaxation. Since a real material includes defects this restoring force may contribute to the non-detection of tilts in experiments.

The calculations of the defect strain tensor using GGA were found to agree reasonably well with the experimental values by Eriksson Andersson *et al.* [49]. The defect radii for both the vacancy and the proton interstitial are somewhat too small in the dilute limit compared to the experimental data. However, the defect radii are smaller than the oxygen ion in both cases which is consistent with the experiments. All point defects investigated exhibit a diagonal strain tensor in the dilute limit with the principal axes parallel to the coordinate axes. This is not true for combinations of single point defects which, with exception for the YV-defect, all cause shear strain, or equivalently a rotation of the principal axes of the strain tensor. It seems that two defects as next nearest neighbours cannot be properly described as the sum of two non-interacting defects but has to be treated as one defect.

The strain tensors for combinations of defects have all been computed from calculations in a 4x4x4 supercell. This size was chosen as the smallest supercell where both defect self-interaction and error in the strain tensor calculation were sufficiently small, but already at this system size the values of the strain tensor may differ by as much as half a unit in either direction. The increasing sensitivity is partly due to large cancellation effects caused by the decreasing values of ε_{ij} at larger system sizes in combination with the amplification caused by the multiplication by the volume of the supercell in equation (3.6).

At higher concentrations the strain tensor is no longer diagonal even for purely doped material with substitutional yttrium without other defects. No two yttrium ions were nearest neighbours in either of the two test configurations but there were several instances of second nearest neighbours. Defect-defect interaction between second nearest neighbours in this inhomogeneous distribution of the dopant atoms is believed to be the origin of the sheer strain. The rather small values of the Tait-Bryan angles describing the rotation of the principal axes of the strain tensor depend intrinsically on the configuration and it is expected that a perfectly homogeneous distribution of the dopants would make the Tait-Bryan angles equal and zero.

While the YP configuration with yttrium and hydroxide ion as nearest neighbours give rise to a non-isotropic strain tensor an YV configuration is almost isotropic. Sundell *et al.* [16] showed that the nearest neighbour

YV configuration is energetically more favourable and thus more likely. This would imply that a dry (or non-hydrated) material has an isotropic strain. Also the YP configuration as nearest neighbours is energetically more favourable [17] and would thus cause trapping of the proton, but this configuration also gives rise to a non-isotropic strain tensor. The effect of an externally applied stress field has not been investigated but it is possible that an applied stress in one direction could make a certain orientations of the YP elastic dipole more favourable than the others or even cause a repulsion between the proton and the yttrium ion. It is then possible that a pre-applied stress field may alter the proton conduction through the electrolyte membrane.

The strain tensors for the point defects in the dilute limit are all diagonal and can never be added linearly to reproduce the strain tensors in the finite concentration. The assumption of linearity in equation (3.3) is thus not strictly true, although the result in some cases is rather good. This probably depends on the defect-defect interaction between defects on neighbouring sites but also on the lattice distortions caused by the defects, e.g. the hydroxide ion is rarely oriented parallel to the coordinate axes in the finite concentration system. It has not been investigated sufficiently in the present work whether it could be possible to compensate for the non-linearity of two neighbouring defects through the difference between the nearest neighbour defect strain tensor and the strain tensors for the individual defects or if two neighbouring defects can be described as one. This remains to be investigated in future studies.

7 Conclusions

The vacancy formation volume in $BaZrO_3$ has been investigated using three different computational models. The rigid ion model based on the Buckingham potential was shown to be insufficient also with an inclusion of the shell model, since it produces a positive vacancy formation volume while both experimental data and data from DFT calculations give negative values.

The DFT calculations in the LDA exhibited a lower energy at a tilt of the oxygen octahedra consistent with a Gazer rotation $a^0a^0c^-$ of 4.2°. This implies that there is a different symmetry, namely the hettotype I4/mcm, which could be the true ground state in the LDA. However, vacancies formed in this hettotype have a higher energy than the aristotype implying a restoring force to the tilt upon formation of a vacancy. The phonon spectrum also showed some residual imaginary modes and the ground state structure has thus not been satisfactorily determined.

The DFT calculations in the GGA using the PBE functional showed a volume expansion of 0.28 $Å^3$ per unit cell upon hydration of a material doped to $6/64 \approx 10\%$. This in remarkably good agreement with experimental values obtained by Eriksson Andersson *et al.* [49] at a finite concentration of 10%. An ionic radius was defined and a value for the radius of the substitutional yttrium ion in the dilute limit as well as in a finite concentration was found to be in very good agreement with ionic radii tabulated by Shannon [41].

The calculated defect radii of the vacancy and the hydroxide ion were somewhat smaller than the experimental values, both in the dilute limit and in the finite concentration simulations. However, while the difference in ionic radius between the vacancy and the hydroxide ion was somewhat larger in the dilute limit compared to experimental data, they were in very good agreement in the finite concentration simulations. Since the experimental data were obtained in the finite concentration the data should primarily be compared at this concentration.

The strain tensor calculations showed that all point defects studied were diagonal in the dilute limit, but when defect-defect interaction is taken into account yttrium dopant and hydroxide ion as nearest neighbours were found to exhibit shear strain. This effect was even more pronounced in the finite concentration and the strain tensor for a combination of defects is not a linear combination of the strain tensors for the individual defects as assumed by Nowic and Berry [18].

The strain tensor formalism showed that point defects have a certain orientation and that the strain in general is non-isotropic. Future studies might investigate if an externally applied stress field could influence the preferred site and orientation of the hydroxide ion and if this in turn can influence the diffusivity of protons in $BaZrO_3$ as well as investigate if chemical expansion can be limited to one dimension only such as not to cause cracking of the electrolyte membrane.

References

[1] Boudghene Stambouli, A. and Traversa, E. "Solid oxide fuel cells (SOFCs): a review of an environmentally clean and efficient source of energy". *Renewable and Sustainable Energy Reviews* (2002).

[2] J.M. Andújar and F. Segura. "Fuel cells: History and updating. A walk along two centuries ". *Renewable and Sustainable Energy Reviews* **13**.9 (2009)., 2309 –2322.

[3] E. Fabbri et al. "Towards the Next Generation of Solid Oxide Fuel Cells Operating Below 600 degrees C with Chemically Stable Proton-Conducting Electrolytes". English. *Advanced Materials* (2012).

[4] H. Iwahara et al. "Proton Conduction in Sintered Osides and its application to Steam Electrolysis for Hydrogen-Production". *Solid State Ionics* (1981).

[5] H Iwahara, H Uchida, and S Tanaka. "High-temperature type Proton Conductor based on SrCeO3 and its Application to Solod Electrolyte Fuel-cells". *Solid State Ionics* (1983).

[6] H. Iwahara et al. "Proton conduction in sintered oxides based on BaCeO3". *Journal of the Electrochemical Society* (1988).

[7] T. Yajima et al. "Proton conduction in sintered oxides based on CaZrO3". *Solid State Ionics* (1991).

[8] H. Iwahara et al. "Protonic Conduction in calcium, strontium and barium zirconates". *Solid State Ionics* (1993). 6TH Eutopean Conf on Solid State Protonic Conductors (SSPC-6), Villard de Lans, France, Sep 06-11, 1992.

[9] K. Kreuer. "Proton-conducting oxides". English. *Annual Review of Materials Research* (2003).

[10] J. Robertson. "Band offsets of wide-band-gap oxides and implications for future electronic devices". *Jourlan of Vacuum Science & Technology B* (2000).

[11] S. Adler. "Chemical expansivity of electrochemical ceramics". English. *Journal of the American Ceramic Society* **84**.9 (2001), 2117–2119.

[12] A. Atkinson and T. Ramos. "Chemically-induced stresses in ceramic oxygen ion-conducting membranes". English. *Solid State Ionics* **129**.1-4 (2000), 259–269.

[13] S. R. Bishop, K. L. Duncan, and E. D. Wachsman. "Thermo-Chemical Expansion in Strontium-Doped Lanthanum Cobalt Iron Oxide". *Journal of the American Ceramic Society* **93**.12 (2010), 4115–4121.

[14] V. Goldschmidt. "The laws of crystal chemistry". German. *Naturwissenschaften* (1926).

[15] A. Glazer. "Simple ways of Determining Perovskite Structures". *Acta Crystallographica Section A* **31**.nov (1975), 756–762.

[16] P. Sundell, M. Björketun, and G. Wahnström. "Thermodynamics of doping and vacancy formation in BaZrO3 perovskite oxide from density functional calculations". English. *Physical Review B* **73**.10 (2006).

[17] M. Björketun, P. Sundell, and G. Wahnström. "Effect of acceptor dopants on the proton mobility in BaZrO3: A density functional investigation". English. *Physical Review B* **76**.5 (2007).

[18] A. Nowic and B. Berry. "Anelastic Relaxation in Crystalline Solids". *Academic Press* (1972).

[19] S. Centoni et al. "First-principles calculation of intrinsic defect formation volumes in silicon". English. *Physical Review B* **72**.19 (2005).

[20] J. Thijssen. "Computational Physics". *Cambridge University Press* (2007).

[21] R. Martin. "Electronic Structure". *Cambridge University Press* (2008).

[22] S. Plimpton. "Fast Parallel Algorithms for Short-Range Molecular Dynamics". *Journal of Computational Physics* **117** (1995), 1–9.

[23] S. Stokes and M. Islam. "Defect chemistry and proton-dopant association in BaZrO3 and BaPrO3". *J. Mater. Chem.* **20** (30 2010), 6258–6264.

[24] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". *Physical Review B* (1964).

[25] W. Kohn and L. Sham. "Self-consistent Equations Including Exchange and Correlation Effects". *Physical Review* (1965).

[26] Becke, Axel D. "Perspective: Fifty years of density-functional theory in chemical physics". *The Journal of Chemical Physics* **140**.18 (2014), –.

[27] J. Perdew, K. Burke, and M. Ernzerhof. "Generalized gradient approximation made simple". English. *Physical Review Letters* (1996).

[28] G. Kresse and J. Hafner. "Abinitio Molecular-Dynamics for Liquid-Metals". *Physical Review B* **47** (1993), 558.

[29] G. Kresse and J. Hafner. "Ab-initil Molecular-Dynamics Simulation of the Liquid-Metal Amorphous-Semiconductor Transition in Germanium". *Physical Review B* **49** (1994), 14251.

[30] G. Kresse and J. Furthmüller. "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set". *Computational Materials Science* **6** (1996), 15–50.
[31] G. Kresse and J. Furthmüller. "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set". *Physical Review B* **54** (1996), 11169.
[32] G. Kresse and D. Joubert. "From ultrasoft pseudopotentials to the projector augmented-wave method". *Physical Review B* **59** (1999), 1758.
[33] P. Blöchl. "Projector augmented-wave method". *Physical Review B* **50** (1994), 17953.
[34] A. Togo, F. Oba, and I. Tanaka. "First-principles calculations of the ferroelastic transition between rutile-type and CaCl2-type SiO2 at high pressures". *Physical Rewiew B* **78** (13 2008), 134106.
[35] J. Bennett, I. Grinberg, and A. Rappe. "Effect of symmetry lowering on the dielectric response of BaZrO3". English. *Physical Review B* **73**.18 (2006).
[36] D. Sahraoui and T. Mineva. "Structural properties of Y-doped BaZrO3 as a function of dopant concentration and position: A density functional study". *Solid State Ionics* **232** (2013), 1–12.
[37] A. I. Lebedev and I. A. Sluchinskaya. "Structural instability in BaZrO3 crystals: Calculations and experiment". English. *Physics of the Solic State* **55**.9 (2013), 1941–1945.
[38] A. I. Lebedev. "Ab initio calculations of phonon spectra in ATiO(3) perovskite crystals (A = Ca, Sr, Ba, Ra, Cd, Zn, Mg, Ge, Sn, Pb)". *Physics of the Solic State* **51**.2 (2009), 362–372.
[39] A. Bilic and J. Gale. "Ground state structure of BaZrO3: A comparative first-principles study". English. *Physical Review B* **79**.17 (2009).
[40] F. Giannici et al. "Long-Range and Short-Range Structure of Proton-Conducting Y:BaZrO3". English. *Chemistry of Materials* **23**.11 (2011), 2994–3002.
[41] R. Shannon. "Revised Effeicive Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides". English. *Acta Crystallographica Section A* (1976).
[42] F. Murnaghan. "The compressibility of media under extreme pressures". English. *Proceedings of the National Academy of Sciences of the United States of America* **30** (1944), 244–247.
[43] C. Hiraiwa et al. ().
[44] S. Yamanaka et al. "Thermophysical properties of BaZrO3 and BaCeO3". English. *Journal of Alloys and Compounds* **359** (2003), 109–113.
[45] R. Khenata et al. "First-principle calculations of structural, electronic and optical properties of BaTiO3 and BaZrO3 under hydrostatic pressure". English. *Solid State Communications* **136**.2 (2005), 120–125.
[46] N. Iles et al. "Atomistic study of structural, elastic, electronic and thermal properties of perovskites Ba(Ti,Zr,Nb)O3". *Computational Materials Science* (2007).
[47] K. Garikipati et al. "The continuum elastic and atomistic viewpoints on the formation volume and strain energy of a point defect". *Journal of the Mechanics and Physics of Solids* **54**.9 (2006), 1929–1951.
[48] B. Puchala, M. L. Falk, and K. Garikipati. "Elastic effects on relaxation volume tensor calculations". English. *Physical Review B* **77**.17 (2008).
[49] A. Eriksson Andersson et al. "Chemical Expansion Due to Hydration of Proton-Conducting Perovskite Oxide Ceramics". *Journal of the American Ceramic Society* (2014).
[50] D. Marrocchelli et al. "Understanding Chemical Expansion in Non-Stoichiometric Oxides: Ceria and Zirconia Case Studies". English. *Advanced Functional Materials* **22**.9 (2012), 1958–1965.
[51] H. Goldstein, C. Pole, and J Safko. "Classical Mechanics". *Pearson* (2002).
[52] D. Freedman, D. Roundy, and T. Arias. "Elastic effects of vacancies in strontium titanate: Short- and long-range strain fields, elastic dipole tensors, and chemical strain". English. *Physical Review B* **80**.6 (2009).
[53] L. E. Malvern. "Introduction to the Mechanics of a Continuous Medium". *Prentice-Hall* (1977).
[54] K. Farahani and R. Naghdabadi. "Conjugate stresses of the Seth-Hill strain tensors ". *International Journal of Solids and Structures* **37**.38 (2000).
[55] H. Callen. "Thermodynamics". *Wiley & Sons* (1960).

A Derivation of the defect strain tensor

One dimensional strain

A rod is a one dimensional object. Assume that the rod initially has a length L_0. It is, by some means, forced into a new length, L. This length can be either larger or smaller than L_0. The elongation is computed as

$$\Delta L = L - L_0 \tag{A.1}$$

such that the elongation is positive (negative) if the length of the rod has increased (decreased). A dimensionless quantity that measures the elongation and is independent of the length of the rod is the strain defined as

$$\varepsilon = \frac{\Delta L}{L_0} \tag{A.2}$$

This is called the *engineering strain* since this is what is often measured in an experiment. There is, in general, no reason to favour L_0 over L and the engineering strain can just as well be defined with L in place of L_0. To relate the elongation to the original length of the rod , i.e. using L_0 in the denominator, is called the *Lagrangian strain measure* and relating the elongation to the final length, L, is called *Eulerian strain measure*.

Under the assumption that the strain is small the Lagrangian strain and the Eulerian strain are approximately equivalent, but under greater strain one would ideally measure the length L at every infinitesimal ΔL in order to compute the *true strain*. The *true strain*, in which the Lagrangian and Eulerian strain are equal, is also defined as

$$\varepsilon = \int_{L_0}^{L} \frac{\mathrm{d}l}{l} = \ln\left(\frac{L}{L_0}\right) \approx \frac{L - L_0}{L_0} \tag{A.3}$$

where the last approximation is valid under small strain and takes us back to the engineering strain, in the Lagrangian description, defined above.

Higher dimensions

In higher dimensions the analogue to the strain ε defined above is the strain tensor ε_{ij}. The derivation requires some elementary continuum mechanics. The following derivation is based on ref. [53].

The Biot strain tensor

In the Lagrangian description of continuum mechanics the final positions $\mathbf{x}$ are functions of the initial positions $\mathbf{X}$. Assuming this function is linear we can write $\mathbf{x}(\mathbf{X}) = \mathsf{F}\mathbf{X}$ or

$$x_i(\{X_j\}) = \mathsf{F}_{ij} X_j = \frac{dx_i}{dX_j} X_j \tag{A.4}$$

A displacement $\mathbf{u}$ of a point $\mathbf{x}$ can be written as $\mathbf{u} = \mathbf{x}(\mathbf{X}) - \mathbf{X}$ and the derivative

$$\frac{\partial u_i}{\partial X_j} = \frac{\partial x_i}{\partial X_j} - \delta_{ij} = \mathsf{F}_{ij} - \delta_{ij} \tag{A.5}$$

It is not obvious at this point that this expression, which is called the *Biot strain*, is the three dimensional analogue of the engineering strain defined above in equation (A.2).

The Lagrangian strain tensor

One potential problem with the Biot strain is that pure rotations can contribute to the strain tensor. A measure that does not contain pure rotations is the increased (squared) length of the infinitesimal distance

$$(\mathrm{d}\mathbf{x})^2 - (\mathrm{d}\mathbf{X})^2 = \mathrm{d}x_k \mathrm{d}x_k - \delta_{ij}\mathrm{d}X_i \mathrm{d}X_j = \left(\frac{\partial x_k}{\partial X_i}\frac{\partial x_k}{\partial X_j} - \delta_{ij}\right)\mathrm{d}X_i \mathrm{d}X_j \tag{A.6}$$

The quantity inside the brackets is called the *Lagrangian strain tensor*[1]

$$\mathsf{L}_{ij} = \frac{1}{2}\left(\frac{\partial x_k}{\partial X_i}\frac{\partial x_k}{\partial X_j} - \delta_{ij}\right) = \frac{1}{2}\left(\mathsf{F}_{ki}\mathsf{F}_{kj} - \delta_{ij}\right) \tag{A.7}$$

For future reference we also include the definition of the *Green deformation tensor* as

$$\mathsf{G}_{ij} = \frac{\partial x_k}{\partial X_i}\frac{\partial x_k}{\partial X_j} = \mathsf{F}_{ki}\mathsf{F}_{kj} \tag{A.8}$$

An advantage by the Lagrangian strain tensor is that it is symmetric. Since $x_i = X_i + u_i$ we get

$$\begin{aligned} 2\mathsf{L}_{ij} &= \frac{\partial x_k}{\partial X_i}\frac{\partial x_k}{\partial X_j} - \delta_{ij} \\ &= \left(\frac{\partial X_k}{\partial X_i} + \frac{\partial u_k}{\partial X_i}\right)\left(\frac{\partial X_k}{\partial X_j} + \frac{\partial u_k}{\partial X_j}\right) - \delta_{ij} \\ &= \left(\delta_{ki} + \frac{\partial u_k}{\partial X_i}\right)\left(\delta_{kj} + \frac{\partial u_k}{\partial X_j}\right) - \delta_{ij} \\ &= \delta_{ij} + \delta_{ki}\frac{\partial u_k}{\partial X_j} + \delta_{kj}\frac{\partial u_k}{\partial X_i} + \frac{\partial u_k}{\partial X_i}\frac{\partial u_k}{\partial X_j} - \delta_{ij} \\ &= \frac{\partial u_i}{\partial X_j} + \frac{\partial u_j}{\partial X_i} + \frac{\partial u_k}{\partial X_i}\frac{\partial u_k}{\partial X_j} \end{aligned} \tag{A.9}$$

which shows that L_{ij} is symmetric.

The Eulerian strain tensor is (after some calculations)

$$2\mathsf{E}_{ij} = \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{\partial u_k}{\partial x_i}\frac{\partial u_k}{\partial x_j} \tag{A.10}$$

When the strain is small $\frac{\partial u_i}{\partial X_j} \approx \frac{\partial u_i}{\partial x_j}$ and the quadratic term can be neglected. In this case $\mathsf{L} \approx \mathsf{E}$ which in some literature is taken as the definition of small strain.

Generalisation of the strain tensor

The Lagrangian strain tensor can be written linearly in terms of the Green strain tensor G as

$$\varepsilon^{(1)}_{ij} = \mathsf{L}_{ij} = \frac{1}{2}\left(\mathsf{G}_{ij} - \delta_{ij}\right) = \frac{1}{2}\left(\mathsf{F}_{ki}\mathsf{F}_{kj} - \delta_{ij}\right) \tag{A.11}$$

In some analogy the Biot strain tensor can be written in terms of "the square root" of the Green strain tensor

$$\varepsilon^{(1/2)}_{ij} = \left(\mathsf{G}^{(1/2)}_{ij} - \delta_{ij}\right) = \left(\mathsf{F}_{ij} - \delta_{ij}\right) \tag{A.12}$$

[1] Note that the name of Lagrange is used in two conceptually different meanings. The *Lagrangian description* of continuum mechanics or the *Lagrangian strain measure* is used for quantities that relate the deformation to the *original positions* (as opposed to Eulerian which relates deformations to the *final* positions) whereas the *Lagrangian strain tensor* is the tensor defined in equation (A.7) which computes the strain tensor from the square of the F-tensor, or equivalently, linearly in G.

A generalisation of this measure to arbitrary powers of the Green strain tensor has been proposed [54] as

$$\varepsilon_{ij}^{(m)} = \frac{1}{2m}\left(\mathsf{G}_{ij}^{m} - \delta_{ij}\right) = \frac{1}{2m}\left(\mathsf{F}_{ij}^{2m} - \delta_{ij}\right) \tag{A.13}$$

where $m > 0$ correspond to the Lagrangian strain measure and $m < 0$ to the Eulerian strain measure. For example, with $m = -1/2$ we get the Eulerian strain tensor

$$\varepsilon_{ij}^{(-1/2)} = -\left(\mathsf{G}_{ij}^{-1/2} - \delta_{ij}\right) = \left(\delta_{ij} - \mathsf{F}_{ij}^{-1}\right) = \left(\delta_{ij} - \frac{\partial X_i}{\partial x_j}\right) \tag{A.14}$$

In the limit of letting $m \to 0$ the definition of the natural logarithm $\lim_{n\to 0} \frac{1}{n}\left(x^n - 1\right) = \ln(x)$ is recognised as a generalisation of the *true strain* in one dimension (cf. equation (A.3)).

$$\varepsilon_{ij}^{(0)} = \ln\left(\mathsf{F}_{ij}\right) \tag{A.15}$$

The transformation matrix F

The definitions of the strain tensors all rely on the matrix F, which contains information about the displacements in the three dimensional body. In the one dimensional case, the length of the rod both prior to and after the elongation, needed to compute the strain, is easily defined. In the three dimensional case F is the matrix containing the relation between the positions of the points that build up a body before and after a deformation.

Assuming the body can be described as a periodic lattice with a unit cell of basis vectors $\mathbf{a}^{(1)}, \mathbf{a}^{(2)}$ and $\mathbf{a}^{(3)}$, the basis vectors can be seen as a coordinate transformation of the kind $\mathbf{a}^{(k)} = \mathsf{L}\,\mathbf{e}^{(k)}$,[2] where $\mathsf{L} = [\mathbf{a}^{(1)}\ \mathbf{a}^{(2)}\ \mathbf{a}^{(3)}]$ is the matrix containing the vectors into which the unit vectors should transform.[3]

Thus the initial lattice of vectors $\mathbf{X}$ can be written as $\mathbf{X}^{(k)} = \mathsf{L}_0\,\mathbf{e}^{(k)}$ and the final lattice as $\mathbf{x}^{(k)} = \mathsf{L}\,\mathbf{e}^{(k)}$. Since the volume of a unit cell is non-zero the determinant of the L_0-matrix is non-zero and L_0 is invertible. We can thus write

$$\begin{aligned} \mathbf{e}^{(k)} &= {\mathsf{L}_0}^{-1}\mathbf{X}^{(k)} \\ \mathbf{x}^{(k)} = \mathsf{L}\mathbf{e}^{(k)} &= \mathsf{L}{\mathsf{L}_0}^{-1}\mathbf{X}^{(k)} \end{aligned} \tag{A.16}$$

or in tensor form

$$x_i^{(k)} = \left(\mathsf{L}{\mathsf{L}_0}^{-1}\right)_{ij} X_j^{(k)} \tag{A.17}$$

which is nothing other than equation (A.4) for the lattice vector number k with $\mathsf{L}{\mathsf{L}_0}^{-1} = \mathsf{F}$. The Biot strain tensor can thus be written as

$$\varepsilon_{ij}^{(1/2)} = \mathsf{F} - \mathbf{1} = \mathsf{L}{\mathsf{L}_0}^{-1} - \mathbf{1} = (\mathsf{L} - \mathsf{L}_0){\mathsf{L}_0}^{-1} \tag{A.18}$$

which shows that the Biot strain tensor is the three dimensional analogue of the one dimensional engineering strain. This is similar to the expression given by Centoni *et al.* [19] albeit with the matrices in a different order. This is of no concern if the matrices commute, which is the case for diagonal matrices. However, if the distortions give rise to non-cubic supercells or if the original supercell is non-cubic the order is of importance. Furthermore, the expression in equation (A.18) is not manifest symmetric and it is possible that rotations are present.

A transformation of a three dimensional body can in general be decomposed in three different actions; translation, rotation and deformation. Translation, as well as rotation, can have no thermodynamical significance, since it does not alter the internal structure of the body, and needs to be removed from the transformation matrix F. Translation are by construction not included and rotations can be removed through a symmetrisation.

[2] The upper index $\mathbf{x}^{(k)}$ denotes a vector while the lower index x_i denotes an element in the vector $\mathbf{x}$.

[3] The letter L, which has hitherto been used to denote the Lagrangian strain tensor will henceforth be used to denote the initial positions for the three dimensional body. The repeated use of the letter is unfortunate but is consistent with the literature.

If F is not already symmetric it can be symmetrized using a polar decomposition.[4] A polar decomposition of a complex matrix F is a decomposition of the form

$$\mathsf{F} = \mathsf{RS} \tag{A.20}$$

where R is a unitary matrix and S is a positive-semidefinite Hermitian matrix. Since only the symmetric S-matrix is of interest, any contribution from rotations must be removed in the calculation of the strain tensor. Consequently, equation (A.18) must be multiplied from the left by R^{-1}. Alternatively, since the square of the F-matrix is related to the S-matrix as

$$\mathsf{F}^2 = \mathsf{F}^\mathsf{T}\mathsf{F} = (\mathsf{RS})^\mathsf{T}(\mathsf{RS}) = \mathsf{S}^\mathsf{T}\mathsf{R}^\mathsf{T}\mathsf{RS} = \mathsf{S}^\mathsf{T}\mathsf{S} = \mathsf{S}^2 \tag{A.21}$$

contributions from rotations can be eliminated through the use of the Lagrangian strain, which is quadratic in F and thus, once again, manifest symmetric by construction.

Writing out the Lagrangian strain tensor in terms of L and L_0 we get the expression

$$\varepsilon^{(1)} = \frac{1}{2}\left(\mathsf{F}^\mathsf{T}\mathsf{F} - \mathbf{1}\right) = \frac{1}{2}\left(\left(\mathsf{L}\mathsf{L}_0^{-1}\right)^T\left(\mathsf{L}\mathsf{L}_0^{-1}\right) - 1\right) = \frac{1}{2}\left(\mathsf{L}_0^{-1}\right)^T\left(\mathsf{L}^T\mathsf{L} - \mathsf{L}_0^T\mathsf{L}_0\right)\mathsf{L}_0^{-1} \tag{A.22}$$

Comparing the different strain tensors

Computing the true strain tensor ε_{ij} requires computation of the matrix logarithm. By diagonalising the S-matrix $\mathsf{S} = \mathsf{Q}\Lambda\mathsf{Q}^T$ we get

$$\begin{aligned}\varepsilon^{(0)} &= \ln(\mathsf{F}) = \ln(\mathsf{Q}\Lambda\mathsf{Q}^T) = \mathsf{Q}\ln(\Lambda)\mathsf{Q}^T \\ &= \mathsf{Q}\ln(\mathbf{1} + \Lambda - \mathbf{1})\mathsf{Q}^T \\ &= \mathsf{Q}\left((\Lambda - \mathbf{1}) + \mathcal{O}(\Lambda - \mathbf{1})^2\right)\mathsf{Q}^T \\ &\approx \mathsf{Q}\Lambda\mathsf{Q}^T - \mathsf{Q}\mathsf{Q}^T = \mathsf{Q}\Lambda\mathsf{Q}^T - \mathbf{1} = \mathsf{F} - \mathbf{1}\end{aligned} \tag{A.23}$$

which is recognised as the Biot strain tensor $\varepsilon^{(1/2)}$. One might then guess that including higher order terms one would retrieve the Lagrangian strain tensor, but this is not true. Including second order terms gives an expression which includes both second and first order terms and is thus not manifest symmetric.

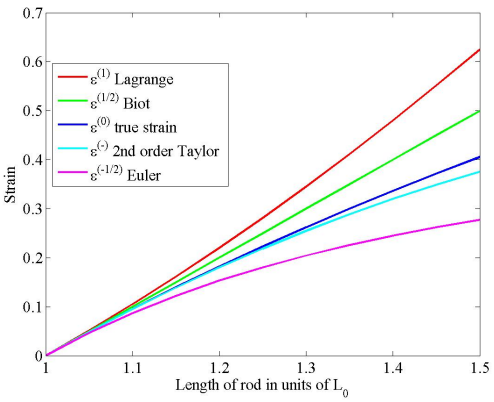


Figure A.1: *Comparison of the different strain tensor measures.*

Figure A.1 shows the strain as a function of final length of a one dimensional rod for several different strain measures. When the final length equals the initial length, $\mathsf{L} = \mathsf{L}_0$, all strain measures are zero. As the elongation increases the strain measures start to deviate from each other and at an elongation of 10% the Biot strain overestimates the strain by about 5% relative to the true strain and the Lagrangian strain overestimates by about 10%. The Eulerian strain on the other hand underestimates the strain by about 9%.

Which strain tensor gives the most accurate result is a still debated issue. It seems to depend on the material and circumstances under which the strain is measured. Despite its name the true strain is not necessarily more true than any other strain. Since the F-matrix appears linearly in both the true strain and the Biot strain, these two strain measures are not necessarily symmetric, which any strain that only describes deformations of a body has to be. Translations can be removed by measuring all positions relative to a point in the body, and rotations can be removed by a polar decomposition described above. However, the simplest way is to use the Lagrangian strain measure which is quadratic in F (see equation (A.21)) and thus is manifest symmetric. This also implies that no polar decomposition is necessary and that the simple formula in equation (A.22) can be used.

[4] Some authors [55] construct a symmetric strain tensor from $\mathsf{F}_{ij} - \delta_{ij}$ by adding its transpose. One then gets

$$\mathsf{F}_{ij} - \delta_{ij} + \mathsf{F}_{ji} - \delta_{ji} = \frac{\partial u_i}{\partial X_j} + \frac{\partial u_j}{\partial X_i} \tag{A.19}$$

which is nothing other than the small strain limit of the Lagrangian or Eulerian strain tensor. The polar decomposition however, does not require small strain and produces a symmetric S-matrix which contains more information.