UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING AND ENVIRONMENT
Engineering Sciences

Ab-Initio Investigation into catalyst supports for Polymer Electrolyte Fuel Cells

Matthew Clive Worsdale
Thesis for the degree of Doctor of Philosophy

December 2016
ABSTRACT

Engineering Sciences

Doctor of Philosophy

AB-INITO INVESTIGATION INTO CATALYST SUPPORTS FOR POLYMER ELECTROLYTE FUEL CELLS

by Matthew Clive Worsdale

Abstract

One of the most promising families of catalyst support materials from the point of view of durability in the PEFC are metal oxides. \(\text{SnO}_2\) can only corrode, in a strong acid, at potentials \(\geq 1.4\text{V}\), progressing through surface hydroxide formation. The bulk stability is backed up by simulated potential cycling, with minimal ESCA loss at 1.6V. The conductivity of pure \(\text{SnO}_2\), which forms the rutile crystal structure, is very small compared to C but can increase by orders of magnitude with formation of certain intrinsic defects which create electron donor levels in the band gap (which, at 3.6 eV, classes the undoped oxide as a wide band gap semiconductor).

The ab-initio study evaluates the potential of extrinsic dopants to increase conductivity through adding \(n\)-type charge carriers to the system, investigating the influence of Ta on the thermodynamic stability and electronic properties through DFT calculations. Particular focus has been given to Ta, as \(\text{TaO}_2\) also forms with the rutile structure. This allows investigation of the full range of Ta doping concentration from an isolated defect to alloy by including it as a substitutional defect. A cluster expansion parameterised in terms of binary occupation variable that equal Sn or Ta is used to calculate the orderings (distributions of Ta:Sn atoms over the lattice at fixed concentration). It is found that alloys are stabilized thermodynamically by lattice distortions, with the creation of donor levels near the conduction band minimum (CBM) that with the relaxation of the lattice turn into states lying deep in the band gap that are calculated to give inadequate donation of electrons to the conduction band. This is indicative of collaborating Jahn-Teller active (JT) distortion modes of the disordered crystallite oxide which is shown through analysis of the spin density. The results suggest that alloying of metals in oxide systems is not a feasible approach to increasing conductivity with thermodynamic stability due to JT effects.

The second part of this thesis explores the potential of hydrogen defects, which have been shown to form at interstitial and host oxygen sites under laboratory conditions. The same
methodology is followed to establish the efficacy of hydrogen as an electron donor and its relation to other defects in SnO₂. Much theoretical work has been devoted to the study of hydrogen in n-type semiconductors and experimental work supports that hydrogen can be expected to act as a conventional n-type donor in SnO₂. It is found that interstitial H not only makes SnO₂ metallic as recorded by the electronic structure, but further that the defect can bring about a large downward shift in the effective band gap. This also raises the possibility of activating Ta-induced deep donor levels.
Contents

1 Author’s Declaration 14

2 Acknowledgements 15

3 INTRODUCTION - AIMS AND OBJECTIVES 16
   3.1 Statement of the research problem ........................................... 16
   3.2 Structure of this thesis ......................................................... 16

4 LITERATURE REVIEW 17
   4.1 The Polymer Electrolyte Fuel Cell - working principles ............... 17
      4.1.1 Fuel Cells - General remarks ............................................ 17
      4.1.2 Polymer Electrolyte Membrane Fuel Cell (PEMFC) .................... 17
      4.1.3 Charge Transport and the Polymer Electrolyte Membrane (PEM) .... 18
      4.1.4 Electrodes and their structure .......................................... 19
      4.1.5 Electrochemical behaviour ............................................... 20
   4.2 Support materials for Pt supports .......................................... 23
      4.2.1 Carbon Corrosion ............................................................ 23
      4.2.2 State of research in the field - overview .............................. 25
      4.2.3 General degradation mechanisms for carbon based supports ......... 25
      4.2.4 Existing support- Carbon Black (CB) .................................... 26
      4.2.5 Carbon Allotropes ............................................................ 27
      4.2.6 Phosphides and Oxyphosphides .......................................... 29
      4.2.7 Nitrides ............................................................................ 30
      4.2.8 Carbides ............................................................................ 31
      4.2.9 Non-carbon based oxides .................................................... 31
      4.3 Density Functional Theory; review and technical introduction ....... 35
         4.3.1 The N-body electron problem ............................................. 35
         4.3.2 The Variational Principle and Hartree-Fock theory .................. 36
         4.3.3 The Kohn-Sham equations ................................................. 38
         4.3.4 Improvements on basic DFT- the exchange correlation functional $E_{xc}$ 42
         4.3.5 The Local Density Approximation (LDA) ............................... 42
         4.3.6 Generalised Gradient Approximation (GGA) ............................ 43
         4.3.7 Hybrid Functionals ............................................................ 44
         4.3.8 Brief evaluation of DFT suitability and outlook ...................... 44
         4.3.9 Plane Wave vs. Gaussian basis functions .............................. 45
         4.3.10 Observable properties of DFT solution- bandstructure and charge density differences ......................................................... 48
   4.4 Crystal structures of $\text{SnO}_2$ ................................................. 50
      4.4.1 Phase competition of disordered $\text{SnO}_2$ hosts .......................... 51
      4.4.2 Rutile structural parameters .................................................. 53
   4.5 Lattice defects of rutile $\text{SnO}_2$ .............................................. 54
## 4.5 Intrinsic Defects of SnO2

- **4.5.1 Intrinsic defects of SnO2**
  - Chapter 4, Section 5.1, page 54
- **4.5.2 Extrinsic Dopants of SnO2**
  - Chapter 4, Section 5.2, page 55
- **4.5.3 Crystal Field Splitting (CFS)**
  - Chapter 4, Section 5.3, page 57
- **4.5.4 Jahn-Teller (JT) effects in the solid state**
  - Chapter 4, Section 5.4, page 58
- **4.5.5 Hydrogen doping of semiconductors**
  - Chapter 4, Section 5.5, page 59

## 4.6 N-Type and P-Type Doping and the Fermi Level

- **4.6.1 Amphoteric doping and the charge neutrality level (CNL)**
  - Chapter 4, Section 5.6, page 66
- **4.6.2 Band filling effects**
  - Chapter 4, Section 5.7, page 68
- **4.6.3 Defect energetics and thermodynamics**
  - Chapter 4, Section 5.8, page 69

## 5 METHODOLOGY

- **5.1 Properties of a replacement support material studied by ab initio methods**
  - Chapter 5, Section 5.1, page 70
- **5.2 Ab initio estimation of conductivity - Drude model**
  - Chapter 5, Section 5.2, page 71
- **5.3 Ab Initio Modelling of Defect Structures**
  - Chapter 5, Section 5.3, page 73
  - **5.3.1 Defect energetics and scaling**
  - Chapter 5, Section 5.3.1, page 73
  - **5.3.2 The Supercell method**
  - Chapter 5, Section 5.3.2, page 74
  - **5.3.3 Correction Methods**
  - Chapter 5, Section 5.3.3, page 76
- **5.4 Charge Difference Maps**
  - Chapter 5, Section 5.4, page 80
- **5.5 Bulk Phase Diagram Methodology**
  - Chapter 5, Section 5.5, page 81
  - **5.5.1 Mixtures and multi-phase behaviour - The Convex Hull**
  - Chapter 5, Section 5.5.1, page 81
  - **5.5.2 Thermodynamics of Pure compounds from DFT data**
  - Chapter 5, Section 5.5.2, page 82
  - **5.5.3 Virtual reference states**
  - Chapter 5, Section 5.5.3, page 84
- **5.6 Exploring phase space near stoichiometric compounds**
  - Chapter 5, Section 5.6, page 84
  - **5.6.1 Cluster expansions (CE) - configurations**
  - Chapter 5, Section 5.6.1, page 85
  - **5.6.2 Implementation of Cluster Expansion (CE) - linking CRYS- TAL09 to custom code (ATAT)**
  - Chapter 5, Section 5.6.2, page 87
  - **5.6.3 Multiphase co-existence, competition, metastability**
  - Chapter 5, Section 5.6.3, page 90
- **5.7 Intrinsic defect calculations in SnO2**
  - Chapter 5, Section 5.7, page 90
- **5.8 Hydrogen defects in SnO2**
  - Chapter 5, Section 5.8, page 92
  - **5.8.1 Hydrogen defects and calculation details**
  - Chapter 5, Section 5.8.1, page 92

## 6 RESULTS

- **6.1 Bulk Defect Calculations on rutile SnO2**
  - Chapter 6, Section 6.1, page 95
  - **6.1.1 Bulk calculations in VASP and CRYS- TAL09**
  - Chapter 6, Section 6.1.1, page 95
  - **6.1.2 Bulk oxygen vacancy**
  - Chapter 6, Section 6.1.2, page 99
  - **6.1.3 Conclusions**
  - Chapter 6, Section 6.1.3, page 102
- **6.2 Ta-doped SnO2**
  - Chapter 6, Section 6.2, page 104
  - **6.2.1 Motivation for study of Ta-doped SnO2 and exploration of Ta-Sn-O ternary phase space**
  - Chapter 6, Section 6.2.1, page 104
  - **6.2.2 DFT Exploration of Ta-O-Sn ternary space**
  - Chapter 6, Section 6.2.2, page 104
  - **6.2.3 Cluster Expansion - Enforcement of Rutile Symmetry**
  - Chapter 6, Section 6.2.3, page 107
  - **6.2.4 Cluster Expansion, no enforcement of symmetry**
  - Chapter 6, Section 6.2.4, page 109
  - **6.2.5 Spin density and electron localisation**
  - Chapter 6, Section 6.2.5, page 115
### 6.2.6 Donor activation as a function of concentration .................................................. 118
### 6.2.7 Carrier Statistics of extrinsic doping ................................................................. 120
### 6.2.8 Conductivity ......................................................................................................... 122
### 6.2.9 Discussion- Effect of symmetry restriction in cluster expansions ......................... 122
### 6.2.10 Crystal symmetry and Jahn-Teller effects ......................................................... 123
### 6.2.11 Conclusions ........................................................................................................ 126

### 6.3 Hydrogen defects in bulk rutile SnO$_2$ ................................................................. 128
#### 6.3.1 Stability of substitutional and interstitial defects ............................................ 128
#### 6.3.2 Defect lattices and bandstructures ................................................................. 129
#### 6.3.3 Electronic structure ............................................................................................ 130
#### 6.3.4 Electron distribution- charge difference maps .................................................. 136
#### 6.3.5 Conductivity ....................................................................................................... 137
#### 6.3.6 Discussion-Defect Stability .............................................................................. 139
#### 6.3.7 Electronic structure and conductivity .............................................................. 141
#### 6.3.8 Evidence for donor state - Electronic structure and charge density ............... 144
#### 6.3.9 Conclusion ......................................................................................................... 146

### 7 OUTLOOK 147
#### 7.1 Rigorous treatment of Jahn-Teller effect ............................................................. 147
##### 7.1.1 Engineering of symmetry and electronic properties of structures in cluster expansion (CE) ................................................................. 147
##### 7.1.2 Alternative extrinsic dopants .......................................................................... 148
#### 7.2 Stability as the oxidation voltage at doping limit ............................................... 149
#### 7.3 Final Concluding Remarks .................................................................................. 151

### 8 APPENDIX - CRYSTAL09 Basis Sets 153
#### 8.1 Hydrogen ............................................................................................................ 153
#### 8.2 Oxygen - Ta$_x$Sn$_{1-x}$O$_2$ .................................................................................. 153
#### 8.3 Oxygen - Oxygen Vacancy and Hydrogen calculations ...................................... 154
#### 8.4 Tin - Oxygen Vacancy and Hydrogen Defect calculations ................................ 155
#### 8.5 Tin - Ta$_x$Sn$_{1-x}$O$_2$ ....................................................................................... 155
#### 8.6 Tantalum ............................................................................................................. 155

**Glossary of key terms** 157

**List of notation and symbols** 159
List of Figures

4.1 Schematic illustration of the Polymer Exchange Membrane of an acidic-mode PEFC. Above-right image shows the basic structure with the hydrophobic fluorocarbon backbone and the sulphonate acid group, taken from reference [19].

4.2 Schematic illustration of the three phase boundary (at the anode) between the three important regions, clockwise from top: Pore of carbon electrode containing hydrogen gas from the flow field plates/gas diffusion assembly (figure (4.3)), the aqueous polymer electrolyte phase consisting of electrolyte phases containing aqueous H\(^+\), and the solid phase with carbon particles and Pt nanoparticles.

4.3 The Gas Diffusion Layer (GDL) and flow field plates (FFP) that aid the diffusion of hydrogen and oxygen gases to the anode and cathode, respectively.

4.4 The principal loss mechanisms of the PEFC based on equation (14), after references [9], [3].

4.5 TEM image of Carbon Black (CB) based catalysts, before corrosion testing left: Carbon Black without graphitisation; right: Carbon Black with graphitisation. Taken from reference [59].

4.6 TEM image of graphene sheet supporting Pt catalyst nanoparticles on surface. Taken from reference [62].

4.7 TEM images of Pt catalysts supported on CNF (left) and CNT (right).

4.8 "Jacob’s Ladder" illustration of successive addition of contributions to the DFT ground state energy from the least accurate approximation to the most accurate. The Hartree Approximation excludes the exchange-correlation energy \(E_{XC}\), which is represented in DFT by the LDA and GGA gradient based approximations. The meta-GGA approximation includes an approximation to the kinetic energy through a system of non-interacting electrons. DFT+U adds the Hubbard parameter U to parameterise the interaction between the spins of electrons.

4.9 Unnormalized Gaussian Type Functions (equation (50)) used in the CRYSTAL09 program for 1s (red line) and 2p (green line) orbitals. Also plotted is a linear combination of these two orbitals (blue line), shifted by a transformation \(r \rightarrow r - 0.19\) to give \(\varphi(r = 0) = 0\), to illustrate the operation of construction of AO from superposition of Gaussian Type Functions used in the CRYSTAL program.
4.10 Stable and possible disordered crystal structures of SnO$_2$. All coordination polyhedra are shaded grey. **Top left:** The rutile structure is derived from a hexagonal close packed (HCP) lattice of oxygen atoms with the unit cell indicated by the dashed yellow line. The tin atoms occupy the octahedral interstitial positions, while the unused tetragonal interstitial sites (in which an Sn would be coordinated to four oxygen atoms) are indicated for comparison. **Centre:** The rutile structure arises due to the extension of two of the six ligands (in green) which gives planes of distorted octahedra offset in the $\bar{c}$ direction and stacked vertically, sharing edges, which makes the structure thermodynamically stable by maximising the closest-sphere packing of atoms. The three fold coordination of oxygen by tin is highlighted in red. **Right:** View of the rutile and CaCl$_2$-type structures in the $\bar{c}$ direction. They are distinguished by the rotations of coordination octahedra about the shared edge, with the upper image (rutile) demonstrating the clockwise and anti-clockwise rotations of alternating planes. The lower image shows the CaCl$_2$ structure in which neighbouring octahedra rotate to maintain an angle between them with respect to the position in the rutile structure. Both images taken from reference [141]. **Bottom Left:** A unit cell of the anatase structure, in which the coordination octahedra have three different vertex lengths highlighted by yellow, green and red lines. The octahedron lying at the centre of the unit cell shown shares two further edges (in yellow) with octahedra either side of it in the $\tilde{a} \tilde{b}$ plane to make the ‘zigzag’ packing seen in the directions $\tilde{a}$ and $\tilde{b}$. The two longest vertices (in red) are not shared.

4.11 Crystal Field Splitting raises energy of two $d$ states $e_g$ in an octahedrally coordinated system, lowering the energy of the other three $d$ states $t_{2g}$. Jahn Teller (JT) distortions introduce an additional fine structure to the $d$ states, removing the degeneracy of the $t_{2g}$ states by splitting them into levels $\delta_{2/3}$ and $\delta_{1/3}$.

4.12 **Left:** Photo-thermal Ionization Spectroscopy (PTIS) spectra of a sample containing acceptor centres A(H,Si) and aluminum acceptors. Taken from reference [180]. **Right:** Schematic illustration of two stage photothermal excitation process at two different temperatures corresponding to the spectra of the first image. In the first stage, incident light at or below a frequency $\omega$ gives photothermal excitation of holes from a shallow acceptor level to the valence band maximum. In the second stage the holes are transferred between energies of excited states of the valence band minimum by thermally induced phonons.

4.13 From left to right: The basic bandstructure of an intrinsic semiconductor with energy gap $E_g$ and fermi level $E_F$. Bandstructure of a semiconductor with only n-type doping, with electrons at the donor level $E_D$ and a raised fermi level. Bandstructure with only p-type doping, with holes at an acceptor level $E_A$ and a lowered fermi level.

4.14 Schematic illustration of the charge neutrality (CNL) level in the presence of amphoteric doping in an extrinsic semiconductor.
5.1 Top: The $xy$-plane of the SnO$_2$ supercell as referred to in the text. The reciprocal lattice vector $K_2$ shown on figure (6.1) is out of the page. Bottom: View of the $z$-layering of the supercell layers.

5.2 The formation energy $G$ of a defect as a function of inverse supercell size is calculated within DFT is fitted to scale factor enlargements 2x2x2, 3x3x3 and 4x4x4 of the unit cell and extrapolated to the infinite size (dilute limit).

5.3 The ground state of systems formed exclusively out of a binary compound $A_xB_{1-x}$ is indicated by the red line. Systems for which $y < z < 1$ including those on the convex hull have a composition of the form $A_yB_{1-y_1}A_yB_{1-y_2}B_k$, where $i + j + k = 1$. The region PQR forms the phase boundary containing all systems of this form.

5.4 Illustration of possible clusters in a face centered cubic (FCC) lattice; considering nearest neighbours (top) and second nearest neighbours (bottom).

5.5 Flow diagram to illustrate procedure for preparing structures in the space Ta$_x$Sn$_{1-x}$O$_2$ calculated in the cluster expansion with ATAT.

5.6 Generalised phases $\alpha$ and $\beta$ in a binary composition space with end limits A and B and Gibbs Free Energy as a function of the molar fraction $X_B$. Phase $\gamma$ exists in the same composition space but has only a narrow homogeneity range relative to $\alpha$ and $\beta$ and will compete with these phases.

5.7 Interstitial hydrogen defect types ‘A’ and ‘B’, as described in text, left and right respectively.

6.1 The path in reciprocal space used to obtain the bandstructure for SnO$_2$ and defected structures throughout.

6.2 Total electronic energy after relaxation as a function of the number of k-points in the Brillouin Zone of the SnO$_2$ unit cell in the LDA and GGA (PW91) [186], [187] functionals as calculated in VASP. The two functionals give energies over different scales, which is arbitrary in DFT as discussed in section (5.5.2). In both cases a k-point density greater than 250 points ($\AA$)$^{-3}$ was needed for convergence.

6.3 Bandstructure of SnO$_2$ as calculated in the GGA approximation (top) and LDA (bottom). The $\Gamma$, X, and M points are described by figure (6.1). The red line on both plots indicates the Fermi level $E_F$.

6.4 Bandstructure of rutile SnO$_2$ as calculated with the CRYSTAL09 code (based on Gaussian Type Functions), using the B3LYP hybrid functional.

6.5 Total electronic energy after relaxation as a function of the number of k-points in the Brillouin Zone of the SnO$_2$ unit cell. Each data point is the difference between energy values for each next successive pair of k-points.
LIST OF FIGURES

6.6 Formation energy of neutral oxygen vacancy defect $\Delta G_f^{vac(O)}$ obtained by removing an O atom and redistributing the valence electrons from the Sn=O with the spatial states of the electron preserved (see text), for unrelaxed geometry (green line) and relaxed geometry (blue line), Scaling is with the effective inverse lattice constant $1/a = x^3$. The supercells used correspond to numbers 1 to 6 listed sequentially in table (3). While the cells are not all of the same shape, they are equivalent from a computational point of view because the vacancy defect is modelled as neutral (see text).

6.7 Bandstructure for the Vac(O) defect in bulk rutile SnO$_2$ calculated in the dilute limit (162 atom supercell, number '6' in table (3)). Charge neutrality of the unit cell was enforced by redistributing the charge of two electrons left at the site of the oxygen vacancy equally between all tin atoms.

6.8 Ternary phase diagram for Ta-O-Sn system obtained from DFT using all known stochiometric compounds (left), and using only the set of compounds for which experimental formation energies were used from reference [215]. (right)

6.9 Formation Energies of all structures in the cluster expansion (CE) performed with enforcement of the preservation of rutile lattice. Rutile TaO$_2$ and SnO$_2$ are the reference states. The minimum energy surface or 'convex hull' for the cluster expansion is indicated by the black line. At each discrete composition is a ground state (gs) whose corresponding structure lies on the convex hull. The greatest stabilisation by 0.09 eV occurs for a Ta loading of 66.6%.

6.10 The true formation energies of structures in the cluster expansion (CE) $E_{DFT}$ and the corresponding values as predicted by the code based on least squares fitting, $E_{CE}$.

6.11 Rutile Symmetry operators belonging to structures in the cluster expansion, and their thermodynamic stability.

6.12 Minimum energy surface 'convex hull', with no restrictions on the symmetry. Meaning of the plotted quantities is the same as in figure (6.9). Dashed red line is 'convex hull' of the stochiometric compound SnTa$_2$O$_6$ whose energy has been calculated with rutile TaO$_2$ and SnO$_2$ as reference states.

6.13 Bandstructure and Density of States (DOS) for the structure in the unconstrained CE having Ta = 50 %. The DOS is projected both onto the atomic species (Ta,Sn,O) and the two main orbital types ($sp$ hybridised and $d$ orbitals). The path in the Brillouin zone is the same as used in SnO$_2$ intrinsic defect calculations, figure (6.1). Dashed and solid lines represent the bands for $\alpha$ and $\beta$ electrons (spin-up and spin-down). The donor level $E_D$ which has been introduced by substitutional doping of Ta in SnO$_2$ lies in the band gap as shown. It can also be seen that there is little covalency between Sn and O, suggesting that the bonding of SnO$_2$ is more ionic than has been suggested elsewhere.
6.14 The bandstructure for the same structure as shown in Figure (6.13) that has not been allowed to relax. Defect states appear higher up in the band gap than in the fully relaxed case and are more disperse, with the Fermi level $E_F$ within the defect states. The density of states projected onto atomic species shows that the defect states are mostly related to Ta electronic states.

6.15 **Left:** Spin density for the CE structure in the Ta lean limit with composition Ta : Sn = 1 : 5 with view parallel to the (111) plane of the untransformed rutile lattice, with view chosen to give an orthogonal projection. **Right:** For the structure in the Ta-rich limit with composition Ta : Sn = 3 : 1, as viewed in the (00$_n$) plane ($n = \frac{c}{2}$) with an orthogonal projection.

6.16 Spin density around the Ta-O polyhedra shown in the Ta-lean limit on figure (6.15). **Left:** Indicated are the apical bond $a_1$ and equatorial bond $c_1$ that both appear to have distorted; **Right:** The same equatorial bond, with apical bond out of the page.

6.17 Depth of the defect state $E_C - E_D$ introduced by Ta in the band gap which may act as a donor state, as a function of Ta concentration. The range from 0-16.67% has been calculated in an approximation of the dilute limit (isolated point defect in bulk rutile SnO$_2$, with preservation of rutile symmetry enforced). The range with Ta > 16.67% has been calculated with the cluster expansion (CE). One can see that the CE can be directly extrapolated to the dilute limit, the point where the depth of the level has converged within the supercell approximation of that limit. The energy of the SnO$_2$ intrinsic oxygen vacancy defect Vac(O) of 2.5 eV is also indicated; for $|E_C - E_D| < $ Vac(O) Ta doping will not be feasible.

6.18 Depth of the defect state as calculated in the supercell approximation to the dilute limit. Second-order polynomial fit was used to fit the points for the three low Ta concentrations calculated using 2 x 2 x 2, 3 x 3 x 3 and 4 x 4 x 4 supercells.

6.19 Ionised Ta concentration $x$ in the dilute limit at T = 300K. A peak in carrier concentration occurs at $x = 0.014$.

6.20 Convex hull for the interstitial (OH group) and substitutional (‘sub(H)’, O replaced by H) defects as a function of chemical potentials $\mu_H$ and $\mu_O$. The polytrope is plotted over the ranges (eV) $\frac{1}{2}\mu_{H_2}^{DFT} + 18 \leq \mu_H \leq 0$, $\frac{1}{2}\mu_{O_2}^{DFT} + 3 \leq \mu_O$. The elemental chemical potential references, based on the DFT energies of H$_2$ and O$_2$ are respectively $\frac{1}{2}\mu_{H_2}^{DFT} = -15.9$ eV, $\frac{1}{2}\mu_{O_2}^{DFT} = -2046$ eV.

6.21 Formation energy of the neutral interstitial defect and the formally charged (jellium) interstitial defect as a function of Fermi energy (referred to the VBM). This is calculated based on equation (69) and assumes linear dependence of the formation energy on $E_F$, which is potentially a deficient assumption given the electronic structure predicting $E_F$ inside the conduction band. Implications of this are discussed in section (6.3.7).

6.22 Nearest neighbour bond lengths for the substitutional H defect (left, figure (6.22a)) and interstitial defect (right, figure 6.22 b). For comparison, the undistorted apical bond length is 2.14303Å, and undistorted equatorial bond length is 2.12706Å.
6.23 **Right:** Bandstructure for bulk interstitial $H^0$ defect in rutile $SnO_2$ with atomic state shell initialisation. DFT calculation performed for one hydrogen atom in a 48 atom cell of rutile $SnO_2$ assuming defect position 'A' defined in section (5.8.1). **Left:** Total and hydrogen projected density of states (arbitrary units).  

6.24 **Right:** Bandstructure for interstitial $H^+_{jel}$ defect in rutile $SnO_2$, modelled as formally charged with a jellium background. Unit cell neutrality was enforced by redistributing the charge of the single hydrogen electron equally amongst all other tin atoms, with all other computational details the same. Density of states (not shown) is identical to that in figure (6.23).  

6.25 **Right:** Bandstructure for $H^0$ substitutional defect. DFT calculation performed for a 48 atom cell of rutile $SnO_2$ with one of the 32 oxygen atoms substituted by one hydrogen. Unit cell neutrality was maintained by redistributing the charge of the two electrons left at the site of the replaced oxygen atom equally amongst all tin atoms. **Left:** Total and hydrogen projected density of states (arbitrary units).  

6.26 Electron density difference map plotted as isosurface (xz plane) for the substitutional defect in the initialised neutral charge state $H^0$. Blue isosurface components correspond to negative density difference, cyan to positive density difference. The right side image shows the immediate environment of the substitutional $H$ site (white circle).  

6.27 Electron density difference map plotted as isosurface (xz plane) for the interstitial defect in the initialised neutral charge state $H^0$, as defined and explained in section (5.8.1). Blue isosurface components are negative density difference and cyan are positive density difference. For sufficiently low isolevel parameter (determining the extent of isosurfaces of equal density normal gradient), projections are limited to the site of $H$ atom and nearest neighbour $O$ indicating the charge transfer leading to the formation of an $OH^-$ group. See figure (6.29) and text. Atom types are given by the legend in figure (6.26).  

6.28 As for figure (6.27) but for a formally charged defect $H^+$, obtained in the same way. Atom types are given by the legend in figure (6.26).  

6.29 Molecular charge difference map between $O^2-$ and $OH^-$, used for comparison to Figure (6.27).  

---

**List of Tables**

1. Summarised performance characteristics of possible PEMFC support materials in addition to Carbon Black (CB). All thermodynamic stability data provided by Pourbaix diagrams [25]. Figures for corrosion stability in terms of ESCA loss and ORR activity correspond to the same experiment in all cases but BET surface area data in some cases from a different source.  

2. Summary of the properties of extrinsic dopants in $SnO_2$.  

12
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Supercells used to examine the scaling of defect formation energy in SnO$_2$, obtained by the respective translation matrices applied to the unit cell.</td>
<td>92</td>
</tr>
<tr>
<td>4</td>
<td>Physical properties of rutile SnO$_2$ as calculated in different DFT codes and functionals, in comparison with experimental values. Experimental band gap measured using optical absorption spectroscopy at $T = 80$K [144]</td>
<td>99</td>
</tr>
<tr>
<td>5</td>
<td>DFT calculations on the Ta-Sn-O system</td>
<td>105</td>
</tr>
<tr>
<td>6</td>
<td>Lengths of all Sn-O and Ta-O bonds throughout 6 layers in $\bar{a}b$ for the GS. Each layer is the direction of decreasing $\bar{c}$. All lengths in Angstrom.</td>
<td>113</td>
</tr>
<tr>
<td>7</td>
<td>Calculated formation energies of interstitial and substitutional H defects in different charge state initialisations (formally neutral, except H$<em>{jel}^+$ which is a formally charged defect modelled with jellium). In these calculations, the reference chemical potentials correspond to the DFT energies of the dimers, $\frac{1}{2}\mu</em>{O_2}^{DFT} = -2046$ eV and $\frac{1}{2}\mu_{H_2}^{DFT} = -15.9$ eV.</td>
<td>130</td>
</tr>
<tr>
<td>8</td>
<td>Possible donor metal elements $M'$ listed by the valence in which they occur in an oxide that is stable at elevated temperature and in an acidic environment</td>
<td>148</td>
</tr>
</tbody>
</table>
1 Author’s Declaration

I declare that this thesis is a presentation of my own original research work and has not to the best of my knowledge previously been submitted to wards the requirements of any academic qualification at this university or another university. I declare also that to my knowledge that the investigation or results presented herein have not been previously published in an academic journal by an independent team of authors. The work presented in this thesis was conducted under the guidance of Dr Denis Kramer at the University of Southampton.

Certain results and figures presented in this thesis appear in an identical or a similar form in journal articles published to this date of which I am a lead or a non-lead author. Among the authors listed in the list of publications, the author with whom I collaborated in order to make my contribution to the published work was again Dr Denis Kramer.


In addition, reproduced in this thesis as figure (4.12) are two graphs from a journal article published by E.E.Haller in Physical Review B on 15th May 1980. I acknowledge this article here:


Finally, I wish to state that the copyright of this thesis lies with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives license.
2 Acknowledgements

Firstly I would like to thank my supervisors Dr Denis Kramer and Prof. Phillippa Reed, for originally presenting me with the opportunity to study for a PhD and their attention paid to my progress over the past four years and invaluable mentoring and guidance. In particular I thank Denis for the training and support he has provided on ab-initio materials science. I also wish to thank the team who made possible the MSSC2012 training school in September 2012 on the CRYSTAL code that I have used extensively within this work, and Roberto Orlando in the CRYSTAL support team.

I thank many individuals in the High Performance Computing (HPC) team at University of Southampton for the technical support they provided, in particular Ivan Wolton and David Baker.

I would also like to thank the students throughout the Engineering Sciences research groups, who I have worked alongside over the last few years, for many discussions and moral support. Finally I would like to thank friends Dominic Chalcraft and Frank Browne and my family for support during my undergraduate degree and my PhD.
3 INTRODUCTION - AIMS AND OBJECTIVES

3.1 Statement of the research problem

This thesis investigates alternative durable catalyst supports for Polymer Electrolyte Fuel Cells (PEMFC) using ab-initio methods. The central research objectives are,

- To identify compounds for the synthesis of supports that meet the requirements of the PEMFC, most fundamentally in terms of durability and electronic conductivity
- To understand mechanisms for electronic conductivity and thermodynamic stability from the ab-initio description of the solid
- Obtain physical quantities pertaining to stability and conductivity and establish the limits of accuracy of these quantities as determined from the methodology used

3.2 Structure of this thesis

The first section (4.1) gives a brief technical introduction to the PEMFC to set the context of the problem of cathode corrosion motivating this research. Following in section (4.2) is a literature review covering performance and degradation of existing standard PEMFC supports and the state of research into replacements with optimised properties.

In section (4.3) is a review of the ab-initio framework for investigating material properties, setting the context for the methodology that is adopted for the calculations based on Density Functional Theory (DFT) and its applications, followed by a review of the types of DFT codes that have been written for the purpose.

Starting section (5.3) is the methodology employed for calculations. This opens by addressing the defects of solid state crystals as a point of departure into modelling the interaction of dopants with host materials. The methodology is general and theoretical before material-specific detailing of the calculations.

The results are likewise outlined system by system with discussions confined to each respective case, before the outlook opening section (7) which considers the respective analyses together. In this part of the thesis, suggestions for extensions and changes to the work are also given.
4 LITERATURE REVIEW

4.1 The Polymer Electrolyte Fuel Cell - working principles

4.1.1 Fuel Cells - General remarks

A fuel cell refers to a device that produces electricity from reacting two fuels in a two stage process via catalysed electrochemical half reactions. The ions are transported through the device between the two electrodes in an electrolyte, driving electron flow in the external circuit. This technical introduction is specific to the PEMFC (Polymer Electrolyte Membrane Fuel Cell), the motivation behind the research proposal to be outlined in the ensuing discussion. This type of cell has already been put to extensive use in high power applications [1].

4.1.2 Polymer Electrolyte Membrane Fuel Cell (PEMFC)

In general, the reactants chosen for the fuels need to be ionised. A reaction involving water is a simple example. Hydrogen and oxygen react in the gas state to form H$_2$O,

$$\text{H}_2 + \frac{1}{2} \text{O}_2(\text{g}) \rightarrow \text{H}_2\text{O}$$  \hspace{1cm} (1)

but only heat is produced as a by-product. By physically separating the two gases into compartments with electrodes connected externally by a metal contact and placed across an electrolyte, electrons can be separated from hydrogen, leaving protons, and flow from the electrode on the hydrogen side to the other with oxygen ions. The reaction of oxygen with protons and electrons, to give H$_2$O, can therefore occur if the protons can be efficiently separated from the electrons to travel across the electrolyte. The oxidation of hydrogen (HOR) takes place at the anode (A) and the reduction of oxygen (ORR) takes place at the cathode (C),

$$\text{A} : 2\text{H}_2 \rightarrow 4\text{H}^+ + 4e^-$$  \hspace{1cm} (2)
$$\text{C} : \text{O}_2 + 4\text{H}^+ + 4e^- \rightarrow 2\text{H}_2\text{O}$$  \hspace{1cm} (3)

At thermodynamic equilibrium, the Gibbs free energy of the reactions given by equations (2) and (3) is zero. Hence equation (3) implies,

$$\Delta G_R = 2G_f(\text{H}_2\text{O}) - [4\mu(\text{H}^+) + 4\mu(e^-) + \mu(\text{O}_2)] = 0$$  \hspace{1cm} (4)

Rearranging, the electrical potential energy $\mu(e^-) = -F \cdot E_0$, where $E_0$ is the equilibrium potential and $F$ is the Faraday constant, can be written in terms of the enthalpy of formation of H$_2$O, $G_f(\text{H}_2\text{O})$, and the chemical potentials of the protons $\mu(\text{H}^+)$ and oxygen $\mu(\text{O}_2)$,

$$-F \cdot E_0 = \frac{1}{2}G_f(\text{H}_2\text{O}) - \frac{1}{4}\mu(\text{O}_2) + \mu(\text{H}^+)$$  \hspace{1cm} (5)
Under standard conditions, \( \mu(\text{H}^+) = \mu(\text{O}_2) = 0 \) and the value of \( G_f(\text{H}_2\text{O}) \) of \( G_f(\text{H}_2\text{O}) = -2.37 \times 10^5 \text{J} \) found from experiment [2] gives the value for the ORR equilibrium potential, 

\[
E_{\text{ORR}} = \frac{-\Delta G_f(\text{H}_2\text{O})}{2F} = 1.229 \text{V} \tag{6}
\]

The ORR is initiated by the adsorption of an oxygen molecule onto the surface of the electrode, which then dissociates to atomic O at a site on the surface (\( \text{O}^* \)) before being protonated first to give \( \text{OH}^* \) and then again to give \( \text{H}_2\text{O} \) [4],

\[
\frac{1}{2} \text{O}_2 + * \rightarrow \text{O}^* \tag{7}
\]

\[
\text{O}^* + \text{H}^+ \rightarrow \text{HO}^* \tag{8}
\]

\[
\text{HO}^* + \text{H}^+ \rightarrow \text{H}_2\text{O} + * \tag{9}
\]

where the notation * denotes a site on the catalyst surface. At the equilibrium potential, the adsorbed (\( \text{O}^* \)) is the lowest energy state, with the latter two processes in the mechanism being endothermic [4]. Therefore the overall energy difference between the adsorption of (\( \text{O}^* \)) and the final process in (9) acts as the rate determining step, i.e. has the largest influence on the reaction rate overall. The ORR rate thus depends on the binding energy of adsorbed (\( \text{O}^* \)). For the case of Pt, the latter two processes in (9) have approximately the same energy barrier [4]. For other metals such as Ni, OH is so strongly bound that it cannot dissociate following the proton transfer, inhibiting the reaction [5]. On the other hand, for metals like Au the oxygen will bind too weakly to the surface. The use of Pt as catalyst, in the case of the ORR, therefore represents the best compromise between over and under-binding of the adsorbed molecules in agreement with Sabatier’s principle [6]. However, there are a number of problems with Pt based catalysts in the PEFC that to date have limited its viability in automotive applications as an alternative to petroleum vehicles. These are discussed in the section (4.2).

### 4.1.3 Charge Transport and the Polymer Electrolyte Membrane (PEM)

The multistep ORR reaction described in section (4.1.2) requires the protons to flow across the device in one direction, from the hydrogen to the oxygen side, which necessitates the use of the Polymer Exchange Membrane (PEM) as a dividing electrolyte between anode and cathode. It acts effectively as an ionic 'valve', allowing transfer of \( \text{H}^+ \) from A to C. The typical membrane structure consists of long fluropolymer chains with shorter side chains of alternate O atoms and pairs of CF\(_2\) molecules, at the end of which is attached a sulfonic acid group HSO\(_3\). The adjacent fluropolymer chains are thus either side of a channel dominated by the presence of the side chains, see figure (4.1). The fluropolymer "backbone" is hydrophobic due to the high electronegativity of flourine. In the channel regions, however, the acid group is hydrophilic towards the aqueous \( \text{H}^+ \) [16], on account of the mechanism of proton "hopping" which facilitates their transport through the membrane. This operates as follows. The protons are able to move
from one water molecule to the next within the electrolyte on account of their attraction to the molecular dipole of $\text{H}_2\text{O}$ that is directed from between the two hydrogen atoms to the oxygen anion. This forms an $\text{H}_3\text{O}^+$ molecule, and with rearrangement of the charge density, a proton then detaches from one of the $\text{H}$ sites alternative to the one initially protonated and moves to the neighbouring water molecule, which assumes the role of the $\text{H}_3\text{O}^+$ [2]. The coexistence of the hydrated channels and the hydrophobic "backbone" fluropolymer connect the available water containing cavities and hence favour high permeability for $\text{H}_2\text{O}$ [17].

Because charge transport across the membrane takes place through the migration of protons between neighbouring water molecules, the whole of the membrane needs to remain at or close to full water saturation [15]. The proton transfer mechanism, involving the formation of the $\text{H}_3\text{O}^+$ species, gives an electro-osmotic 'drag' of $\text{H}_2\text{O}$ from anode to cathode [14], and hence can result in dehydration of the anode side. To avoid this, the membranes are designed to have a thickness of between 15 – 20$\mu$m to enable back diffusion of water from the cathode side to the anode side when an excess arises [3]. On the other hand, the humidity inside the membrane relative to the exhaust stream should not exceed 100% [9]. Otherwise, the porous layers of the PEM will contain liquid water [10] that can lead to flooding of the cathode [11]- [13], and deleterious expansion, or swelling, of the membrane pores, such that the protons are no longer in close enough contact to be transported by hopping. In the ideal case, membrane conductivity increases with the degree of hydration, when the interactions of the water with the sulphonate acid groups dominate over the hydrophobic nature of the fluropolymer, giving the membrane a resistance to swelling. Past the point of over-saturation, however, the chemical potential of the water in the vapour stream becomes sufficient to overcome this resistance and the average distance between $\text{SO}_3^-$ groups abruptly increases, which hinders the proton-hopping between two points and greatly reduces membrane conductivity [40].

4.1.4 Electrodes and their structure

The electrode design must serve the purpose of isolating ions from electrons as noted in section (4.1.2) whilst maximising electrochemical reaction rates, and the material used must have properties commensurate with this. Carbon has high electrical conductivity in the range of $10^3$ – $10^8$ Sm$^{-1}$, is highly porous and is available in many morphologies [18], which makes it feasible to fabricate an electrode that achieves a three-phase boundary at which the catalyst, electrode pore, conducting electrode and polymer membrane are all in contact [20]. This ensures that with diffusion of $\text{H}_2$ molecules, the $\text{H}^+$ and $e^-$ resulting from the hydrogen oxidation travel respectively into the electrolyte and electrode as desired, illustrated schematically in figure (4.2). In practice, thin membrane polymer layers may cover catalyst particles, diminishing the contact between the catalyst surface and the hydrogen in the electrode pore. However, such obscured surfaces can still contribute to the reaction through the transfer of mobile hydrogen or hydroxyl species [21]. In the electrically conducting phase, the Pt catalyst on which $\text{O}_2$ and $\text{H}_2$ are adsorbed respectively at the cathode and anode initiates the electrochemical reactions.
4.1 The Polymer Electrolyte Fuel Cell - working principles

Figure 4.1: Schematic illustration of the Polymer Exchange Membrane of an acidic-mode PEFC. Above-right image shows the basic structure with the hydrophobic fluoro carbon backbone and the sulphonic acid group, taken from reference [19].

The catalyst is supported in nanoparticle form on the carbon which may be used in a number of different allotropes that are discussed in turn in section (4.2.2).

The remainder of the electrode consists of two components, the Gas Diffusion Layer (GDL) and the flow field plates (FFP), figure (4.3). \( \text{O}_2 \) and \( \text{H}_2 \) flow through inscribed channels in the plates from the gas inlet to the main body of the electrode. The diffusion layer is inserted between carbon electrode and flow field plate so that the full area exposed to the channels is utilised for electrochemical reactions, rather than the carbon sitting flat against the plate.

4.1.5 Electrochemical behaviour

To understand the charge transfer process quantitatively it is necessary to introduce the fundamental equations of electrochemistry. At a given cell potential, the oxidation and reduction reactions of equations (3) and (HOR) are initiated by the excitation of electrons from the reactant or product species, the charge transfer at the atomistic level. The Boltzmann factors associated with the probabilities of \( e^- \) excitation depend on the activation barrier between the reactant and product states, whose shape determines the ease with which excited electrons will make the transition between the two states in an arbitrary single electron transfer process. The
4.1 The Polymer Electrolyte Fuel Cell - working principles

Matthew Worsdale

Figure 4.2: Schematic illustration of the three phase boundary (at the anode) between the three important regions, clockwise from top: Pore of carbon electrode containing hydrogen gas from the flow field plates/gas diffusion assembly (figure (4.3)), the aqueous polymer electrolyte phase, consisting of electrolyte phases containing aqueous H\(^+\), and the solid phase with carbon particles and Pt nanoparticles.

Current density \(i\) is described by the Butler-Volmer equation, for a single electron process [22]

\[
i = i_0 \left( \exp \left( \frac{\gamma F}{RT} \eta \right) - \exp \left( \frac{-(1 - \gamma)F}{RT} \eta \right) \right)
\]

(11)

where \(z\) is the number of electrons, \(R\) is the gas constant, \(T\) is temperature and \(\eta = (E_{\text{cell}} - E_0)\) is the overpotential, or difference between the observed cell potential and the thermodynamic equilibrium potential for the ORR, and \(\beta\) is the symmetry factor describing the standard free energy barrier for the reaction. The exchange current density \(i_0\) is the current at equilibrium, when anodic and cathodic currents are equal, and reflects electron transfer between the electrolyte and electrode. The Butler-Volmer equation only describes charge transport at the electrode surface, and hence its application to determine reaction kinetics is only valid if the reaction rate on the electrode surface is slower than the diffusion of reactants to the surface [23]. If this is not the case, the reaction rate is hindered to a degree defined by the mass transport overpotential \(\eta_{\text{mt}}\). However, for the ORR in the PEFC \(\eta_{\text{mt}}\) only becomes significant at high overpotentials (high current densities) due to slow diffusive transport and liquid phase water blockage in the GDL [24], so that equation (11) is a good approximation to determine the reaction rate throughout most of the range of cell operating potential. The variation of the equilibrium cell potential \(E_0\) with changes in the reactant and product concentrations [\(\text{ox}\)] and [\(\text{red}\)] as a function of the temperature and inlet gas pressures is described by the Nernst equation [35]

\[
-E_0 = -E_{\text{ORR}} + \frac{RT}{zF} \ln \frac{[\text{ox}]}{[\text{red}]}
\]

(12)

where it is often convenient to work with hydrogen and oxygen gas partial pressures \(p_{H_2}, p_{O_2}, p_{H_2O}\) expressed as activities \(a_i = \frac{p_i}{p_0}\) with \(p_0 = 0.1\)MPa. In this form the Nernst equa-
Figure 4.3: The Gas Diffusion Layer (GDL) and flow field plates (FFP) that aid the diffusion of hydrogen and oxygen gases to the anode and cathode, respectively.
where it is assumed that the activity of $O_2$, $H_2$ and water vapour dissolved in the aqueous electrolyte are proportional to their respective partial pressures in the gas phase, in situations at or close to the thermodynamic equilibrium. Equation (13) describes the evolution with thermodynamic state variables temperature $T$ and activities $a_{H_2}$, $a_{O_2}$, $a_{H_2O}$ of the equilibrium potential $E_0$, assuming no loss mechanisms. A reduction in the working cell potential $E_{cell}$ from the equilibrium value $E_0$ determined by equation (13) occurs with increasing current density in the cell as a function of ORR rate, corresponding to the overpotential $\eta_K$ of this reaction, which leads to a voltage drop across the cell. There is also an overpotential for the HOR, although negligibly small even at the highest current densities [36]. Increasing cell current density gives additional losses from the electrical resistance to electrons of the electrode $R_{elec}$ and the resistance to protons of the electrode $R_{H^+ elec}$ and of the PEM $R_{H^+ mem}$. The ORR overpotential increases to account for approximately 70% of the overall voltage loss at the highest current densities with the remaining losses from electrical resistance being dominated by ohmic resistances [3]. Therefore the contributions to the voltage loss can be represented by the following expression,

$$E_{cell} = E_0 - \eta_K - \eta_M - \eta_{mt}$$

$$E_{cell} = E_0 - \eta_K - iR_{elec} - i(R_{H^+ elec} + R_{H^+ mem}) - \eta_{mt}$$

where $\eta_M$ is the total ohmic overpotential, and any non-ohmic terms in the resistances of the membrane and the electrode to proton conductivity are neglected, $\eta_{elec} \approx iR_{elec}$, $\eta_{mem} \approx iR_{mem}$. The cell potential losses are illustrated in figure (4.4). The principal losses of the PEM are represented by the current-potential curves shown by figure (4.4).

### 4.2 Support materials for Pt supports

#### 4.2.1 Carbon Corrosion

While the variety of available forms of carbon and its conductivity are useful properties for fabrication of electrodes, carbon is only thermodynamically stable in within a small range of cell potential up to $E_{COR} = 0.207 E_{SHE}$ (defined relative to the potential of HOR, equation (2)) in acidic environments (pH=0), as seen from the Pourbaix diagram [25]. Carbon will therefore undergo corrosion,

$$C + H_2O \rightarrow CO_2 + 4H^+ + 4e^-$$

taking place primarily at the cathode which is at a high oxidation potential with respect to the anode. The progress of reaction (15) depends both on the cell potential and the kinetics of the reaction. The simplified two stage mechanism involves the formation of surface oxides that are
then oxidised to CO$_2$ with Pt acting as a catalyst for the second process [3], with more detailed steps given in c.f references [26]-[30]

$$C + H_2O \rightarrow CO_{surf} + 2H^+ + 2e^- \quad E > 0.3V \ vs \ SHE$$

$$CO_{surf} + H_2O^{Pt} \rightarrow CO_2 + 2H^+ + 2e^- \quad E = 0.8V \ vs \ SHE \quad (16)$$

The surface groups are observed at potentials higher than 0.8V for a temperature of 65°C, and for potentials higher than 1.0V at room temperature [37], which means that the reaction is kinetically hindered at lower potentials. Since 65°C is approaching the maximum working temperature of the PEFC, the limit of kinetic stability is judged to be a cell potential of approximately $E_{KIN} = 0.8V$ [38]. Working cell potentials are in the range of 0.7-0.9V for the $\approx 300000$ load cycles the system should be expected to deliver over its working life in high power (e.g., automotive) applications [39]. In addition, the potential temporarily surges to values as high as 1.5V under starting and stopping conditions when there is an influx of H$_2$ and air through the anode [45], [99]. These potential cycles, over the lifetime of the PEFC, lead to progressive corrosion of the carbon support [38].

The main detrimental consequences of carbon corrosion can be summarised as the loss in the electrochemically active surface area (ECSA), and detachment and loss of the expensive Pt
4.2 Support materials for Pt supports

Matthew Worsdale

catalyst nano-particles [45], [38]. As discussed in sections (4.2.3) and (4.2.4), this serves to place a limit on the working lifetime of current PEMFCs.

4.2.2 State of research in the field - overview

Materials are evaluated in the following review based on the demonstrated corrosion with fuel cell operating conditions (observed loss of electrochemically active surface area (ESCA)) and ORR activity (in terms of observed cell current densities \( i \)). As discussed in the previous section, corrosion of carbonaceous supports depends on the kinetics of the carbon oxidation reaction (equation (15)) as well as the temperature and pH dependent thermodynamic equilibrium potential. The kinetics are found to be dependent on the cell potential cycling conditions through both the maximum and minimum potentials [52], and the rate of change of potential between these limits [54]. Such conditions are assessed by the method of accelerated durability tests (ADT), in which an electrode is prepared by spraying with a catalyst and placed in contact with an electrolyte to mimic the PEFC electrode-membrane interface [46]. The loss of ESCA due to corrosion is determined by cyclic voltammetry tests (CV), in which the potential applied between the working electrode and a counter electrode is varied between two fixed limits linearly as a function of time. This cycling is maintained over a number of timescales, to calculate the charge transfer corresponding to the adsorption and desorption of hydrogen in the HOR reaction, and the change in ESCA for the process, by calculating the BET surface area. The BET theory determines electrode area for multilayer adsorption/desorption processes, where the surface area depends on the coverage by the \( \text{H}_2 \) monolayers adsorbed, in addition to the electrode geometric surface area [47].

This review starts with a general appraisal of the problems that are the most prevalent in any PEFC, before moving on to consider these and the other most problematic issues in the particular case of carbon black (CB), the support most widely adopted since the inception of the PEFC, and then carbon and non-carbon based alternative supports.

4.2.3 General degradation mechanisms for carbon based supports

Primarily, the electrochemical corrosion of carbon, reaction (15), causes not only a loss in the electrochemical surface area (ECSA) over time, but also aggravates the tendency of the Pt nanoparticles to agglomerate into larger clusters from their initial dispersion over the carbon support [48], due to a high specific surface energy [49], leading ultimately to detachment from the electrode surface. Under potentials closer to the maxima that can be experienced, during the current surges of at start-up and shut-down, the Pt itself can be oxidised which can allow the formation of other chemical species such as oxides of Pt on the electrode surface that reduces the number of available reaction sites for (2) and (3) [44], [48]. Also important for long term durability, in addition to electrochemical corrosion of the electrode surface under oxidation potentials, is the resilience of the carbon to the 80°C+ temperature over periods of
Figure 4.5: TEM image of Carbon Black (CB) based catalysts, before corrosion testing left: Carbon Black without graphitisation; Right: Carbon Black with graphitisation. Taken from reference [59].

sustained use to both thermally driven direct oxidation (gasification) [50] and enhanced kinetics of electrochemical corrosion [51]. This is particularly important when considering an allotrope of carbon that can decompose thermally, as will be discussed in the case of carbon nanotubes and nanofibres below.

Attempts to overcome the above problems have focused on both carbon based and non-carbon based replacements for the problematic existing carbon black (CB).

4.2.4 Existing support- Carbon Black (CB)

Carbon Black is a form of carbon having varying degrees of order, between amorphous and crystalline (graphite content). Figure (4.5) shows a tunnelling electron microscopy (TEM) image of a carbon black catalyst supporting Pt catalyst nanoparticles on its surface. Existing carbon black (CB) based catalyst supports have the advantage of high surface area and porosity which furnishes a large number of reaction sites for the Pt catalysed PEMFC reactions ((2) and (3)). However, it has been shown that under fuel cell conditions, imitated by potential cycling, there is extensive surface oxidation on the CB surface involving the mechanism [46],

\[
C=O + e^- + H^+ \rightleftharpoons C-OH
\]

(17)

involving the formation of the hydroxyl radical. The surface charge density for this process was found to increase almost tenfold over the first 4 hours of potential cycling over the range of 0-1.2V, a factor accelerating the electrochemical corrosion of the electrode through rapid loss of ESCA [52]. This is, however, dependent on the specific CB used, with losses of 40% or 20% in two different supports after 10,000 potential cycles [53]. For the best performing CB, the rate of loss declined to around 5 % of the original ESCA being lost over 600 cycles, which is therefore of no use for a working PEMFC which needs to be designed for approximately 300,000 potential cycles between 0.7V and 0.9V over a lifetime in a working automotive application [54].

26
Assuming that the corrosion cannot be mitigated through treatment or otherwise of the CB, taking this as an approximate best case degradation scenario for the PEMFC, the durability could potentially be increased by a higher starting ESCA, achieved through increasing the loading of Pt catalyst and/or the surface area of the carbon that is electrochemically active for the desired cell reactions. The first approach is unviable not only because of the limiting factor of Pt agglomeration but also because Pt itself is understood to act as catalyst for carbon corrosion, equation (15) [55]. Although theoretical studies propose reaction sites within carbon nanostructures, formed from lattice defects, that favour the ORR reaction [56], the thermodynamic stability of these defects or of the nanostructures containing them is not given. This motivates exploring other carbon allotropes as candidates, so that the desirable conductivity of carbon may be kept within a system that is also inherently more corrosion resistant than CB.

4.2.5 Carbon Allotropes

Carbon exists in a number of forms with different crystal structures which may offer superior performance as catalyst supports.

**Graphitisation of Carbon Black (GCB).** Graphitisation leads to an increase in molecular ordering and degree of interconnection of carbon sheets, with a resultant increase in crystallinity, but with retention of the original mesostructure. After 30,000 potential cycles between -0.21 and 0.8V, about 85% of the original ESCA remained for GCB compared to 60% for the untreated CB, for a temperature of 25°C [58], representing a substantial improvement. However, this is not an entirely fair comparison with the study reported above [52] for carbon black, since the potential range used in the cycling test in that case was 0.6-1.2V. Perhaps a fairer comparison is provided by a second investigation where the measurements were taken under a temperature of 65°C. In this case, the current density (at 0.85V) was initially lower than that for ordinary CB, but after a durability test of 10,000 cycles, GCB had retained around 50% of the original, compared to around 25% for CB making its long term electrochemical performance superior, and similarly for the ESCA. This study proposed that the distribution of the Pt nanoparticles between different sites is an important factor in the enhanced corrosion resistance [59]. Although CB offers a high surface area, which increases the maximum ESCA utilizable by Pt, much of it is found in high density deposits at defect sites of CB, whereas in Pt/GCB, the deposition is more even over the graphite layers. This maximises the utilization of the electrochemical surface area by Pt for a given loading which is an immediate advantage over CB where the accumulation of the initial catalyst coverage at particular sites is due to Pt agglomeration over time. Functionalization of a graphitised carbon can also be undertaken with groups such as phenyls being found to improve both the catalytic activity for the ORR, attributed again to a higher degree of uniformity in dispersion of Pt, and the reduction in migration of nanoparticles and hence the corrosion resistance, with a reduction in the loss of ECSA compared to the unfunctionalized GC. [60].

**Graphene.** This material is formed of a single sheet of hexagonally ordered carbon atoms, which has been explored as a potential support material since its discovery [61]. For graphene
4.2 Support materials for Pt supports

Figure 4.6: TEM image of graphene sheet supporting Pt catalyst nanoparticles on surface. Taken from reference [62].

nanosheets produced by sonic treatment, ESCA was as much as 45% greater than ordinary Carbon Black (CB), and the decrease in the current density and ESCA under sustained potential cycling was also reduced compared to CB, by around 50% and 35%, respectively, if the graphene was functionalized in addition using citric acid and ultrasonic treatment followed by heating at 300°C [62]. Figure (4.6) shows a TEM image of a graphene sheet supporting Pt nanoparticles. More comprehensive studies into both electrochemical performance and corrosion resistance have been hard to come by.

**Carbon Nanotubes (CNT).** Nanotubes, which are 2D nanostructures, have been developed for PEMFC applications over the last couple of decades. The surface of a pristine CNT is chemically inert and therefore functionalization is essential to enable the attachment of Pt nanoparticles, usually achieved using treatment with strong sulphuric or nitric acid [63]. The loss of ORR current of a multi-walled CNT supporting Pt catalyst is reported to be 5.8% after 1000 hours held at 0.9V while the loss of ESCA is 37% after 168 hours [64]. This study also reported reduced rates of Pt agglomeration compared to CB, and a lesser increase in the corrosion current when Pt is used as a catalyst compared to the case of CB. The desirable preservation of the current compared to ESCA could, therefore, reflect the better utilization of Pt than in the case of CB. Increasing loading of Pt is again, however, found to be responsible for the enhanced agglomeration with resultant loss of ESCA and catalytic activity [65], though the effect was found to be mitigated by a technique of Low Temperature Hydrogen Bubbling (LHTB), whereby molecular hydrogen dissociates promoting dehydration of the functional groups on the surface of the nanotube leaving behind oxygen that can readily bond to both Pt and C, which resulted in an order of magnitude reduction in the loss of active surface area and activity [66].

**Carbon Nanofibres (CNF).** The main way that CNF are distinct from CNT is their near total or total lack of a hollow cavity. Limited work on CNF as a support has been reported, and is largely aimed at establishing if the benchmark electrochemical performance is adequate to be considered in the PEMFC. For both CNT and CNF to be viable durable supports, thermal decomposition of the nanostructure must not occur under long term operating conditions. A preliminary indication is that CNF has a higher decomposition temperature in addition to superior electrochemical corrosion resistance [76], attributed to its high graphitisation and crystallinity. The heat treatment involved in the production of CNF requires temperatures of up to
2800°C, which might however negate this stability benefit in terms of the cost of production. Figure (4.7) shows TEM images of CNT and CNF supporting Pt nanoparticles.

4.2.6 Posphides and Oxyposphides

**Paladium Posphide.** While no work on Platinum Posphides can be found, interest in Palladium Posphides PdP$_x$ is motivated by the higher global availability and lower cost of Pd compared to Pt. A recent investigation by Kucernak et al. [67] synthesised three catalysts retaining carbon as support, Pd/C, Pd$_5$P$_2$/C and PdP$_2$/C. The latter gives the best retention of ESCA, comparable to that recorded with Pt/C (after 10 000 cycles) and a 20% improvement on Pd/C (after 5 000 cycles). The sample with the lowest proportion of Pd also showed the most promising ability to withstand thermal oxidation under TGA with around 15% mass loss after 60 minutes of heating at a rate of 10°C/min, and, the highest onset potential for the ORR considering both the geometric and mass-transported corrected specific surface areas. This catalyst was tested under fuel cell conditions in both directions, using Pt/C as the corresponding electrode in both cases. Current density is of the order of 10mA cm$^{-2}$, though the resilience of the ORR activity after a large number of potential cycles specifically is not given. The authors consider the potential of this material to be limited by both surface phosphorous atoms predisposing the formation of H$_2$O$_2$ instead of H$_2$O and anionic sulphonic acid groups. Notwithstanding, it would appear that gains in electrode durability could be a decisive benefit in the use of PdP$_x$ while preserving the relatively high intrinsic conductivity of Pt group metals. However, single crystal work [68] shows that Pd is less active than Pt at the same crystallographic site, and the Kucernak study finds current densities that are generally an order of magnitude lower than for Pt. Furthermore, the loss of ESCA in absolute terms is still at approximately 30% after 10 000 cycles, which is far from the best found. Smaller corrosion rates are found for metal oxides without carbon, as discussed below.
Cobalt Oxyphosphide. This system has been studied [69] by synthesising a number of compositions Co$_2$P$_2$O$_7$/C, CoP$_2$O$_6$/C, Co$_3$P$_2$O$_9$/C and CoO/C. None of these compounds provide particularly favourable current densities, with values typically an order of magnitude less than the PdP$_x$ and Pt based systems throughout the scan potential range. The onset potential for ORR is maximised at 0.69V for Co$_4$P$_2$O$_9$/C. In terms of stability, however, assessed for the most active compound Co$_4$P$_2$O$_9$/C, performance would appear very attractive with hardly any change in potential (with a constant current density of 100mA cm$^{-2}$) after 50 hours. Published figures for the corresponding change in ESCA would increase confidence in this conclusion, particularly for cycling over a potential range, since this is the standard method for testing of electrochemical stability. There is also no data for the conductivity or surface area or a meaningful source for comparison, as synthesis of these compounds is seldom reported. The authors speculate that control over the functionalization brought about by P and O could allow control in turn of the oxygen adsorption on Co which needs to be optimised for the given surface, and even improve charge transfer at the three phase boundary as a whole. Such optimisation require much costly experimental trials, though the possibilities through varying the composition and surface properties probably make it worth keeping these materials within consideration.

4.2.7 Nitrides

Titanium Nitride. Another material with metallic conductivity and a high benchmark stability [70] under oxidising conditions [71]-[73], the electrochemical oxidation of TiN under fuel cell conditions proceeds with changes in the surface chemistry due to there being alternate thermodynamically favourable reactions depending on the cell potential [70], [74]. In the range 0.5-0.9V, ionisation and also oxidation of Ti and production of oxide/oxynitride films self-passivates. Further oxidation in the range 1.0-1.5V produces hydroxides that are more soluble and diminish the passivating effect as a result. The simultaneous production of N$_2$ in both regimes and its diffusion through the surface layers further modifies the electrochemistry by screening the titanium ions. In a study under PEMFC operating conditions [75], the low oxidation current densities that are found (in the range of 0 - 0.1mA cm$^{-2}$), therefore, are initially due to the inert bulk nitride and with the progress of the aforementioned processes, increase within only a small range to the peak value. Not surprisingly, the rate of corrosion over a shorter timescales up to 24 hours as a consequence of these processes (related approximately to the rate of current density decrease) is low. However, with the different temperature dependences of the coexisting processes, it happens that at 80°C the dissolution of the oxynitride layer takes place in the early stages, which leads to greater formation of the passivating hydroxides at higher potentials. The TiN nanoparticles are thus unable to maintain adequate transfer of electrons from Pt, and hence the majority of the ESCA is lost effectively immediately at that temperature. The rate of loss is still unsatisfactory, though, at lower temperatures. It should be emphasised that the long term ESCA loss is not corrosion but electrochemical inactivity due to surface chemistry. Perhaps the possibility exists to produce surfaces of the oxynitride where the formation of hydroxides and other passivating compounds is minimised, raising the
4.2 Support materials for Pt supports

prospect of high durability due to the resistant TiN. The current density rate would also need to be improved, and there is no well defined indicative rate of that offered by bulk TiN.

4.2.8 Carbides

Tungsten Carbides. Interest in Tungsten Carbide (WC) is motivated by evidence that it promotes the ORR reaction [77]. One investigation reports a loss in ORR current density of around 50% between 0.6V and 0.8V after only 100 potential cycles over this range, together with about a 25% loss in ESCA [78]. This is further supported by a recent study that detected oxides of W on the basis of X-ray Photoelectronic Spectroscopy to establish the electron states that were indicative of W-O bond, and concluded that the material had at least partly oxidised for an oxidation potential of \(0.8\) V [79]. There is therefore reason to expect that oxidation of carbides is a routine occurrence at commonplace cell potentials.

4.2.9 Non-carbon based oxides.

A number of oxides of rare earth and transition metals have been used in place of carbon based materials. One can expect a good benchmark stability if the metal is in its highest oxidation state, so the metal element itself will not be able to donate any more electrons in an oxidation reaction of the material as a whole.

Tungsten Oxide \(WO_x\) This is an n-type semiconductor with a band gap in the range of \(2.6 - 2.8\) eV, which can form a wide variety of oxides due to the ability to adopt all oxidation states up to +6, which in general coexist as samples are nonstoichiometric. The material is understood to have an intrinsically high corrosion resistance, motivating trials both as a support in its own right, and an integrated support material used in conjunction with carbon, in both cases generally employing Pt to enhance performance. Unlike carbon based systems, it has not been proposed that Pt catalyses or otherwise contributes to the corrosion of \(WO_x\), and may instead promote durability in the case where carbon is included, since the oxidation potential of CO is shifted towards lower values, alleviating CO poisoning of the Pt catalyst [80]. The methods used to synthesise Pt–\(WO_x\) are of importance to the outcomes of electrochemical activity and corrosion resistance. As an example, in the case of Pt–\(WO_x\) as a stand alone support, in which the existing chemical species was revealed to be \(WO_3\) by XRD analysis, under accelerated oxidation cycling, involving 100 cycles at 80°C and with potential steps between +0.6 and +1.8V, the oxidation activity remained relatively stable whereas that for CB decreased after just 10 cycles [81]. Given that the potential cycling was between 0.6V and 1.4V, with a 20s excursion to 1.8V, this could be taken as quite promising electrochemical durability of the oxide. In addition almost no weight was lost under an applied temperature up to 1000°C, which demonstrates thermal stability comparing favourably to CNT and CNF which had decomposition temperatures of 500°C and 600°C. However, the overall current density for the ORR remained lower than for CB, chiefly because \(WO_x\) is less conductive than carbon but also attributed to the use...
chloroplatinic acid in the synthesis which reduces to Pt, even in the absence of WOₓ, resulting in a greater proportion of Pt clusters that are not supported properly on the oxide.

Using Pt–WOₓ in conjunction with carbon, tungsten oxide nanostructures can be employed. One investigation used carbon coated nanowires grown directly onto carbon paper [82]. The ORR current density was of the same order as that reported in the first study, that used Pt nanoparticles, but after 50 potential cycles from 0.0V to 1.2V the maximum current had decreased by a factor of 0.06, which is less promising as a durability indicator. Assessments of longer term electrochemical corrosion using a larger number of potential cycles were not reported in either of the studies discussed. However, according to Pourbaix diagrams, passivation makes WOₓ resistant to corrosion under the oxidation potentials of the PEMFC in an acid [25].

**Titanium Dioxide TiO₂.** Stochiometric titanium oxides demonstrate high electrochemical stability even in acidic environments [83] and are relatively low cost materials [84]. However, TiO₂ is a wide band gap semiconductor. Electronic conductivity could be achieved if Ti³⁺ ions were introduced, either by oxygen deficiency or by n-type donors incorporated into the sample, but it has been reported that under fuel cell operating conditions, the stoichiometry is regained and a resistive dioxide layer is formed at the three phase interface [85], [86]. Confined to systems with Pt nanoparticles, the maximum activity for the ORR is impressive, about the same levels as reported for WOₓ, but unfortunately this activity is lost exponentially as the potential increases from 0.75V to 0.95V ultimately by two orders of magnitude. In addition, the retainment of ESCA after 2500 cycles was only around 50 percent when the support was cycled under potentials from 0.0 to 1.4V [87]. Although this was not decreasing as fast with the number of cycles in the end, this and the rapid loss of activity at high cell potentials are both indications of acceleration of corrosion under the harshest conditions of the PEFC and is an important reminder that metal oxide must be stable under both pH = 0 and oxidation potentials up to 1.5V. According to Pourbaix diagrams, passivation of hydrated TiO₂ gives way to corrosion above 1.3V in a pure acid [25], while tungsten oxides will not corrode below pH = 4, which may be the reason for the particularly high reported corrosion of TiO₂ with Pt, especially given that much of the loss of ESCA occured over the first 200 potential cycles. The results can be compared to a similar study in which the loading of Pt was 40 percent by weight, rather than the 60 percent used in the first study. The retainment of ESCA after 1000 cycles was 75.6%, but the current density for the ORR was only around 10% of what it was for the first study. [88]. After 1000 cycles, the ESCA loss in the first case was in the range of 60-70%, thus it appears that the electrochemical performance can be greatly improved using a higher Pt loading, but at the expense of sacrificing some corrosion stability.

**Tin Dioxide SnO₂.** Another electron deficient semiconductor (see section (4.6) for the meaning of this term), SnO₂ in the rutile structure is stable in an acid up to around 1.3V, according to the Pourbaix diagram, similar to tungsten oxide. The band gap is around 3.5 eV [98], placing the material in the class of wide band gap semiconductors, nevertheless, the intrinsic conductivity (when negligible dopants or impurities are present) is higher than other materials in this category [158], and possible reasons for this are discussed in section (4.5.2). The material also acts as a transparent conducting oxide (TCO) (if fabricated in films of thicknesses of 10⁻¹ µm), so that the primary mechanism for the electrical conductivity cannot be optical excitation
of electrons across the band gap except for the case of absorption and re-emission of light of the same frequencies. In an investigation into SnO$_2$ nanoclusters with Pt, when the sample was held at a potential of 1.6V for 10 hours at 60°C, the loss of ESCA was only 11.3% [89]. In an additional long term test of durability assessed instead by potential cycles from 0.6V to 1.2V, the loss was 23%, which suggests interestingly that change of potential over the range of the onset of the ORR is more damaging to the electrode than the high potential range, where the stability of the bulk material would be expected to be compromised by the direct reduction of the Sn in support. However, there is negligible change in the current density in the higher potential range where the Pt oxidation should be observed, as the number of potential cycles increases, which the authors attribute to the strong interaction between Pt and SnO$_2$ as well as the low conductivity of SnO$_2$ versus Pt. These observations suggest that the low conductivity of SnO$_2$ does not prevent its corrosion in the range 0.6V to 1.2V, which might be linked more strongly to the load placed on the cell during operation with which the rate of change of potential corresponding to the scan rate in the experiments is associated. The long term durability is promising nonetheless with 77% of ECSA still in place after 12000 cycles. In addition to the Pourbaix diagram, which reports stability for metal oxide systems at 25°C, it is further suggested that SnO$_2$ remains stable at 80°C [90], [91], which puts this strong result for corrosion stability on a firm basis. However, if SnO$_2$ is to find use as the PEMFC support, the electrochemical activity towards the ORR also needs to improve as the current densities are only in the range of 0 – 1mA cm$^{-2}$. Hence, more conductive samples of the material are sought, that improve the current density for the ORR, without accelerating the reduction of Sn in the oxide.

Conclusions The properties of the investigated materials are summarised in table (1). In general, the conductivity is superior for the carbon based supports. However, the metal oxides SnO$_2$ and WO$_3$ have better corrosion stability in terms of the retention of ESCA than the other materials, with the exceptions of GCB and graphene which were tested at a low temperature as seen in table (1). TiO$_2$ appeared to suffer most of its degradation at high oxidation potentials, possibly due to the loss of a limiting passivation effect, which is retained in WO$_{X}$, though this compound quickly lost ORR activity. SnO$_2$ shows the most promising corrosion resistance overall. Looking at ESCA retention and ORR activity retention together, the former showed a loss of 23% after 1800 potential cycles, then after seven times as many cycles the loss of the maximum ORR activity was 45% suggesting that the rate of overall degradation had substantially reduced over the 12,000 total cycles. The important point to take from this investigation is that metal oxides can demonstrate high corrosion resistance under the most extreme (most strongly oxidising) PEMFC conditions, while offering the potential to enhance their electrochemical properties by engineering their properties through doping and other means. A first principles approach can be employed to reduce the complexity of finding good candidates within the space of compounds based on SnO$_2$ and other metal oxides.
4.2 Support materials for Pt supports

<table>
<thead>
<tr>
<th>Candidate support material type</th>
<th>Thermodynamic stability range [25] (Range $V_{H2E}$, pH = 0)</th>
<th>Surface Area (BET) ($m^2/g$)</th>
<th>Conductivity (without Pt) ($Sm^{-1}$), 358K</th>
<th>Corrosion stability (% ESCA loss)</th>
<th>ORR activity (mA cm$^{-2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB</td>
<td>0-0.2</td>
<td>56.9 [93]</td>
<td>5.58 x 10$^3$ [93]</td>
<td>20.6% 30,000 cycles [53] 20°C 0.6-1.2V 20% Pt</td>
<td>22 (initial), 50 [54] 25 (10,000 cycles)</td>
</tr>
<tr>
<td>Graphitised CB (Mesoporous)</td>
<td>0-0.2</td>
<td>300 [93]</td>
<td>10$^3$ [93]</td>
<td>14%, 28,800 cycles 25°C -0.2V-1.8V [58] 18% Pt</td>
<td>100 (initial), 50 [58] 28,800 cycles, 25°C (at 0.9V)</td>
</tr>
<tr>
<td>CNT</td>
<td>0-0.2</td>
<td>272 [93]</td>
<td>10$^3$ [93]</td>
<td>37% after 168 hrs, 0.9V 25°C, 30% Pt [64] 25°C, 30% Pt [64]</td>
<td>0.01, 5.8% loss after 168 hrs, 0.9V 25°C, 30% Pt [64]</td>
</tr>
<tr>
<td>Graphene</td>
<td>0-0.2</td>
<td>180 [93]</td>
<td>10$^3$ [93]</td>
<td>14%, 1000 cycles 0.1V 20% Pt, 25°C</td>
<td>44, 12.3% loss after 1000 cycles [62]</td>
</tr>
<tr>
<td>TiN</td>
<td>0-1.1 [70]</td>
<td>20-50 [96]</td>
<td>40 (bulk) [97]</td>
<td>555 (s.crystal) [70]</td>
<td>1 (initial) 100 cycles [75]</td>
</tr>
<tr>
<td>PdP/C</td>
<td>Assumed 0-0.6V [67]</td>
<td>42-46 [67]</td>
<td>-</td>
<td>30%, 10000 cycles 0.6-1.0V, 25°C</td>
<td>10 (initial)</td>
</tr>
<tr>
<td>C$_3$O$_2$P$_2$O$_7$/C</td>
<td>Assumed 0-0.69V [69]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TiO$_2$</td>
<td>0-1.2</td>
<td>266 [87]</td>
<td>3.0 x 10$^6$ [84]</td>
<td>50%, 2500 cycles 0-1.4V, 80°C 60% Pt [87]</td>
<td>10 (at 0.75V H$_2E$) to ≥ 0.1 (at 0.95V H$_2E$), 2500 cycles [87]</td>
</tr>
<tr>
<td>WO$_3$</td>
<td>&gt; -0.2V</td>
<td>47 [92]</td>
<td>3.17 x 10$^{-4}$ [95]</td>
<td>60%, 100 cycles 0.12V [52] 40% Pt, 30°C</td>
<td>0.001, approx 5% loss after 100 cycles [52]</td>
</tr>
<tr>
<td>WC</td>
<td>Assumed 0-1.3</td>
<td>14.6 [78]</td>
<td>-</td>
<td>25% 100 cycles 0.6-1.8V 80°C, 40% Pt [78]</td>
<td>0.002, no change after 100 cycles [78]</td>
</tr>
<tr>
<td>SnO$_2$</td>
<td>0-1.3V</td>
<td>127.7 [69]</td>
<td>5 x 10$^{-1}$ [98]</td>
<td>11.3%, 10hrs 1.6V 23%, 1800 0.6-1.2V cycles 60% Pt, 60°C [89]</td>
<td>0.2 (initial), 0.025% after 2 hrs at 1.6V [89]</td>
</tr>
</tbody>
</table>

Table 1: Summarised performance characteristics of possible PEMFC support materials in addition to Carbon Black (CB). All thermodynamic stability data provided by Pourbaix diagrams [25]. Figures for corrosion stability in terms of ESCA loss and ORR activity correspond to the same experiment in all cases but BET surface area data in some cases from a different source.
4.3 Density Functional Theory; review and technical introduction

4.3.1 The N-body electron problem

The starting point into an ab initio investigation of stable compounds must be precise knowledge of the electron states over one repeating unit of the crystal structure in which the material is to form. The total energy of a system of electrons in the atomic potential is properly defined by the Schrödinger equation, given here in SI units,

\[
\frac{-\hbar^2}{2m_e} \sum_{i=1}^{n} \nabla^2_i + \frac{1}{4\pi\varepsilon_0} \left( \sum_{i,j} \frac{Z_i e^2}{|r_i - R_j|} + \frac{1}{2} \sum_{j \neq j'} \frac{e^2}{|r_j - r_{j'}|} \right) \psi_e(\vec{r}, t) = i\hbar \frac{\partial}{\partial t} \psi_e(\vec{r}, t) \quad (18)
\]

\[
[H_{KIN} + V_{Ze} + V_{ee}] \psi_e(\vec{r}, t) = i\hbar \frac{\partial}{\partial t} \psi_e(\vec{r}, t) \quad (19)
\]

where \( m_e \) is electron mass, \( e = 1.6 \times 10^{-19} \text{C} \) is elementary charge, \( \varepsilon_0 \) is permittivity of free space, \( Z_i \) are nuclear charges, \( r_i, r_j \) are position vectors of the \( i \) and \( j \)th electrons, and \( R_j \) is the fixed position of the \( j \)th nucleus. The three main terms of the Hamiltonian operator are the kinetic energy \( H_{KIN} \), given by the squared derivative operator, and the potential energies of the electron-nuclear interaction \( V_{Ze} \) and the electron-electron interaction \( V_{ee} \). The object of interest is the many-body wavefunction \( \psi(r_1,...,r_n) \). By setting the right hand side of this equation equal to zero, the time-independent form is obtained,

\[
[H_{KIN} + V_{Ze} + V_{ee} - E] \psi_e(\vec{r}) = 0 \quad (20)
\]

where \( E \) is the energy of the state \( \psi_e(\vec{r}) \). This reflects the application of the Born-Oppenheimer approximation [101], which can be expressed as the constraint that the motion of the nuclei with atomic charge \( Z \) is negligible compared to that of the electrons, so that the wavefunction for the whole system is separable into two wavefunctions, one describing only the nuclei and the other describing only the electrons. The nuclear position vectors \( R_j \) enter into the equation as constants, as equation (18) is the wave equation for the electron wavefunction \( \psi_e \).

The corresponding many-electron energy \( E \) would describe the total electronic energy, a continuous function of the space defined by \( (r_1,...,r_n) \). The wavefunction \( \psi \) in principle contains all the physical information about the physical system that equation (18) represents. However, (20) would have to be solved exactly, requiring simultaneous solution of each one of the \( n \) Schrödinger equations for the \( n \) electrons. Even if a solution for \( \psi \) could be obtained, its representation as a single function of \( n \) position variables is not very useful in a periodic crystal system. For practical purposes, electron states are constructed from the eigenfunctions obtained from the Schrödinger equation for just one electron in a coulomb potential, the only problem that has an analytical solution. Bloch’s theorem guarantees that coefficients for the expansion possess the periodicity of the space lattice in which the crystal solid forms [102], so that total wavefunctions and energy are obtained as a function of the reciprocal space position coordinate \( k \) by applying symmetry conditions on the wavefunction that apply throughout the lattice. DFT proceeds by providing single-particle wavefunctions, incorporating the \( N \)-body physics of the problem in its...
formulation and integrating the solution as a function of reciprocal space $\mathbf{k}$ to obtain extrinsic quantities such as the total electronic energy.

DFT is principally based on the idea of a self-consistent field (SCF) which is a concept introduced below, also the foundation of Hartree-Fock (HF) theory to which DFT is in some respects a successor - hence, HF will first be reviewed.

### 4.3.2 The Variational Principle and Hartree-Fock theory

An alternative statement of the many-body electron problem is to find the wavefunction $\psi$ that minimises the total energy $E$ of the whole system. Since the total energy of a system of interacting electrons will be a functional of the wave function,

$$\langle \psi | \hat{H} | \psi \rangle = E \geq E_0 = \langle \psi_0 | \hat{H} | \psi_0 \rangle$$

(21)

an alternative formulation of the eigenvalue problem posed by equation (20) is to find the wavefunction $\psi_0$ that minimises the total energy. Wavefunctions corresponding to excited states cannot be calculated on the basis of (21), but are in principle obtainable from the ground state as linear combinations of the mutually orthogonal eigenfunctions from which it is constructed as long as there exists a basis in terms of these functions that is complete. The ansatz

$$\Psi_{HF} = \frac{1}{\sqrt{N}} \text{det} [\phi_1 \phi_2 \ldots \phi_n]$$

(22)

is attributed to Hartree [103], where the $\phi_i$ are independent single-particle wavefunctions with $\text{det} [\phi_1 \phi_2 \ldots \phi_n]$ the Slater determinant and the factor $1/\sqrt{N}$ enforces normalization. The Slater determinant is constructed so as to vanish when the same orbital is occupied by two electrons of parallel spin. The result is the Hartree-Fock (HF) equation, reflecting in its name the contribution of both physicists [104], given here in atomic units,

$$\left( -\frac{1}{2} \sum_i \nabla_i^2 + V_{\text{ext}} \right) \phi_i(\mathbf{r}) + \sum_{j=1}^{N} \left[ \int d\mathbf{r}' \phi_j^*(\mathbf{r}') \phi_j(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \phi_i(\mathbf{r}) \right]

- \sum_{j=1}^{N} \left[ \int d\mathbf{r}' \phi_j^*(\mathbf{r}') \phi_j(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \phi_i(\mathbf{r}') \right] = (\epsilon_i + J_{ij} - K_{ij}) \phi_i(\mathbf{r})$$

$$= \sum_{j=1}^{N} \epsilon_{ij} \phi_i(\mathbf{r})$$

(23)

where $V_{\text{ext}}$ is the external or nuclear potential. Each of the $N$ equations is for one single-particle wavefunction $\phi_i$. The second and third terms are the action on $\phi_i$ of two operators $J_{ij}$ and $K_{ij}$, so that the Hartree-Fock operator is $\hat{O}_{HF} = \hat{h} + \hat{J} - \hat{K}$. The $\epsilon_{ij}$ are coefficients of a matrix whose diagonal elements are the energies of the single particle wavefunctions

$$\epsilon_i = \epsilon_{ii} = \langle \phi_i | \hat{O}_{HF} | \phi_i \rangle$$

(24)
For completeness, it is worth noting at this point that as is usual in quantum theory, one obtains the total Hartree-Fock energy by taking the inner product of the Hartree-Fock wavefunction with the total Hamiltonian operator

$$E_{HF} = \langle \Psi_{HF} | \hat{H}_{HF} | \Psi_{HF} \rangle$$

This definition is useful because the Hartree-Fock approximation not only informs Density Functional Theory, but is one assumption used in the particular formulation relied upon to obtain many of the results in this study. The physical meaning of the terms in (23) is also helpful to understand.

The integral inside the second term is the modulus square of the wavefunction of an electron at an alternative location \( r' \), divided by the distance between points where this quantity and the \( \phi_i(r) \) are evaluated. From the statistical interpretation of QM, the numerator is the charge density of an electron in the state \( j \), so we conclude that the second term represents the interaction of an electron at \( r \) with the charge density at \( r' \), that has a contribution from each electron \( j \). Note that the integrated quantity calculates a local potential over the range \( r - r' \). The third term in (23) does not have a classical interpretation but its physical significance can be understood from the corresponding term in the expression for \( E_{HF} \), equation (25). This term contains a product of two wavefunctions \( \phi_j(r)\phi_i(r') \) for electrons in different spin states \( i, j \). By the Pauli principle, two electrons can have the same position variable only if one is in the spin-up state and the other in the spin-down state. More generally, the wavefunction of a fermion (of which an electron is an example) has to be \textit{antisymmetric}, meaning that its sign changes on the exchange of any two position coordinates or spin states of the particle \([105]\). This requirement can be expressed as

$$\Psi(r_1, r_2, \ldots, r_n) = -\Psi(r_2, r_1, \ldots r_n)$$

The Hartree Fock wavefunction (22) is decomposed into a product of single-particle orbitals, each one corresponding to a distinct electron state. The antisymmetry of such a wavefunction can also arise from exchanging the states,

$$\phi_1(r_1)\phi_2(r_2)\ldots\phi_n(r_n) = -\phi_2(r_1)\phi_1(r_2)\ldots\phi_n(r_n)$$

We see that for the relevant term in the Hartree-Fock energy, which represents the unique case of a two electron system, these possibilities are equivalent. Two electrons can have only the 'opposite' state or the same state, corresponding to spins anti-parallel or parallel. Anti-parallel
spins are energetically favoured. The physical consequence is reflected in the fact that moving two electrons of the same state and having the same spin together in space has an increasing energetic cost associated with it. The third term in (25) accordingly accounts for the electron interacting with the potential of two electrons in the same spin state and calculates the resulting energy for all different states of the system. It might be named the \textit{exchange} term to reflect the energetic consequence of spin swapping. The second term of (25) contains two terms equivalent to the mod square of the electron wavefunction $\psi^*\psi$. The integral hence always calculates a positive energy. This is the \textit{coloumb} repulsion of electrons, which always increases the potential minimum of the system. On the other hand the alignment of spins may result in an increase as well as decrease in potential energy. The exchange energy is an important concept in a many electron system, that has to be incorporated in DFT, though in a somewhat different way which is discussed following introduction to the theory.

The 'bare' Hartree-Fock equation (that is, without any refinements or additional approximations) has been shown to describe semiconducting transition metals such as Silicon, and simple ionic crystals, though with significant error on the experimental value of calculated parameters such as the binding energy \cite{110}. In general, the HF approximation will not produce reliable results for the binding energies, or other properties that are physical consequence of the valence electrons, although it is sometimes satisfactory for calculating total energies of systems where only the inner-shell electrons are relevant \cite{111}. This is because the HF equations completely disregard electron \textit{correlation}, which affects less tightly bound electrons over long distances and appears in the equations of DFT. Nevertheless, the term relating to electron exchange in (23) gives an exact exchange energy, $E_{\text{FOCK}}^x$, as it sums over all electron pairs, which means HF theory is still useful in parallel with DFT (see section 4.3.7).

The electron correlation can be treated perturbatively, which is the motivation for Møller-Plesset theory an example of a post SCF method whose relation to DFT and other methods is shown in an overview figure (4.8). This will be examined in the next section.

### 4.3.3 The Kohn-Sham equations

Modern DFT originates from the hypothesis of Kohn and Hohenberg that the ground state of a many-electron system may be characterised completely by just the electron density $\tilde{n}(\mathbf{r})$ in the ground state. This was formulated in a theorem \cite{101} stating that $\tilde{n}(\mathbf{r})$ uniquely determines the potential in which the electrons move. Since the density also provides the number of electrons $N$ in the system, it completely determines the Hamiltonian $H$ and hence, via the original Schrödinger equation (18), the wavefunctions. The relation should thus be invertible, producing a representation for the ground state $\tilde{n}(\mathbf{r})$ as a function of the ground state $\psi(\mathbf{r})$. Then the variational principle can be rewritten for the total energy as a functional of $\tilde{n}(\mathbf{r})$, minimisation with respect to $\tilde{n}(\mathbf{r})$ producing the equations that define $\psi$ \cite{112}. The object of minimisation is the ground state energy $E$ obtained by the integration over $\Psi$,

$$E = \min_{\Psi} \langle \Psi | H | \Psi \rangle$$

(28)
Figure 4.8: "Jacob’s Ladder" illustration of successive addition of contributions to the DFT ground state energy from the least accurate approximation to the most accurate. The Hartree approximation excludes the exchange-correlation energy $E_{XC}$, which is represented in DFT by the LDA and GGA gradient based approximations. The meta-GGA approximation includes an approximation to the kinetic energy through a system of non-interacting electrons. DFT+U adds the Hubbard parameter $U$ to parameterise the interaction between the spins of electrons.
It can additionally be assumed that there exists a functional $F[n(r)]$ for the remaining contributions to the potential energy for which variations of $n(r)$ will interpolate between the non-interacting and interacting system. The minimisation is then performed by constraining the trial wavefunctions to a sub-group that give a defined $n(r)$ upon integration $[101]$. Within the constrained space, the minimisation with respect to $n(r)$ is performed,

$$E = \min_{n(r)} \left\{ \int v(r)n(r)dr + F[n(r)] \right\}$$

(29)

defining the ground state energy $E$ which is obtained when $n(r)$ assumes the ground state density. In the proceeding discussion, the integral over $\Psi$ is implied by ket bra notation $| \rangle$ but is not performed explicitly because in practice, it is done in a way that is specific to the trial functions as per the DFT formulation. To demonstrate rigorously the connection between the interacting and non-interacting system, we can write the inequality $[112]$, $[104]$

$$E_v[\tilde{n}] \geq F[\tilde{n}] + \int v_{\text{ext}}(r)\tilde{n}(r)dr$$

(30)

where $\tilde{n}(r) = \sum_{i=1}^{N} |\psi_i(r)|^2$ is the electronic density of non-interacting electrons, since it is proven by Hohenberg and Kohn that the potential and hence the wavefunctions $\psi_i$ are completely determined by $\tilde{n}(r)$, which must therefore be reproduced by a sum over the modulus square of the set of $\psi_i$. The notation $E_v[\tilde{n}]$ denotes the energy functional that depends on the external potential $v(r)$, and $F[\tilde{n}]$ is the unknown functional of the charge density which distinguishes the non-interacting system in general from the case of the ground state, where the electrons do not interact between themselves and the potential energy is minimised by the optimal configuration of electrons in the presence of $v(r)$. Without yet elucidating the form of $F[\tilde{n}]$, we can continue the heuristic argument of the previous section and write that, since the electron interactions take the form of repulsions, and the interactions imply in addition an associated kinetic energy of the system, the minimum-energy principle for an interacting system thus becomes $[114]$, $[113]$, $[104]$

$$\langle \Psi_n | \hat{H} | \Psi_n \rangle \geq \langle \Psi_{\tilde{n}} | \hat{H} | \Psi_{\tilde{n}} \rangle = E_v[\tilde{n}]$$

(31)

where the subscript $\tilde{n}$ distinguishes the system with the non-interacting as opposed to the interacting electron density $n$. Since the total Hamiltonian for the fully interacting system is of the form $\hat{H} = \hat{T} + \hat{U}$, this can be expanded as

$$\langle \Psi_{\tilde{n}} | \hat{T} + \hat{U} | \Psi_{\tilde{n}} \rangle + \int v(r)\tilde{n}(r)dr \leq \langle \Psi_n | \hat{T} + \hat{U} | \Psi_n \rangle + \int v(r)n(r)dr$$

(32)

However, for the case where the wavefunctions of the non-interacting system $\Psi_{\tilde{n}}$ and the interacting system $\Psi_n$ produce exactly the same density, we simply have

$$\langle \Psi_{\tilde{n}} | \hat{T} + \hat{U} | \Psi_{\tilde{n}} \rangle \leq \langle \Psi_n | \hat{T} + \hat{U} | \Psi_n \rangle$$

(33)

and we see that the condition for the genuine ground state $\Psi_0$ is that it minimises the expectation value $\langle \hat{T} + \hat{U} \rangle$ for a given system. While this allows one to find a unique ground state $\Psi_0$ amongst any $N$-electron system of a given density, it also allows an interacting example of such a system
to be mapped onto a non-interacting one. Clearly, given a non-interacting system with density $n(r)$ and external potential $v(r)$, for an alternative potential $v'(r)$ the system can be equivalent to an interacting system for a value of the same $n(r)$. Equation (33) states that the ground state $\Psi_0$ can be found if the kinetic and potential energies $T$ and $U$ can be represented as a functional of $\tilde{n}(r)$. It has already been seen that the minimisation of the total energy with the most general scalar product of wavefunctions leads, in the third term of the Hartree-Fock energy expression (25), to a potential energy of all electrons interacting individually with the averaged local potential of the electron density. This contributes to the potential energy $U$ in DFT for the case that the single particle wavefunctions are in the same spin state,

$$U[n] = \frac{1}{2} \int \frac{n(r)n(r')}{|r-r'|} drdr'$$  \hspace{1cm} (34)

The non-interacting contribution to the kinetic energy $T$ is retained by an introducing an artificial system of $N$ non-interacting electrons [109],

$$T[n] = -\frac{1}{2} \sum_i^N \langle \psi_i | \nabla^2 | \psi_i \rangle$$  \hspace{1cm} (35)

and is in terms of $n(r)$ owing to the symmetry of the squared gradient operator. However, the remaining contributions to the kinetic and potential energies are a consequence of the effects of the many-body electron interactions and can only be approximated in terms of electron density. In addition to the exchange energy associated with the spins of an electron pair, a fraction of the total energy is due to electron correlation, whereby a certain amount of information about the system is contained within interactions between the electrons at point in space that are not incorporated by calculating the interaction of each electron with the mean field. The exchange correlation functional $E_{xc}[n(r)]$ incorporates the energy contributions that cannot be written as functionals of the charge density, and needs to be written as a function of $n(r)$ that best reproduces the many body effects of the system studied. It contributes inexact exchange energy (as opposed to exact, in the case of the Hartree-Fock equations) in addition to correlation.

Accounting for the four terms, kinetic, external, electron-electron and exchange correlation, the total energy functional is therefore

$$E_v[n(r)] = \int v(r)n(r) + T[n(r)] + \frac{1}{2} \int \frac{n(r)n(r')}{|r-r'|} drdr' + E_{xc}[n(r)]$$  \hspace{1cm} (36)

The exchange correlation functional is unknown and has to be approximated. The exchange correlation potential is introduced, [101]

$$v_{xc}(r) = \frac{\delta}{\delta n(r)} E_{xc}[n(r)]|_{n(r) = \tilde{n}_0(r)}$$  \hspace{1cm} (37)

which is a local quantity due to the explicit dependence on $r$ in $E_{xc}$, treating all the non-classical exchange and correlation effects.
Using the derivative defining $v_{xc}$ in (37), functional derivatives are evaluated at the minimum density $\tilde{n}_0(r)$ corresponding to the ground state. Of all the variations of $n(r)$ one is interested only in those that keep the total number of particles fixed. In terms of $v_{xc}$ the Kohn-Sham (KS) equations can be written [101],

$$\left(-\frac{1}{2} \nabla^2 + v(r) + \int \frac{n(r')}{|r-r'|} dr' + v_{xc}(r) - \epsilon_i\right) \psi_i(r) = 0$$

which take the form of $n$ single-particle eigenvalue equations, where $\epsilon_i$ have the appearance of single particle energies, though this interpretation is not valid as the energy of any single electron state is not defined in DFT, only the total electronic energy. Each electron experiences the potential of the mean coulombic interaction due to all other electrons, as well as the nuclear potential $v(r)$ and the exchange correlation potential of unknown nature defined by (37). Since the coulombic interactions are defined by $n(r)$, which is itself defined from the single particle wavefunctions, $\tilde{n}(r) = \sum_{i=1}^{N} |\psi(r)|^2$, the set of KS equations could be solved by a self-consistent method if $v_{xc}(r)$ was known. Starting from a trial $n(r)$ and obtaining the single particle wavefunctions $\psi_i(r)$ produces the next estimate of $n_0(r)$, which is substituted to yield the next estimate $\psi'_i(r)$, this process continuing until both wavefunctions and density have converged to the required accuracy; the self-consistent condition is thereby reached. For this to be possible, of course, a representation of $v_{xc}(r)$ in terms of $n(r)$ is required.

### 4.3.4 Improvements on basic DFT- the exchange correlation functional $E_{xc}$

#### 4.3.5 The Local Density Approximation (LDA)

The most primitive treatment of the exchange correlation potential $v_{xc}(r)$ is based on the homogeneous electron gas. In this simple model, $n$ electrons fill a volume $V$ with half in the spin-up state and half in the spin-down state. The charge neutrality condition is maintained by assuming that a positive counter charge is uniformly distributed throughout the space [117]. The electrons then obey the single-particle Schrödinger equation and the density of electron states over space is also uniform. Since what is needed for $E_{xc}[n]$ is ideally some function of $n(r)$, we can find a locally uniform approximation by looking for functionals that depend only on the density itself and not its derivatives or integrals. Dimensional considerations fix the general functional forms [116]

$$T_{local}[n] = \frac{3}{5} A(N) \int n^\frac{3}{2} dV, \quad V_{ee}[n] = \frac{3}{4} B(N) \int n^\frac{3}{2} dV$$

The most useful parameterisation of the exchange correlation separates it into two terms, one reflecting purely the correlation energy, and the other the exchange energy,

$$E_{xc}[n] = E_{c}[n] + E_{x}[n]$$

The motivation comes from the fact that there is a simple, exact result for the exchange energy which can be derived for the case of the homogeneous electron gas. Since the ratio of electron
spins in the up and down state is assumed to be exactly unity, one can approximate the spin-polarised charge density $|n_\pm|$ of one electron that has been exchanged with another, or switched from a spin-up state to a spin-down state, as a quantity which falls off as a point-like potential inside a region where the total charge is $e$. This volume of space is the exchange hole with radius $r_0 = \left( \frac{3e}{4\pi |n_\pm|} \right)$.

Since the exchange charge density represents the redistribution of the average charge density into local concentrations of spin-up and spin-down charge, there should be a dependence on the average charge density. In the regime of the homogeneous electron gas, the states that each electron is assumed to occupy fall within the 'Fermi sphere' of reciprocal space $\sim k$. The density of states within is uniform and depends on the number of electrons $x$,

$$ k_F = \left( \frac{3x}{8\pi V} \right)^{\frac{2}{3}} $$

which are assumed to be uniformly distributed throughout the space $\vec{k}$, and therefore represented by $n/V$. Combining (41) and (42) produces the result for the homogeneous exchange potential energy $E_{\text{homo}}$,

$$ E_{\text{homo}}[n(r)] = 3 \left( \frac{e^2}{4\pi \varepsilon_0} \right) \left( \frac{3n(r)}{8\pi V} \right)^{\frac{1}{3}} $$

where we have introduced a local function $n(r)$ which is now to be determined by the functional variation of $E_{\text{xc}}[n(r)]$, c.f (37). The energy measuring just the correlation of two spatially separated electrons $E_c[n]$ as opposed to the energy of exchanging two states, now has to be approximated. Useful parameterisations can be obtained as expansions in terms of the Wigner Seitz radius $r_s = (4\pi n(r)/3)^{1/3}$ [111], [118], which is the radius of a sphere equal in volume to the average volume per atom and therefore has the same dependence on the local density function as the exchange potential energy.

### 4.3.6 Generalised Gradient Approximation (GGA)

Another approximation based on the LDA but often producing significantly different results takes the form of $E_{\text{xc}}[n(r)]$ as a function of not only the local density $n(r)$ but also its gradient,

$$ E_{\text{xc}}[n(r), |\nabla n(r)|] $$

Physically, this means that one is changing the domain of $n(r)$, the region inside which the approximation of the homogeneous electron gas is valid. This ensures that the analytical features of the unknown exchange correlation functional will be maintained. Functionals involving higher order derivatives under the nomenclature of meta-GGA approximations have also been constructed [119]. Although the GGA approximation and its extensions further the logic of the LDA, it is not the case that GGA is more refined, or produces better results; the suitability of
each functional depends on the physical system that is being calculated, as explained in more
detail in sections (5.5.2) and (5.5.3). A summary of the accuracies attained for a variety of func-
tionals will shortly be discussed. It is important to note that one can also use a combination of
the broad approximation schemes discussed so far.

4.3.7 Hybrid Functionals

It has been found that often the performance of DFT is optimised by using a functional made up
of a chosen weighting of the forms used in the Hartree-Fock and LDA approximation schemes.
Although Hartree-Fock theory does not account for electron correlation, the third term in the
Hartree Fock equation (25) as discussed in section (4.3.2) represents the exact exchange energy if
the summation over all electron state pairs can be evaluated, whereas in DFT outright exchange
effects are incorporated along with those due to electron correlation and do not receive exact
treatment as a result. For this reason, usage of a pure Hartree-Fock potential as a component in
the overall functional may be preferable. In addition other near-exact exchange potentials have
emerged, constructed so as to reproduce the behaviour of the exchange correlation potential
over the regimes of differing interaction strengths [120]. The B3LYP functional proposed by
Becke [120] is defined as a linear combination of a number of exchange, correlation and exchange-
correlation functionals,

$$E_{xc} = E_{xc}^{LSDA} + a_0(E^{exact} - E_{xc}^{LDA}) + a_x \Delta E_{x}^{B88} + a_c \Delta E_{c}^{PW91}$$  \hspace{1cm} (45)$$

where $E^{exact}$ is the exact exchange energy, as is calculated in Hartree-Fock theory, $E_{x}^{B88}$ is the
gradient correction to the exchange energy also introduced by Becke [121], $E_{x}^{LDA}$ and $E_{xc}^{LDA}$
are exchange and exchange correlation energies within the LDA approximation and $\Delta E_{c}^{PW91}$ is
the gradient correction to the correlation energy introduced by Perdew and Wang [127]. The
semi-empirical coefficients as determined by a fit to experimental data [120] take the values
$a_0 = 0.20$, $a_x = 0.72$ and $a_c = 0.81$.

4.3.8 Brief evaluation of DFT suitability and outlook

We have seen that for a system with correlations and exchange interaction between electrons,
the lower bound on the total energy will be the ground state of the non-interacting system,
because the unique definition of the ground state by the charge density means that there will be
an interacting system defined by an alternative potential for which the same electron density will
produce the ground state energy. If a sufficiently accurate form for $v_{xc}(r)$ can be found, it should
be possible to determine the ground state up to a satisfactory degree of accuracy. However,
there remain some systems which are not adequately represented by DFT nonetheless. One
reason for this is that, in allowing the average density as mediating the interactions between
electrons, a fraction of energy of an $e^-$ in the potential $v(r)$ will be due to that electron effectively
interacting with itself. The Hartree Fock potential evaluates each pairwise interaction and hence
by definition there is no self-interaction, which is why it is often desirable to include it in a hybrid DFT functional.

The shortcomings of DFT can be significant for the systems studied by applying it in this work, due to the strong spatial localisation of electron states in $d$ and $f$ orbitals and consequent correlation. In particular, the band gaps of transition metal oxides are often underestimated by conventional DFT [123]. By using an appropriate hybrid functional with some Hartree-Fock contribution, both band gaps and the description of localised states is much improved [123], [122]. One particular study by Martínez et al. [124] into the formation energies of rutile metal dioxides is noted here, as this family of materials is of primary interest. The investigation calculated theoretical Gibbs Free Energies of formation, calculated as the differences between the DFT total energy bulk dioxide and the bulk metal and between water and hydrogen gas [124], [125] in order to give an appropriate reference value for the free energy of the oxygen atom. The functionals used were the PBE (Perdew-Berke-Erzenhof), [126], [127] a commonly used functional of the GGA type which builds on a number of previous versions, and its own extension, the revised (RPBE) functional [128]. Using the revised form of the functional the average absolute error on the formation energies was 0.29 eV, but the more primitive GGA functional gave much poorer results. Unfortunately, the RPBE functional has not been developed for the DFT codes used in this report. A similar study for LDA functionals or hybrids involving either LDA or GGA could not be found, so the preliminary results presented in the results section were calculated with hybrid functionals based on both of these two broad approximation regimes. In a study using the B3LYP functional for the La–Mn–O system [129], the formation energies of binary and ternary compounds studied was accurate within an absolute error range of just 5.4%, giving confidence in the use of the functional for systems whose electron interactions require accurate treatment of exchange and correlation.

4.3.9 Plane Wave vs. Gaussian basis functions

To solve the single-particle Kohn-Sham (KS) equations (38), DFT codes expand the single-particle wavefunction as a linear combination of basis functions. For the case of a periodic crystal where the electrons move in a periodic lattice potential, Bloch’s theorem guarantees that the wavefunction in reciprocal space $\mathbf{k}$ can be expanded in terms of functions $u_{\mathbf{k},i}(\mathbf{r})$ that are periodic over the crystal lattice [102],

$$\psi_{\mathbf{k},i}(\mathbf{r}) = \sum_i u_{\mathbf{k},i}(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}}$$  (46)

where $u_{\mathbf{k},i}$ meets the requirement

$$u_{\mathbf{k},i}(\mathbf{r}) = u_{\mathbf{k},i}(\mathbf{r} + \mathbf{T})$$  (47)

with $\mathbf{T}$ a lattice translation vector. The wavefunction of an electron in state $i$ can, therefore, be expressed in terms of the function $u_{\mathbf{k},i}(\mathbf{r})$. Equation (46) is the starting point for the self-consistent solution of the KS equations. According to the band description of crystalline
solids, there is partial occupancy of each band by each state $\psi_i$. A representation for the total wavefunction can be obtained valid for each of $n$ points on the reciprocal lattice, and the KS equations are solved at each point until the self-consistent condition is reached at all points. This gives the fictitious ‘single-particle’ energies $\eta_i(\tilde{k})$ at each point, so that the total energy can be obtained by an interpolation technique, taking into account the electronic density of states obtained from the wavefunction. Different DFT formulations, however, use different types of basis functions, and it is known that for a particular type of solid, some give reliable results and descriptions of physics for similar systems that correlate with each other whereas others do not describe the physics correctly, resulting in inconsistent or contradictory physical properties.

The calculations documented in the results section used two codes: one, VASP [130], uses plane-wave basis functions and the other, CRYSTAL09 [131], uses gaussian-type functions (GTFs), which are optimised for different systems.

**Plane Wave basis sets.** With this approach the wavefunction is expressed as a linear combination of plane waves having the periodicity of the reciprocal lattice vector $\tilde{K}$,

$$\psi_{\tilde{k}}^I(\tilde{r}) = \sum_{\tilde{K}} C_{I,\tilde{K}} e^{-i(\tilde{k} + \tilde{K}) \cdot \tilde{r}}$$

which are a mathematically complete set of functions. This means that in principle the summation extends over all vectors ($\tilde{k} + \tilde{K}$) though in practice, the required convergence of the total energy can be obtained if one uses the set of vectors having a kinetic energy of less than $E_{\text{cut}}$. In addition to the wavefunction, the nuclear potential must also be approximated in terms of the same set of plane waves, to satisfy Bloch’s theorem. A suitable pseudopotential is defined to have the property that the integral from 0 to $r$ of the charge density calculated from it is equal to that obtained using the true potential for $r > r_c$, where $r_c$ is a cut-off radius, whatever valence state the electrons are in [132]. Pseudopotentials should be ‘smooth’, (‘soft’), meaning that the total energy converges rapidly with the number of basis functions used [133], as the expansion in terms of a larger set of functions increasingly approximates the full character of the true potential. The number is effectively governed by the atomic radius [133], which determines the range of $r$ over which the pseudopotential is valid before one encounters the cut-off radius $r_c$. Since the number of plane waves included in the expansion is cut off for a certain value of the kinetic energy $E_{\text{cut}}$, this means that one wants to find a set of plane wave basis functions that minimises the kinetic energy, but since in a mean-field theory the kinetic energy is related to the derivative of the wavefunction, as given by (35), the $\tilde{k}$ components of the (pseudo) wave function with a higher rate of oscillation are higher in kinetic energy, so that to optimise the convergence one has to selectively minimise the kinetic energy of the high $\tilde{k}$ components [134]. A plane-wave method that applies the same logic is the projector augmented wave method (PAW), which was used in VASP for the calculations documented in this report. This method transforms the problematic high $\tilde{k}$ components of the wavefunction into a different ‘space’, [135] preserving the requirements that are placed on a wavefunction existing in the true Hilbert space.

**Gaussian Type Functions (GTFs).** The second broad category of basis functions are based on local functions (atomic orbitals, AO) constructed as linear combinations (‘contractions’) of
GTFs, which are centered on atomic sites within the unit cell of the lattice. Each takes angular momentum quantum numbers \(l\) and \(m\), and is defined in CRYSTAL09 as [136]

\[
\varphi_{l}^{m} = N \sum_{j} d_{j}^{l} c_{j}^{m} \chi(\alpha_{j}; \vec{r})
\]

where \(\vec{A}_{i}\) is the nucleus of the atom in the region where the Gaussian type function, \(\chi\), is centred, \(c_{j}^{lm}\) and \(N\) are normalization constants where \(c_{j}\) is defined for each pair of angular momentum quantum numbers \(m,l\), and \(d_{j}^{l}\) are expansion coefficients for the \(i\)’th function in terms of the set \(j\) of the local functions. The Gaussian type functions \(\chi\) are separated into radial and angular parts \(G\) and \(X_{lm}\),

\[
\varphi_{l}^{m} = N \sum_{j} d_{j}^{l} c_{j}^{m} X_{l}^{m}(\vec{r}) G(\alpha_{j}; \vec{r})
\]

The angular part of the function is a linear combination of solid harmonic functions \(X_{lm}\). The radial function is a Gaussian function \(G\) with a coefficient \(\alpha_{j}\) governing the diffusivity as a function of distance (in real space) from the atomic site,

\[
G(\alpha_{j}; \vec{r} - \vec{A}_{i}) = \exp(-\alpha_{j}(\vec{r} - \vec{A}_{i})^{2})
\]

From the AO given by equation (49), one can construct the periodic Bloch functions as in (46),

\[
\phi_{i}(\vec{r}, \vec{k}) = \sum_{\vec{K}} \varphi_{i}(\vec{r} - \vec{A}_{i} - \vec{K}) e^{i\vec{k} \cdot \vec{K}}
\]

Finally the total wavefunction is a further linear combination of the Bloch functions,

\[
\psi_{\mu}(\vec{r}, \vec{k}) = \sum_{i} a_{\mu,i}(\vec{k}) \phi_{i}(\vec{r}, \vec{k})
\]

where \(a_{\mu,i}(\vec{k})\) is the Bloch expansion coefficient. Returning to the AO, equation (50), using the definition of the harmonic functions [137], enables to write the function in full as

\[
\varphi_{l}^{m}(\vec{r}) = N X_{l}^{m} \sum_{j=1}^{N_{\text{prim}}} d_{j}^{l} c_{j}^{m} e^{-\alpha_{j} r^2}
\]

where \(X_{l}^{m}\) are spherical harmonic functions and \(N_{\text{prim}}\) is the number of GTF used. The value of the angular momentum quantum number \(l\) corresponding to the shell of the orbital \(s,p,d\) results in shifting the maximum of the function to different values of \(r\). In practice, the AO is constructed by a sum over the functions for different shells weighted by the coefficients \(d_{j}^{l}\), as represented in figure (4.9).

The Gaussian functions are not a complete set and are not mutually orthogonal functions, unlike plane waves. However the degree of quasi-linear dependence between the basis functions can be measured by the overlap between Bloch functions \(|\phi_{i}\rangle\langle\phi_{k}|\) [136]. Physically, this means that when the Gaussian functions are superimposed on eachother in space, the character of distinct functions can be lost, so it is as though one GTF can be expressed as a linear combination of others even though formally this linear combination does not exist. The spatial representation
4.3 Density Functional Theory: review and technical introduction

Matthew Worsdale

Figure 4.9: Unnormalized Gaussian Type Functions (equation (50)) used in the CRYSTAL09 program for 1s (red line) and 2p (green line) orbitals. Also plotted is a linear combination of these two orbitals (blue line), shifted by a transformation \( r \to r - 0.19 \) to give \( \varphi(r = 0) = 0 \), to illustrate the operation of construction of AO from superposition of Gaussian Type Functions used in the CRYSTAL program.

of the GTF is parameterised by the coefficient \( \alpha_j \), which is a free parameter that may be adjusted to change the basis set used for a particular atomic species. However, this requires the dependence of the total energy on \( \alpha_j \) to be known so that the rescaling of energies can be computed. As with the plane-wave method, a pseudopotential must also be used and can easily be represented in terms of GTFs [138].

### 4.3.10 Observable properties of DFT solution- bandstructure and charge density differences

In order to make quantitative predictions on the availability and stability of defects in SnO₂, it is fundamental to have a description of the mechanism by which they are absorbed into the host, and diffuse and bond with the lattice, in terms of the electronic states involved in the mechanism, with their relation to the host and defect states. Accurate values for the band filling in the zero temperature regime of DFT enable to make a prediction of electron transfer through defect activation and free electron transport as a function of temperature \( T \). The solid state calculations employed for defects can give properties obtained from the DFT solution from which one can distinguish the alteration to the lattice. One of the most informative is the bandstructure that describes qualitatively the states introduced. Energy bands in semiconductors are determined by the AO overlap throughout the BZ as a function of a given path over lattice vectors in the reciprocal space, and in combination by the electron repulsion mediating the transfer between sites.
To show how these elements are treated, the procedure for solving the periodic system in DFT is summarised for the case of CRYSTAL, since this program is used for most of the results in this work. The two elements are integrated in an implementation of LCAO, as written in equations (49) to (53),

$$\psi_i(\vec{r}, \vec{k}) = \sum_{\mu} a_{\mu,i}(\vec{k}) \phi_{\mu}(\vec{r}, \vec{k})$$

as contributions to the Bloch functions $\phi_i$ from the overlapping of AO, weighted by the density matrix elements in terms of the expansion coefficients of Bloch functions $a_{\mu,i}(\vec{k})$ integrated over the BZ [131],

$$P_{\mu,\nu}^n = 2 \int_{BZ} d\vec{k} e^{i\vec{k} \cdot \vec{n}} \sum_j a_{\mu,j}^*(\vec{k}) a_{\nu,j}(\vec{k}) \delta(E_F - E_j(\vec{k}))$$

From this one can define the total electronic energy (per unit cell) by summing over all expansion coefficients and lattice vectors,

$$E_T = \frac{1}{2} \sum_{\mu,\nu} \sum_K P_{\mu,\nu}^K (H_{\mu,\nu}^K + F_{\mu,\nu}^K)$$

where $H_{\mu,\nu}^K$ contains the one electron nuclear potential and kinetic energy densities and $F_{\mu,\nu}^K$ are the exchange and coulomb contributions. The observables (density of states and bandstructure) are then given by the crystalline orbitals constructed by the coefficients $a_{\mu,i}(\vec{k})$ which determine $E_T$ to within the required convergence criterium (making the system self-sufficient). The combination of AO and electrostatic interactions are composed of coulomb, exchange (Hartree Fock) and other exchange correlation corrections fixed by the DFT functional determining the form of interaction between states through orbital mixing. As a result, DFT is closely related to the Tight-Binding Approximation (TBA) in its treatment of solid state electronic structure [193], in which electrons interact between nearest neighbours through AO centred on the atomic sites. One important difference is that in DFT AO overlap is calculated between sites at all lattice vectors (though expressed in each case in terms of the basis $\mu$ local to the given site) whereas in most formulations of TBA each overlap integral considers up to just three nearest neighbour sites [195]. Consequently in DFT, electrons can interact through and are correlated by properties of AO effective over more than one length scale simultaneously. The considerations include orbital symmetry, diffusivity and the energy difference between a pair. Coordinating these features is potentially important in a DFT description of a system with many possibilities for the orbital and site occupation by the valence electrons, such that many related configurations are possible thermodynamically on introduction of defects to the host.

The limit of validity of DFT and TBA becomes important when one studies materials that are substantially disordered by defects or site substitutions. In these situations, electrons can become strongly correlated to the extent that the coulomb repulsion, between electrons occupying the same AO on the relevant neighbouring atoms, needs to be included in the Hamiltonian, weighted by the Hubbard parameter $U$, to complement the AO overlap on which TBA band theory is based. This is important because this on-site interaction can be strong enough for electrons in spin-up and spin-down states to force filling of bands exclusively by spin-up or spin-down electrons if $U$ dominates over the electron ‘hopping’ between the overlapping integrals,
depending on the atomic spacing that is a function of the thermodynamic conditions. The choice of standard DFT or DFT+U can, therefore, be important to the prediction of the conductivity as a function of temperature. Other models such as DMFT are used to address the more finely grained effects of electron correlation by attempting to include the instantaneous correlation of electrons as they interact with the average potential. Corrections may also be used to overcome the breakdown of the pseudopotential approximation for systems with unpaired electrons due to overlap of core and valence basis functions [194]. These extensions to DFT are beyond the scope of this work.

In ab initio calculations based on DFT, each energy band is determined as a solution to an eigenvalue problem with $\mathbf{k}$. Periodic boundary conditions enforced on the solution determine the domain of each band within the computational space. Assuming that the bandstructure has been computed in this way for a bulk system without defects, one is then interested in the moderation that occurs when a defect is introduced.

### 4.4 Crystal structures of SnO$_2$

A crystal structure consists of a space lattice describing the geometry of a repeating unit of the structure. This unit cell contains a number of lattice points at which atoms are sited. A basis for generating the crystal structure, by translating the lattice points by the lattice vectors, is then defined by the set of atoms associated with each of these points. The first consideration behind the crystal structure of SnO$_2$ is the bonding arrangement formed when the Sn and O atoms approach each other to form a repeating unit. Because of the relatively high ionization energy of Sn compared to most transition metals, it is energetically favourable to form covalent bonds even though O has a significantly different (lower) electronegativity (in the Pauling scale) compared to Sn, which promotes ionic bonding for most metal oxides. Accordingly, tin is six-fold coordinated to oxygen on account of sharing of the six 2$p^4$ electrons of O with the 2$p^2$ electrons of Sn resulting in a bonding plane of four oxygen with one set of ligands $c$ directed between $p_1$ orbitals, and two lone pairs forming ligands $a$ perpendicular to this plane, parallel to the orbitals $p_2$ [139]. This octahedral coordination geometry with two ligand lengths $a,c$ can be accommodated by a lattice having two different axes lengths. The most stable polymorph of SnO$_2$ is the rutile structure which can be described within the framework of a hexagonal close packed (HCP) crystal structure with a tetragonal unit cell geometry. This structure is taken heuristically to be the most stable by the principle of closest packing of spheres, which is maximised for the HCP lattice [102].

**Rutile** can be visualised as a lattice of hexagonal close packed oxygens, with Sn placed at the octahedral interstitial positions, of which half the available positions are filled. This is indicated in figure (4.10). Because there are two different Sn-O ligand lengths, however, four of the oxygen in the coordination octahedra are the nearest neighbours of an Sn atom and each is coordinated to three Sn (as opposed to six, since half the Sn positions are filled). In addition, there are three different oxygen nearest neighbour lengths. The polyhedra are stacked in the $c$ axis of a tetragonal unit cell, reducing the number of oxygens per lattice point from six to...
three. The increased length of the $c$ ligand with respect to the shorter one $a$, then allows one additional oxygen to be shared between two polyhedra offset in the $c$ direction, an example of edge sharing. In SnO$_2$ having perfectly uniform geometry, as viewed in the $\bar{a}b$ plane, each next nearest octohedron in the $\bar{c}$ direction is rotated by $90^\circ$ with respect to the previous one.

**Anatase.** The sharing of oxygen atoms between lattice points to give the required stochiometry can alternatively occur if each polyhedra shares six of its twelve edges with neighbouring octohedra. This takes place in the anatase structure by the adoption of ‘zigzag’ packing in which two of these edges are shared with polyhedra in the bonding plane offset in the direction $\pm a$ and two pairs of edges are shared with polyhedra offset in the direction $\pm b$, where the axes refer to a tetragonal unit cell. In this case there are three different ligand lengths, and three different lengths of vertices of the octohedra, as shown in figure (4.10). While pure anatase SnO$_2$ is not found experimentally, it is possible that anatase crystallisation could occur by changing the composition of pure SnO$_2$, increasing the Gibbs free energy of the rutile SnO$_2$ phase.

### 4.4.1 Phase competition of disordered SnO$_2$ hosts

In order to study how disorder arises thermodynamically by varying the chemical composition of SnO$_2$, it is necessary to consider the phases that are associated with the crystal structures. The perfect structure minimises its free energy by maximising the packing of the atoms in a given volume, and for SnO$_2$ this is achieved by crystallising in the rutile structure, whereas experimental databases of bulk SnO$_2$ crystal structures report other structures only being associated with high pressure or temperature phases [140]. One way higher energy states can be realised within a lattice is through the propagation of the phonons, which govern the directions in the crystal in which atoms can oscillate simultaneously and which depend on the interatomic forces due by the interactions between atoms in the system as a function of distance. For small oscillations, the free energy $G$ is symmetric as a function of volume of the crystal $V$ about the minimum $V_0$, so that increases in $G$ can excite the system into a higher energy state but this will be unstable and the ground state $G_0$ will be recovered.

In the limit of a single substitution of Sn to another atom, the volume of the coordination polyhedra can be expected to be similar, since the atom experiences a rutile bonding environment and it will be favourable for the distortion of ligands from equilibrium positions to be distributed throughout the lattice. As the number of sites substituted increases, a bonding geometry of the dopant atom and ligands sufficiently different from that of SnO$_2$ will promote change in the coordination octohedra volume, and localised lattice distortions. These distortions become coordinated with increasing ways that the ligands within a given polyhedra can change under lattice distortions, while maintaining the original crystal structure. It becomes increasingly likely that length changes of ligands will coincide along the given crystal directions along which phonons propagate, resulting in coupled distortions of the crystal. If the crystal is sufficiently large, the effect of replacement of atoms of Sn with other species in the bulk will be a directionless strain, and the distance between two equivalent atoms $O_x$ can be used as an order parameter to characterise the phase of the system. The strength of the effect of
4.4 Crystal structures of SnO$_2$

Matthew Worsdale

Figure 4.10: Stable and possible disordered crystal structures of SnO$_2$. All coordination polyhedra are shaded grey. **Top left:** The rutile structure is derived from a hexagonal close packed (HCP) lattice of oxygen atoms with the unit cell indicated by the dashed yellow line. The tin atoms occupy the octahedral interstitial positions, while the unused tetragonal interstitial sites (in which an Sn would be coordinated to four oxygen atoms) are indicated for comparison. **Centre:** The rutile structure arises due to the extension of two of the six ligands (in green) which gives planes of distorted octahedra offset in the $c$ direction and stacked vertically, sharing edges, which makes the structure thermodynamically stable by maximising the closest-sphere packing of atoms. The three fold coordination of oxygen by tin is highlighted in red. **Right:** View of the rutile and CaCl$_2$-type structures in the $c$ direction. They are distinguished by the rotations of coordination octahedra about the shared edge, with the upper image (rutile) demonstrating the clockwise and anti-clockwise rotations of alternating planes. The lower image shows the CaCl$_2$ structure in which neighbouring octahedra rotate to maintain an angle between them with respect to the position in the rutile structure. Both images taken from reference [141]. **Bottom Left:** A unit cell of the anatase structure, in which the coordination octahedra have three different vertex lengths highlighted by yellow, green and red lines. The octahedron lying at the centre of the unit cell shown shares two further edges (in yellow) with octahedra either side of it in the $ab$ plane to make the ’zigzag’ packing seen in the directions $\bar{a}$ and $\bar{b}$. The two longest vertices (in red) are not shared.
a given substitution can be taken as the radius ratio of Sn and the dopant, which determines the strength of distortions resulting from the pressure. The pressure or the temperature can act as the variable of state characterising the phase. Then the system can undergo a phase transition that may be continuous (second-order) or discontinuous (first-order). In the discontinuous case, the transition corresponds to a transformation between crystal structures such that the order parameter $O_x$ changes discontinuously as a function of $p$ or $T$ at the point of the transition, which only occurs with a transfer of latent heat to or from the environment. There can also be metastable phases in the transition to the new phase, where the solid has some domains of one crystal structure and other domains of a second crystal structure. However, no crystal structures for binary alloys Sn$_{1-x}$D$_x$O$_2$ (D = Ta, Sb, Nb) are reported experimentally, so the consideration is limited to the stability of single phase crystal structures perturbed by doping.

In the continuous phase transition, the distortion of the crystal keeps it in the same phase until a critical transition temperature (or pressure) $T_c$. The system then transitions to a new phase defined by the breaking of a symmetry of the phase before the transition. More specifically the space group of the phase formed is a sub-group of the original phase space-group (first Laundau criterion). Such a transition can occur between rutile symmetry and the CaCl$_2$-type structure [141] (space group $Pnmm$) which both have coordination octohedra stacked in one plane, whose relation can be seen from figure (4.10).

The octohedra can undergo mutual rotational vibrations in the $\vec{a} b$ plane about their shared edges. In the rutile structure, this gives synchronous motion of alternating planes of octohedra in the $\vec{a}$ and $\vec{b}$ directions, i.e. clockwise and anti-clockwise. This, however, is not the most symmetric mode of oscillation possible; in the CaCl$_2$-type structure, each octohedra stack rotates about its neighbour so that every point in the crystal is invariant at any point in the vibrational motion. There exists a continuous phase transition taking the rutile phase to the CaCl$_2$ phase, and is observed experimentally as a function of pressure [142]. Changing the species of Sn sites in rutile SnO$_2$ could therefore induce this phase transition in the disordered structures studied in this work. However, the possibility of a more than one crystal structure is not included in the computational approach, described in the methodology section (5.5), as the calculations are not performed at a finite temperature that would enable the phase transitions to be represented and identified.

### 4.4.2 Rutile structural parameters

Since the rutile structure is taken to be the thermodynamically stable one for SnO$_2$ the structural parameters are described here. The unit cell is simple tetragonal with lattice parameters $a = b \neq c$, belonging to the space group $P42/mnm$ [143]. The unit cell contains four non-equivalent O and two non-equivalent Sn atoms. The Sn are located at the Wyckoff positions (0,0,0) and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, at the unit cell corner and at the centre, and there are two O atoms in the plane of the central Sn atom having the Wyckoff positions $(x + \frac{1}{2}, x + \frac{1}{2}, \frac{1}{2})$ and $(x + \frac{1}{2}, x + \frac{1}{2}, \frac{1}{2})$ and the other pair lying on $z$ faces of the cell having the positions $(x, x, 0)$ and $(\bar{x}, \bar{x}, 0)$ [143]. The $x, \bar{x}$ are two possible translation coordinates that will map an atom to its symmetry equivalent.
4.5 Lattice defects of rutile SnO$_2$

Within a bulk crystal solid, any change to the perfect periodic arrangements of the atoms in the lattice constitutes a defect, and a point defect is a change affecting only a single atomic site. The physics of the interactions between the defect and host crystal are studied in this work by DFT based methods to deduce the impact on conductivity (section (5.2)) and thermodynamic states (see section (4.6.3)). A defect is intrinsic if only the atomic species of the host are involved, and extrinsic if a non-host species is introduced either through doping or as an impurity. With regard to electronic properties, the significance of a defect state is introduced here in terms of providing additional charge carriers. As is discussed in section (4.6), semiconductor conductivity is dominated by acceptor (p-type, positive charge) or donor (n-type, negative charge) carriers, and so for n-type conductivity, one is generally interested in promoting additional charge carriers (electrons) while reducing the concentrations ('killing') of those that are acceptors (such as holes). For p-type conductivity, electron 'killers' and hole producers would be sought. SnO$_2$ is considered to an n-type material, since a p-type oxide is considered inherently difficult to realise due to the localisation of holes associated with oxygen 2$p$ energy states that results in a deep acceptor state in the band gap (see section (4.6)) [154], [155]. However, some work has explored the feasibility of inducing p-type conductivity using elements such as I [156] and N [157].

The point defects that are considered in this work are of two types.

**Vacancy** defects atomistically represent a deficit of one of the host atomic species. In the case of SnO$_2$, an oxygen vacancy acts as an n-type donor since by removing an O$^{2-}$ from the lattice two $e^{-}$ are left over from breaking the Sn=O covalent bond.

**Interstitial** defects are additional atoms at sites not occupied in the host crystal. A surplus Sn in SnO$_2$ acts as an n-type donor by having the same coordination as a host, so that electrons are left over when each bond to the extra Sn has formed equivalently to the perfect crystal.

4.5.1 Intrinsic defects of SnO$_2$

In terms of its electronic structure, SnO$_2$ is a direct gap semiconductor. Experimental investigations put the estimated band gap in the range 3.5eV to 3.8eV [144], [147], firmly in the category of what is considered a wide band gap semiconductor - in which the highest quantised energy state occupied, in the thermodynamic ground state of the system, is the valence band maximum (VBM). In spite of the large band gap, the intrinsic conductivity reported, that is, without any dopants or impurities, can be as high as 0.5 Sm$^{-1}$ (at room temperature) for thin films [144], which compares favourably to other n-type oxides with a wide band gap such as ZnO (0.15
Sm\(^{-1}\)) \cite{145} and TiO\(_2\) (10\(^{-5}\)Sm\(^{-1}\)) \cite{146}. Moreover the value can vary by orders of magnitude depending on the chemical environment used for crystal growth and processing, with a (temperature dependent) decrease accompanying an increase in oxygen partial pressure \cite{159}, \cite{160}. This characteristic is attributed to oxygen vacancies, since these have been detected in a high concentration in the material \cite{158} and would therefore contribute a diminishing amount of conductivity with increasing oxidation, their effect as electron donors also becoming outweighed at elevated temperatures by compensating acceptor levels forming near the VBM and increased hole transport \cite{161}. Kılıç and Zunger \cite{162} carried out a comprehensive theoretical investigation into the possible influence of Sn interstitials, motivated by the observation that SnO\(_2\) is capable of supporting high concentrations of intrinsic defects that make it off-stoichiometric, when many other materials with similar thermochemistry (CaO,MgO,SiO\(_2\)) are generally found to be stoichiometric. Combinations of tin interstitials and oxygen vacancies are stable, it is argued, because the relaxation of the lattice evenly affects all bond lengths at a given distance from the defect site, and the new bonds formed in the case of an Sn interstitial defect, hence minimising the lattice distortion. However this explanation proposes that Sn interstitials are the predominant source of n-type conductivity, which has been disputed by other researchers on the basis of the DFT functional used for the computations \cite{165}. It has also been proposed from first principles calculations \cite{163} that hydrogen (H) defects act as an n-type donor, an example of an impurity which can be incorporated into the electric structure as a function of the processing environment \cite{164}, although the stability of the defect states, with which the electron donation is associated, is not quantified, nor the populations of charge carriers expected under such a doping scenario. From DFT calculations with local functionals (LDA,GGA) \cite{165} it is asserted that it is not intrinsic defects themselves that are the origin of conductivity, but instead the interstitial H defect that can form in the rutile SnO\(_2\) lattice. However, because the interstitial hydrogen in SnO\(_2\) bonds to a host oxygen in the lattice, the formation energy of the defect and its stability may depend not only on the hydrogen present in the environment but also the oxygen chemical potential, which also governs the formation of the oxygen vacancy defects. The possibility that both occur in parallel makes it important to analyse oxygen vacancy defects in this work, which is done in section (6.1). Hydrogen defects in semiconductors are further discussed in section (4.5.5).

### 4.5.2 Extrinsic Dopants of SnO\(_2\)

Extrinsic defects potentially allow a greater degree of control than intrinsic defects over the behaviour of the charge carriers, by introducing both n-type and p-type dopants, independently varying the concentrations of both types. Here, many experimental reports can be found on a wide variety of candidates. One of the most widely used is Indium Tin Oxide (ITO) (In\(_2\)O\(_3\)-SnO\(_2\)), a proto-typical transparent conducting oxide (TCO) having the best available optical and electronic conductivities, though the falling global availability of In motivates a search for alternatives. In this case, an n-type dopant is provided by substituting Sn\(^{4+}\) for the I\(^{3+}\) in the host In\(_2\)O\(_3\); instead, SnO\(_2\) as host is more appealing because conductivity would presumably...
come from adding an increasing concentration of a dopant as a replacement for the low cost Sn. The most investigated direct dopants, which substitute for Sn at lattice sites, are F, As, Sb, Ta and Nb as n-type dopants, while In has been trialled as a p-type donor [156]. Flourine, F replaces a double negative charged ion $O^{2-}$ with its single negative ion, thus keeping one electron from participation in bonding, which results in a linear increase in the effective carrier concentration to 50% of the undoped concentrations for flourine concentrations up to 10%, together with a 150% increase in carrier mobility [166]. However, increasing carrier scattering with the number of F ions places a limit on usable doping concentration [167]. With some dopants, the efficacy as a donor depends on the charge state, as in the case of Sb$^{3+}$ replacing Sb$^{5+}$ as the predominant state beyond a threshold of increasing doping concentration [170]. In such cases the relative stability (formation energy) becomes important and a thermodynamic framework is required in which the concentrations of concurrent defects and their properties can be known as a function of the same thermodynamic conditions. For the cation dopants, in the case of Nb, Ta and Sb-doped thin films, an increase in conductivity is frequently observed for low doping concentrations, maximised anywhere in the range of around 1-10%. The data is summarised in table (2). For the case of Sb, the optimum doping concentration is very low with measured decreasing conductivity above 1%, as listed in table (2). For the same study, a significant increase in mobile charge carriers was realised with temperature rise from from zero to 100°C, yet no corresponding increase in conductivity, which activates within the higher range of 100°C – 400°C. The band gap increases over the same range of increasing Sb content up to around 6% beyond which it falls; this may therefore be the primary limiting factor on the mobile carrier concentration and the conductivity. Higher doping concentrations than 8% are likely not desirable due to the change in the predominant valence from Sb$^{5+}$ to Sb$^{3+}$ as observed by other researchers noted above. For Ta doped thin films, one study has reported very positive data on the conductivity, though it is interesting that as with the Sn case conductivity decreases for concentrations beyond the best reported of 3.75% Ta. While carrier concentration increases appreciably at low $T$ for undoped SnO$_2$, according to this investigation, the change is insignificant for the doped samples, while data for higher $T$ is not given in this study. A more pertinent experimental investigation produced both polycrystalline and epitaxially grown thin films [169]. Again a maximum in conductivity was seen for 5% Ta, with half an order of magnitude greater total conductivity for films grown on the (0001) surface of alumina. Temperature dependence of the properties was also assessed for the growth of the films on a rutile substrate (TiO$_2$). The resistivity behaviour had the signature of a weakly metallic system with a slow increase of resistance with $T$ towards room temperature, as in the first study. The mobility for the epitaxial film is around twice that of the rutile seeded growth one, on average over the conditions sampled, and carrier concentration is greater by a small factor. Both studies arrive at similar conclusions, namely that Ta-doped SnO$_2$ behaves as a degenerate semiconductor. The second study adds that the mobility appears to be limited by the Ta concentration, which is attributed to a greater degree of scattering of the ionized impurities, which is the expected result for n-type semiconductors with a high concentration of mobile charge carriers. It is unfortunate that data for higher temperatures is not given in either study, since the PEFC working temperature is closer to 80°C. The two studies together do however raise the importance of the structure of a doped oxide to making
it conductive. The first study highlights the growth direction of the films changing to different degrees with the orientation of the initial plane, which could suggest that the nucleation rate in the films at the low Ta concentrations has more influence than the ordering tendency of the crystal structure. The second study brought out the result of maximised conductivity for growth on an alumina surface, which possibly favours oxygen vacancies as contributors towards the $n$-type conductivity, given that a theoretical study [172] found formation energies of these vacancies varying significantly with the crystal direction, and as cited above the dependence on the oxygen partial pressure is well known.

### 4.5.3 Crystal Field Splitting (CFS)

A metal with $d$ electrons in the outer shell surrounded by anions brings with it an effect known as crystal field splitting. If the $d$ orbitals surround a metal cation, which is assumed to form electrostatic (ionic) interactions with the anion neighbours, then some of the orbitals will be directed between the anions as they approach to form bonds, whereas some will be directed parallel. This will cause the energy levels of one subset of the $d$ orbitals to be raised in energy and the other subset lowered. The sizes of the total energy gap is defined by the crystal field stabilisation energy ($\text{CFSE}$), $D_q$, which depends on the ionic radius of the cations [173]. While the value of the total energy difference $D_q$ can change with variations in the cation-anion distances, the relative energy separation of the two subsets is fixed by the coordination. In the case of Ta-doped SnO$_2$ considered in this work, Ta is octahedrally coordinated by six oxygen atoms. Two of the orbitals, the $d_{x^2-y^2}$ and $d_{z^2}$, are parallel to the Ta–O ligands. These states are denoted $e_g$, and their energy is $6D_q$ above the barycentre while the other three $d$ states $t_{2g}$ are directed between the ligands and lie $4D_q$ below the barycentre. The situation is illustrated in figure (4.11). The pairing energy $U$ relates to the spin states of the $d$ electrons, penalising occupation of an orbital by one spin-up and one spin-down electron together (just as in the DFT+U formulation of DFT). The relative strength of the pairing energy $U$ and the CFSE, $D_q$, determines whether it is energetically preferable for states to be occupied by two electrons of anti-parallel spin or by single electrons with a net spin of $S = \frac{1}{2}$, and the possible configurations resulting from the two situations are determined by Hund’s rules. These state

<table>
<thead>
<tr>
<th>Dopant</th>
<th>n/p-type</th>
<th>Band gap (eV)</th>
<th>Conductivity ($\text{S m}^{-1}$)</th>
<th>Electron mobility (\text{cm}^2/\text{Vs})</th>
<th>Carrier concentration ($\text{cm}^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sb (Sn$<em>x$Sb$</em>{1-x}$O$_2$) [170]</td>
<td>n</td>
<td>2.3-3.1, $z=0.01-0.08$</td>
<td>$9 \times 10^{20} - 2.71 \times 10^{21}$, $x=0.01-0.08$</td>
<td>$10^3 - 10^2$, $T=0°C - 400°C$</td>
<td>1.5-2.1 x 10$^{20}$</td>
</tr>
<tr>
<td>Ta (Ta$<em>x$Sn$</em>{1-x}$O$_2$) [168]</td>
<td>n</td>
<td>$10^9$, $x = 0.0375$</td>
<td>50 (epitaxial)</td>
<td>50 (epitaxial) ($x=0.0375$, 10°C)</td>
<td>10$^{20}$</td>
</tr>
<tr>
<td>or Al$_2$O$_3$-grown, $x=0.0375$</td>
<td>10 (epitaxial) ($x=0.0713$, 80°C)</td>
<td>$x = 0.0375$ or $x = 0.07135$</td>
<td>0.8 x 10$^{21}$ (epitaxial)</td>
<td>0.85 x 10$^{21}$ (polycrystalline)</td>
<td></td>
</tr>
<tr>
<td>Sb</td>
<td>n</td>
<td>3.94-3.71, $z=0.01-0.04$</td>
<td>$10^{11-9.4}$</td>
<td>$x = 0.02-0.04$</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Summary of the properties of extrinsic dopants in SnO$_2$. 

Matthew Worsdale
that the spin quantum number $S$ and the angular momentum quantum number $L$ take their maximum values, and that the total angular momentum $J$ must also take values consistent with the conservation of momentum depending on the direction in which spins $S$ are aligned. Two cases that must be considered are $U < |D_q|$ and $U > |D_q|$. Which one applies can depend on the arrangement of the ligands around the metal ion. In the case of FeO, pressure can induce a transition whereby the on site repulsion of $d$ electrons is reduced and a low spin state is adopted (the diamagnetic Fe$^{2+}$ with a fully occupied $d^6$ configuration) [174]. At lower pressures the dominance of Fe$^{3+}$ ions can give paramagnetic states with some amount of unpaired $d$ orbitals distributed between the $e_g$ and $t_{2g}$ levels whose configuration is determined by the value of $U$ that decreases with pressure. However, in the case of Ta ions in octahedral coordination, only one electron is left in the $5d$ orbital when the Ta$^{4+}$ oxidation state is adopted. Therefore there is only one possible electron configuration for the ion in this oxidation state, whereas the Ta$^{5+}$ state has no non-bonding electrons. The change in ionic radius between the Ta$^{4+}$/Ta$^{5+}$ states could affect the value of $D_q$, but there is only one configuration possible whatever the value of $U$.

### 4.5.4 Jahn-Teller (JT) effects in the solid state

The crystal field splitting produces two sets of degenerate $d$ states. In the solid state, each coordination polyhedra shares edges and/or vertices with its neighbours, and if the composition of the oxide changes by introducing more than one cation species, the cation radius may change on average throughout the crystal, affecting the value of the crystal field stabilisation energy $D_q$. The cation radius also depends on the oxidation state. If a cation has an unpaired electron in the electron configurations in one or more of its oxidation states, the electron configuration for the crystal may change, governed by Hund’s rules. The Jahn-Teller (JT) theorem states that any non-linear molecule with a spatially degenerate electronic ground state will undergo a geometrical distortion that lifts the degeneracy by lowering the overall energy of the system [175]. Thus, the JT effect makes the two energy levels non-degenerate, leading to a further energy gap between two of three $t_{2g}$ states and the third which is lowered in energy with respect to the other two. This can be expected to occur in an alloy system where a non JT active compound is progressively doped with JT active ions. With the case of octahedral coordination in the crystal, as the number of these ions increases, the size of the energy gap created in the $t_{2g}$ sub-shell can be expected to increase in size as a primary consequence, making one $d$ state lower in energy. For a JT active ion having more than one $d$ electron, there could also be changes to electron configuration of the $t_{2g}$ states; the level higher in energy $\delta_{2/3}$ (as demonstrated in figure (4.11)) could be fully occupied by two electrons of anti-parallel spin, or this level could be half filled by a single electron and the lower energy level $\delta_{1/3}$ also half occupied. A transition may occur from three electrons occupying the $\delta_{2/3}$ state to two electrons in $\delta_{2/3}$ and one electron in $\delta_{1/3}$ if the value of the total angular momentum quantum number $J$ can be reduced by the transfer of one electron to the lower state which has less angular momentum.
4.5 Lattice defects of rutile SnO$_2$

Matthew Worsdale

However, in the alloy system Ta$_x$Sn$_{1-x}$O$_2$, there is only one electron configuration associated with the Ta$^{4+}$ ion, so occupation of either state by the single electron gives the same value of $J$. However, occupying the lower energy state may depend on the relative number of sites adopting the Ta$^{4+}$ or Ta$^{5+}$ state, and the location of these sites, since the potential energy of a given Ta$^{4+}$,Ta$^{5+}$ pair depends on their separation.

The JT effect in the solid state can cause coupled distortions of the crystal structure due to the existence of phonons. The origin of the effect is the interaction of the electronic states with the phonons which can be modelled as two coupled harmonic oscillators [176]. The short range JT effects that are driven by orbital-orbital interactions of $d$ states through electron exchange can thus become coordinated by the propagation of phonons through the crystal. These distortions can become cooperative if the concentration of JT active sites in the crystal is sufficiently high at of the order of one per unit cell [177]. There are several ways in which this can occur. In all cases the mechanism responsible for JT is a symmetry breaking operation that may be accompanied by a phase transition. In some cases, a discontinuous transition of the entire crystal structure may occur, affecting a group of lattice sites. The set of crystalline orbitals having a given symmetry is reduced, such that electronic states that were previously equivalent become distinct. Another possibility is for the electronic orbitals themselves to distort continuously from their equilibrium positions about fixed atomic coordinates while maintaining a symmetry in a given phase, as in the case of the rotations of coordination octahedra of the rutile and CaCl$_2$-type crystal structures, discussed in section (4.4.1).

4.5.5 Hydrogen doping of semiconductors

Hydrogen can be incorporated in the structure of a bulk semiconductor as an impurity where it can form a number of defects of the host structure owing to its high reactivity and ability to adopt a number of charge states (H$^+$,H$^0$,H$^-$). Hydrogenic defects [179] have been identified as donors in elemental intrinsic semiconductors such as Ge and Si. For compound semiconductors, hydrogen can act as both a $p$-type and an $n$-type donor depending on the material, and the cases are discussed in section (4.6.1).

Hydrogen was originally proved to exist in Si on the basis of the deepening of an acceptor ground state when the growth environment was changed from a hydrogen atmosphere (of 1 atm) to a deuterium one, by Hall [181], [182]. In the case of Ge, the existence of H in the lattice can be demonstrated experimentally owing to H forming shallow donor or acceptor levels, allowing the transition of an electron from the corresponding state to be detected through photothermal ionisation spectroscopy (PTIS) as performed by Haller [180] and reproduced in figure (4.12). In that experiment, absorption of radiation of infra-red (IR) frequency promotes absorption of the phonon corresponding to the transfer of electron or hole to the valence or conduction band in the case of an acceptor or donor complex. In a sample of Ge containing substitutional impurities of Si, as well as of aluminium and boron. This experiment performed at two low temperatures of 4.2K and 8.0K gave observation of two acceptor level spectra $A_1$ and $A_2$ that are shown in the left side of figure (4.12). However their equivalence can be seen from the fact that they are
4.5 Lattice defects of rutile SnO$_2$

Matthew Worsdale

Figure 4.11: Crystal Field Splitting raises energy of two d states $e_g$ in an octahedrally coordinated system, lowering the energy of the other three $d$ states $t_{2g}$. Jahn Teller (JT) distortions introduce an additional fine structure to the d states, removing the degeneracy of the $t_{2g}$ states by splitting them into levels $\delta_{2/3}$ and $\delta_{1/3}$.

distinguished only by shifts of all the lines by the same energy. It was found, moreover, that the ratio of the intensities of equivalent lines in the two series was independent of temperature, so each equivalent line corresponds to the same electronic transitions of a shallow acceptor centre in Ge. However, the separation in energy between any two equivalent lines in one of the series taken at a series of temperatures is found to be well fitted by the Boltzmann factor with an energy $E = 1.1$meV, suggesting that while the dispersion of the series of energy levels changes with temperature, the two series $A_1$ and $A_2$ belong to the same set of states which are the bound states of a shallow acceptor state in Ge. The two ground state energy levels and their series of thermally excited states are therefore split by 1.1meV. These observations do not demonstrate the involvement of hydrogen, but its association with the electronic structure was inferred by further details.

The acceptor state was attributed to the interstitial Sn atom in the tetravalent elemental crystal Ge, since host electrons are taken to form covalent bonds with the substitutional atom, leaving a population of holes. A negatively charged hydrogen H$^-$ can occur at the Si site which is a positively charged centre in the system, but in the ground state the hydrogen electron is attracted to both the Si and H atom centres giving a neutral overall complex. In the excited states detected by PTIS, the degrees of freedom of the Si-H complex and unbound hole from the Si impurity and their interactions in an effective Hamiltonian $H_{Si-H}^{Ge}$ were shown to give tunnelling of the hydrogen atom around the Si center, separating two ground state levels equivalent by
symmetry whose spectra were observed in the PTIS experiment, with a splitting magnitude given in the model by the strength of the tunnelling. This study is therefore verification of the ability of hydrogen to create shallow donor and acceptor levels in elemental semiconductors by forming linear combinations of the host states. Despite being a localised impurity at low temperatures, the model based on $H_{Si-H}$ suggested H alters the crystal and electronic structure due to the symmetry of the bound states derived from the host orbitals. Accordingly the value of the tunnelling transition probability found in the study on this basis was high at around 0.8, which reflects the disperse wavefunction of the hydrogen nucleus that is able to exert a significant influence on the electronic wavefunction of the system.

This experimental study does not, however, detect absolute donor or acceptor levels by electronic transitions, a verification of the effect of H that has been reported more recently through muon spin rotation and relaxation spectroscopy ($\mu$SR) to observe the electro-paramagnetic spin resonance (EPR) of the hydrogen atom with an unpaired electron [202]. This experiment utilises the fact that hydrogen can be detected when in a paramagnetic state. However, the short lifetime in a semiconductor makes it necessary to simulate H through Muonium (Mu), a bound state of $\mu^+$ and $e^-$ that can represent a light isotope analogue of H.

In the experiment, samples of SnO$_2$ were bombarded by spin-polarized muon beams in a transverse applied magnetic field over a temperature range of $T = 5-110K$. The muon spin precession was examined through the positrons emitted during its decay. Two signals were detected, one corresponding to the precession of diamagnetic Mu$^+$ and another due to that of Mu$^0$ created through the binding of the muon to the electron, a paramagnetic state that precesses due to the hyperfine interaction at a different frequency to Mu$^0$. The precessions were used to fit the relative amplitudes of the diamagnetic and paramagnetic signals, allowing for the deviations due to the change in (Larmor) frequency [234]. The creation and detection of Mu$^0$ therefore is equivalent in this experiment to establishing a shallow donor state occupied by an unpaired electron, at a depth of less than 0.1 eV. Fitting the temperature dependence of the relative spin amplitudes with an ionization model [235], and comparing the results to a hydrogenic model of a localised impurity (that is, modelled as a nucleus with orbiting donor electrons to represent H as donor in the bulk semiconductor) gives good agreement for the spectrum of defect energy levels. Where H is not associated with a shallow donor level, however, spectroscopic techniques like ($\mu$SR) cannot be applied, motivating study of the electronic structure by DFT.

The theoretical arguments in favour of H as ready donor/acceptor can be heuristically summarised as the observed ease of formation of H defects which suggests low formation energies at low concentrations. Despite the small atomic radius of H it is nevertheless able to exert large influence due to the bonds it can form over large distances [233], and the different valence states, high diffusivity and the possibility to occupy different lattice sites simultaneously. In addition, H is argued to be mobile and possessing a low migration barrier in different charge states within a number of materials, behaviour which has also been directly observed [183], although H thermal trapping and increasing entropic losses with diffusion of H through the lattice also represent great potential limits on the conductivity induced at increasing temperatures. Keeping in mind the importance of the thermodynamic stability of the compound, migrating H that is highly reactive could be regarded as detrimental given the ability of H to form substitutional defects,
which involves creating O vacancies.

In cases such as interstitial hydrogen defects bonding to oxygen in SnO$_2$, induced states will compete thermodynamically with host defect states associated with the oxygen vacancies and perhaps those associated with other extrinsic dopants or defects. The properties can be expected to change with doping concentration and the chemical environment. Control over the induced electronic properties may be achievable by controlling the lattice sites where the defects occur, and hence the possibility may arise to control independently those defect states that promote thermodynamic stability, and those that create electronic conductivity.

In a compound semiconductor H can act as an amphoteric donor, i.e either p-type or n-type, and the discussion in section (4.6.1) forms the theory that must be applied to describe the electronic structure of H-doped compounds. The multiple valences and high reactivity can generally predispose the formation of bonds with all of the chemical species present in the case of a compound semiconductor, and in addition, H may be capable of interacting with other defects through the cooperative influence of these and itself on the material and electronic structure. For these reasons, in analysing the defect properties to suggest the likely position of the bands formed due to the bonding and interactions of H with the host it is important to understand all of the states present in the sample and the position of H due to its effect on the host structure. Hydrogen can have polar opposite effects on conductivity as a dopant- capable of passivating some systems that possess conductivity while activating and providing sources of conductiv-
4.6 N-Type and P-Type doping and the Fermi level

When a solid crystallises, electronic orbitals overlap as the atoms approach one another and bands are formed from the energy levels of the individual atoms, describing an energy state of all atoms in the solid state. For some ranges of energy, there is a gap where there are no energy levels. A semiconductor has a band gap between the highest band that is occupied by electrons (all the occupied bands known collectively as the valence band) and the remaining unoccupied bands (collectively, the conduction band), such that for the ground state of the system at $T \approx 0K$, the material is an electrical insulator. At higher temperatures, the electrons in the valence band acquire kinetic energy such that their maximum energy is increased from that of the highest permitted energy band, by amount called the Fermi energy $E_F$. The band gap $E_g = |E_C - E_V|$, where $E_C$ is the energy of the lowest state in the conduction band, and $E_V$ the energy of the highest state in the valence band, ranges from just greater than 0.0 to around 4 eV at room temperature [102]. In ab initio studies of electronic structure, the energies between which a band exists, and its shape, are consequences of the forms of the crystalline orbitals as they are represented by basis functions in DFT, as described in section (4.3.9). Semiconductors fall into two categories on the basis of their bandstructure; direct gap or indirect gap. In the former case the maximum of the valence band and the minimum of the conduction band coincide at the same value of the reciprocal space vector $\vec{k}$, so electron promotion is a straightforward excitation requiring energy $|E_g|$. An indirect band gap requires a transfer of crystal momentum equal to the difference $|\vec{k}_c - \vec{k}_v|$. The translation of electrons must be counteracted by a small change in crystal momentum in the opposite direction, in order to conserve the wavevector.

N and P-type doping can be achieved by introducing intrinsic defects (an increase or decrease in the atom species of the pure crystal) and extrinsic defects (introduction of a foreign atom species) or a combination of both. The distinction between N-type and P-type refers to the majority charge carrier involved. Within the quantum description of solids the concept of hole states necessarily arises to explain how electrons move. The bands containing electron states exist in regions of reciprocal space ($\vec{k}$) with boundaries defined by Bloch’s theorem [102]. Each band contains $2N$ independent states, where $N$ is the number of valence electrons in the
primary cell (containing one atom), due to the two possible values of electron spin. Whenever an electron makes a transfer between bands, each electron moves to occupy the next state in its band. If an electron is promoted from a band, a hole is left behind. When the next electron excitation happens, each successive electron moves into the hole, effectively creating a distinct quasiparticle state moving ‘backwards’, i.e. with crystal momentum opposite to that of the electrons, and opposite charge. In an intrinsic semiconductor with no doping the number of electrons \( n_e \) has to equal the number of holes \( n_h \). However, where additional atoms are introduced through doping, the populations of \( n_e \) and \( n_h \) can affect the \( n \)-type and \( p \)-type conductivity in different ways.

**N-Type** doping can come about with the addition of a group V element to a compound where the dominant constituent has four electrons in its outer shell and normally shares these valence electrons with four other atoms in covalent bonds to get a full shell. The group V element has five valence electrons and if introduced to the system will donate four of these to form bonds, leaving a spare \( e^- \). The spare electrons exist at an energy level \( E_D \) below the conduction band. This is illustrated in figure (4.13). With the increased concentration of \( e^- \) upon \( n \)-type doping, \( E_F \) also rises relative to its undoped value. The net result is an increased availability of electron states near the top of the band gap, that can be thermally (or optically) excited into the conduction band. Simple approximations for the new position of \( E_F \) can be derived by applying equations expressing conservation of charge and the number of dopant atoms

\[
\begin{align*}
N_e &= N_{D^+} + N_h \\
N_D &= N_{D^0} + N_{D^+}
\end{align*}
\]

(58)

where \( N_d \) is the total number of donors and \( N_d^+ \) and \( N_d^0 \) are respectively the ionised and non-ionised donors. This is combined with the assumption that ionized dopants obey an expression of the Maxwell Boltzmann form,

\[
N_d^+ = N_d \left( 1 - \frac{1}{e^{(E_D - E_F)/k_B T}} \right)
\]

(59)

and exploiting the limiting low \( T \) case when it is assumed that \( N_d^+ \gg n_h \) because only electrons from the donors have been ionised at that point. The details are omitted here since the electron concentration \( n_e \) is unknown and the equations (58) and (59) can be replaced in practical calculations by a method that takes into account the thermodynamic potentials of different states, discussed below in section (4.6.3)

**P-Type** doping introduces a concentration of holes into the valence band at energy \( E_A \) near to the VBM. This can be achieved through introducing a group III impurity atom to a system with atoms of valency four. In order to form bonds to the host atoms an additional electron is needed, taken from the host valence band states. This results in an unoccupied level near the VBM that can instead provide mobile holes to the valence band, locating the Fermi level closer to it, \( E_V < E_F < E_A \). The process occurs through the ‘sinking’ of the holes to the valence band rather than the rise of electrons to the conduction band [102]. Holes are replaced by electrons at each state in the band gap and the valence bands, following the argument given above. Thus electrical conductivity increases with an increased number of acceptor atoms \( N_A \).
4.6 N-Type and P-Type doping and the Fermi level

Matthew Worsdale

Figure 4.13: From left to right: The basic band structure of an intrinsic semiconductor with energy gap $E_g$ and fermi level $E_F$.
Band structure of a semiconductor with only n-type doping, with electrons at the donor level $E_D$ and a raised fermi level.
Band structure with only p-type doping, with holes at an acceptor level $E_A$ and a lowered fermi level.

and temperature $T$, brought about by liberating holes rather than electrons. See figure (4.13).
Whether a semiconductor shows N-Type or P-Type conductivity depends on the populations of defect states in addition to those of electrons and holes. Such defects may be of the intrinsic category, taking the form of imperfections of the ideal crystal structure such as vacancies of atoms at their regular sites, and interstitials where additional atoms exist at non-regular sites.
A third category is antisite defects, where in a binary or ternary crystal an atom of one species occupies a site normally taken by another species. Extrinsic defects fall into two main categories; interstitial and substitutional, referring respectively to atoms that are located at regular and irregular sites of the pure semiconductor, and in both cases that are different elements to those in the pure composition.
The effects on electrical conductivity due to doping are on $n$ and $p$-type charge-carrier concentration in the conduction and valence bands. As is discussed in section (4.6.3), all the processes that contribute to the conductivity can be understood thermodynamically by assuming Maxwell-Boltzmann statistics for the charge carrier concentrations in terms of the chemical potential $\mu$ of charge carriers, the position of the Fermi level and temperature $T$. Hence, the relative importance of processes such as ionisation and compensating re-absorption of charge carriers can be determined by knowing the populations of each particle involved as a function of the same thermodynamic conditions. Scattering of the charge carriers by ions impedes the transport of electrons in the solid and with ab initio approach would be treated by calculating the excited state of the system formed by the charge carrier and the target, which is a problem
Shallow and Deep donors.

It is also necessary to understand the redefinition of the bandstructure that may take place. The introduction of any defect breaks the periodic arrangements of charge of a perfect crystal solid. The resulting potential $V'$ experienced by other electrons consists of a long-range Coulomb contribution $V_r$ and, in the case where the defect is an impurity atom, a short-ranged part $V_R$ representing the effects of its core electrons [198]. The form of the two potentials in conjunction distinguishes two limiting cases of defect states which are induced in the bandstructure, since $V_r$ varies over the extent of the lattice while $V_R$ is confined to a unit cell. However, in practice $V_R$ is translated by most DFT codes when modelling a large cell as per the supercell method, which is described in the next section ($V_R^T$ becomes a periodic component) which produces an overall perturbing potential given by the variations due to the details of the host crystal (crystal potential) and the periodically varying impurity potential. The defect state is bound by the potential (both parts separately) as a function of $k$. Where $V_r$ is more significant, the binding potential places the eigenvalues that determine the points on the energy bands in consistency with the crystal field, which tends to produce states that are derived from host eigenfunctions of the DFT solution for the undoped system, physically because the potential receives only a small perturbation from the screened coulomb potential of the charged dopant atom [199]. A state bound primarily by $V_R^T$ is generally not very disperse over $k$-space since the potential is only effective over a lattice constant of distance and so is split from the main bands in the bandstructure inside a band gap. This is known as a deep level. Those defect states that are physically described more accurately by $V_r$ are frequently discussed in the analysis of the results of this work in terms of perturbed host states (PHS), whereas deep states ($V_R^T$) are discussed either by that name, or alternatively as ‘gap’ states.

4.6.1 Amphoteric doping and the charge neutrality level (CNL)

The modification to electronic structure due to a defect takes another level of complexity in cases where the dopant is neither $n$ or $p$-type. Here, H forms a bond with either the host cation or anion, resulting in the formation of a bonding state and an antibonding state. With both of these states, the $s$ state of hydrogen is expected to be hybridised with those of the host and will form a resonance to a different degree with the states of the host cation/anion to which it is bonded, some of which lie in the valence band and some in the conduction band, assuming that there is no over or under coordination in the host structure.

The construction of the H orbitals has been linked to the properties of defect states. For example, multi-vacancy defect states that can support configurations of H in the Si lattice (for example, Si-H, Si-H$_2$) were detected on the basis of vibrational modes of the bonds [200] whose frequencies were found to depend on the amount of $s$- and $p$- mixing [201].

\[ ^1 \text{Muonium was used to model H due to the experimental difficulties detecting it on account of the very short lifetime in a semiconductor [202].} \]
when the host bulk semiconductor is made inhomogeneous by a defect, surface or interface. A state in the band gap is formed to maintain continuity of the energy state at the interface within the bulk semiconductor host. The position of these states in the band gap is related to a concept of charge neutrality level (CNL). In the perfect semiconductor charge neutrality occurs if the valence band is completely filled and the conduction band completely empty. The gap state has mixed conduction and valence band character, being composed of a weighted average of host valence and conduction bands in proportion to the position in the band gap. Occupation of these states therefore gives a degree of charge imbalance, \( i.e., \) a local deficit if the state is unoccupied, or excess if it is occupied. This governs the scale of charge neutrality over the bulk material. The prevailing behaviour observed for hydrogen in a compound semiconductor, on the basis of theoretical studies to date, is an ability to form either acceptor levels or donor levels, depending on whether it bonds to the host cation or anion respectively [204]. The strength of the cation and anion bonds determine the position of the acceptor and donor levels. The charge state, \( \text{H}^+ \) or \( \text{H}^- \), depends on the position of the fermi level \( E_F \) in the band gap as given by equation (69). At a unique position in the band gap, the charge neutrality level (CNL), the formation energies of the defect in the two charge states will be the same. This level is well defined, \( i.e. \) 'pinned' within a small range of band gap energy by the dipole interaction between charges of opposite sign at the two sides of the interface [203]. In an ideal material corresponding to this picture where the interaction of hydrogen is the same for any anion or cation, the relative positions of the \( \text{H}^+ \) energy band in the conduction band and the \( \text{H}^- \) energy band in the valence band are the same, determined only by gradient of the energy with respect to the position of the Fermi level. This is shown schematically in figure (4.14). The energy difference between \( E_F \) and the formation energies of \( H \) acceptor (donor) level in or near the edge of the valence band (conduction band) is the same, with \( E_F \) adjusted between the two cases. This means that the energy needed to move an electron from \( E_F \) to the acceptor level is equal to that needed to remove an electron from the donor level to \( E_F \) (thus forming an unoccupied bonding state). Thus the creation of the donor/acceptor levels and the bonding which gives rise to the \( \text{H}^- \) and \( \text{H}^+ \) are two parts of a single process that is characterised by the alignment of \( E_F \) when the \( \text{H} \) defect is created. Therefore, both charge states \( \text{H}^- \), \( \text{H}^+ \) can be stable depending on the position of the CNL and the defect type, substitutional or interstitial, can act as either a donor or acceptor state depending on the prevailing conductivity. However, the CNL does not necessarily lie in the band gap but can instead fall inside the conduction band [202].

In addition, it is possible that \( \text{SnO}_2 \) with hydrogen defects forms an ionic system, where the orbitals of hydrogen are not shared with those of the host. In this case hydrogen would represent a formally charged defect in either the \( \text{H}^+ \) or \( \text{H}^- \) charge states. The methods for the study of charged point defects could therefore be relevant to hydrogen, which are evaluated in section (5.3.3), as is the bonding, which can be studied in a first principles way through the electronic structure predicted by DFT, and the mixing of the atomic orbitals of hydrogen and the bonding atom (as explained in section (4.3.10)). Furthermore, when the position of the VBM or CBM changes, the question arises of whether the bands that consequently lie within the VB or CB should be re-expressed in terms of the CBM/VBM, or conversely, whether the CBM/VBM
states should be re-expressed in terms of the host orbitals. This ambiguity can result in a
degree of uncertainty on the estimation of the band gap [188]. However, it is assumed that the
Fermi level and the position of the energy bands and their dispersion are influenced by factors
that have different relative importance depending on the defect modelled, as explained in the
following section.

\[ \text{Figure 4.14: Schematic illustration of the charge neutrality (CNL) level in the presence of}
\text{amphoteric doping in an extrinsic semiconductor.} \]

4.6.2 Band filling effects

The charge carrier statistics of electron donors and acceptors can depend appreciably on the
change in the band gap \( \Delta E_g \) brought about through the effect of defect states on the host charge
density which brings into play band filling effects (such as Moss-Burstein [230]) at increasing
concentrations, depending on the electrons per additional atom varying as a quantity over
different symmetry points of the Brilloin zone. In this case, one must compare the quasiparticle
gap defined as the difference between ionization potential and electron affinity in the ground
state, with the BZ average of single-particle Kohn-Sham (KS) energies [199]. This approach
is often necessary in defect calculations because of the incorrect calculation of the electronic
structure around the band gap attributed to the underestimation of this quantity using LDA or
GGA functionals on the host-defect system. Potentially significant in this work is the disperse
conduction band edge of SnO\(_2\), in which the location of acceptance of carriers can vary over a
range of about 1 eV. This can correspond to an order of magnitude change in carrier density for
nominal concentrations of the order of \( 10^{19} \text{ cm}^{-3} \), which could make it important to compare the
carrier concentrations for the cases where defect states are and are not activated, as a function
of the Ta doping concentration, in order to decouple the dependence of the conduction band
minimum and the defect level positions on the electron concentration.

68
4.6.3 Defect energetics and thermodynamics

The coexistence of populations of $n$ and $p$-type donors which are connected through the Fermi level $E_F$, and the case of amphoteric doping makes it necessary to consider the defect problem from a unified thermodynamic angle. In the general case, the activation of defects increases with temperature $T$. In addition to the statistical sampling of states of a system with given defects, as a secondary effect there can arise in the material concentrations of electron and holes, which will have some dependence on the temperature and also the concentration of extrinsic defects that have been introduced in the material. To understand defect concentrations quantitatively, one needs an understanding of all the mechanisms by which charge carriers are absorbed and excited and those by which the concentration of intrinsic defects changes. First it is needed to consider how the formation energy of an arbitrary defect varies as the concentration of the defect in the crystal increases, thereby increasing its disorder. At a finite temperature, there is an increasing probability that a crystal will move towards a system disordered by the appearance of intrinsic defects, as there is in general a large number of configurations of the atoms in the crystal that give rise to one defect structure. In the simplest example, if there is one irreducible crystallographic sites $d$ of multiplicity $N$, the total number of arrangements of the all the sites is $N!$. If there are two possibilities for the occupation of $d$, the number of arrangements of one composition is equal to the product of the arrangements of the number of sites of one of the species, $n!$, multiplied by the number of arrangements of the sites occupied by the other species, $(N-n)!$. The number of arrangements of the composition must therefore be equal to the ratio of the total multiplicity of the system and that for the composition $n$. The multiplicity of the configurations of a given composition, the statistical mechanical macrostate of the system, is the number of microstates $\Omega$ having associated entropy $S = k_B \ln \Omega$. This together with the relation $\Delta G = H - T \Delta S$ implies that [198]

$$\Delta G = nH_f - k_B T \ln \left( \frac{N!}{n!(N-n)!} \right)$$

(60)

where $\Delta G$ is the change in free energy $H_f$ is the formation enthalpy of the substitution between the two species at a given site at constant pressure. This equation expresses the balance of free energy between a site substitution and the entropy of disorder in the system associated with the combinatorial distributions that can be produced from the substitution. The form of the second term is valid only where the occupation of each crystallographic site is binary, and equally when there is one irreducible crystallographic site. For other cases, the number of configurations will still be related to $N$, though the number of defects that are not independent but equivalent by symmetry may be higher (so $n$ is reduced). In addition, it is important to understand the processes that occur when the extrinsic defect is stabilised in the host by changes in the geometry from the perfect crystallographic position where it is introduced. The processes involved can all be understood in terms of the relative predominance of different charge carriers by finding expressions for their concentrations as functions of thermodynamic variables $\mu$, $T$ and $E_F$ [205], [206], [207]. Then the question of which carrier production process is compensated is addressed, because they are specified by a particular thermodynamic state of the system.
5 METHODOLOGY

First principles calculations are important to extending the methodology used to search for desirable dopants to form the basis of an alternative support material for the PEFC. In recent years there have been advancements in numerical approximations to solution of the many body Schrödinger equation that can now provide accurate ground state properties of solid state crystals that are likely to enter into consideration. Density Functional Theory (DFT) has emerged as a highly accurate approximation of the electronic structure and the energetics of electronic interactions at the scale relevant for crystalline solids and their defects. Not only is it possible to quantify the individual energy of particular defined defects in the crystal lattice and to calculate a wide variety of candidate dopants swiftly in a high throughput way, thanks to the configuration of DFT programs for supercomputing platforms, but performing calculations on systems of varying sizes with choice over the constraints on the crystal structure permits determination of the importance of atomistic interactions over varying length scales and hence it is possible to extract the role played by structure from progressive doping of the lattice, which may be an essential consideration in selection of the appropriate host material, as discussed above. The scope of this work is limited to ground states of electrons in a mean field picture; full theoretical treatment would require a detailed understanding of electron transport mechanisms in the electrode fine structure.

5.1 Properties of a replacement support material studied by ab initio methods

Throughout this work, materials are evaluated based on the bulk properties of the material from which a candidate replacement catalyst support would be fabricated, using DFT calculations. It should be remarked that in corrosion science, the processes involved are often dependent on the surface of the bulk material in contact with the medium making them occur. However, such surface phenomena are not treated in this work as the scope is limited to analysis of bulk properties (primarily, conductivity and thermodynamic stability) and prediction of potential as a stable electrocatalyst due to these properties. Conductivity is expressed in the standard units of Siemens per metre (Sm$^{-1}$), while formation energies of appropriate materials, modelled as bulk systems, are expressed in electron-volts (eV) per formula unit of the material.

The most fundamental properties that a suitable replacement electrode support material must demonstrate can be summarised at this point as the following,

- High electrical conductivity
- Resistance to PEFC operating conditions (thermodynamic or at least high kinetic stability for cell potentials up to 1.5V)
- Low cost. Rare earth metals should demonstrate preferable conductivity but are pro-
hibitively expensive in anything but the smallest quantities.

- High porosity (and surface area), to maximise electron transfer by maximising ECSA.

The essential physical characteristics of the support material can be reduced to

- Compound must be thermodynamically or at least kinetically stable
- Compound must be a conductor or a semiconductor with the potential to increase conductivity.

### 5.2 Ab initio estimation of conductivity - Drude model

It is assumed that the conductivity of the bulk material will originate from the excitations of electrons from the valence band to the conduction band in the case of a semiconductor and the transport of these electrons in the conduction band. This is due to the propagation of quasiparticles, the electrons and holes, through the lattice under an applied electric field, which the catalyst support will experience when under load in the PEFC. The current $\mathbf{j}$ resulting from the quasiparticle propagation in the field $\mathbf{E}$ is then $\mathbf{j} = \sigma \mathbf{E}$, where $\sigma$ is the conductivity to be determined. As seen in section (4.6), the electrons and holes move within energy states of the solid $\epsilon_k$ that are determined by the potential they experience in the lattice, due to the external potential (atomic nuclei) and the many body electron interactions, that are expressed in the Kohn-Sham equations of DFT (38). If the electrons are sufficiently weakly bound by the potential, their motion may be approximated to take the form of travelling waves through the lattice, behaving as 'nearly free' particles in a quantum mechanical sense [102]. In this approximation the states $\epsilon_k$ are very delocalised so that the interactions between electrons are negligible, and the current formed by the electron transport in this regime is calculated by the Drude model [189],

$$\mathbf{j} = \frac{e}{V_{\text{sample}}} \sum_k \mathbf{v}_k F(k)$$

where $V_{\text{sample}}$ is the sample volume, $F(k)$ is the occupation of the quasiparticle state $k$ at a given value of the reciprocal lattice vector $\mathbf{k}$ as a function of quasiparticle concentration $n$ and $\sigma$, and having energy $\epsilon_k$. The group velocity $\mathbf{v}_k = \frac{\partial \epsilon_k}{\partial \mathbf{k}}$ is the velocity of the wave packet formed from the superposition of travelling wave states. The Drude model assumes that the transport of electrons and holes is dominated by their short ranged interactions with the atomic nuclei, which are governed by the scattering time $\tau$. If this is assumed to be independent of $k$, the conductivity $\sigma$ is approximated by [190]

$$\sigma = \frac{n}{m_{\text{eff}},k} e^2 \tau$$
where the effective mass of electrons or holes in the band as a function of their concentration \( x \), \( m_{\text{eff},k}(x) \) is related to the curvature of the band,

\[
m_{\text{eff}}(k) = \left( \frac{1}{\hbar^2} \frac{\partial^2 E(k)}{\partial k^2} \right)^{-1}
\]

(63)

where notation is switched \( \epsilon \rightarrow E \), to emphasise that this quantity is readily attainable from DFT calculations by fitting the conduction band to an analytic function \( \epsilon'_k \) in the neighbourhood of the minimum, \( i.e. \) in the range that is occupied by the position of \( E_F \), a good approximation for nearly free electrons. The expression for conductivity can be written in terms of a related quantity, the electron/hole mobility,

\[
\mu_{e,h} = \frac{e}{m_{\text{eff}}} \tau
\]

(64)

whence

\[
\sigma = ne\mu_{e,h}
\]

(65)

In addition, the same approximation implies that the charge carrier concentration \( n \) is defined by the probability distribution of the states occupied by electrons (half-integer spin particles) occuring in the conduction band according to the density of states as

\[
n_{e-} = \int_{E_1}^{E_2} \int dE_F(n \epsilon) g_{\text{DOS}}(E)
\]

where the density of states \( g_{\text{DOS}}(E) \) is available at a given \( E \) without approximation from the DFT code.

The electron/hole lifetimes in the material depend on the scattering rate, which is not so simple to estimate with DFT. The charge carriers scatter with impurities, with phonons and with eachother at rates \( \frac{\hbar}{\tau_{\text{imp}}} \), \( \frac{\hbar}{\tau_{\text{ep}}} \) and \( \frac{\hbar}{\tau_C} \). These three contributions scale as

\[
\frac{\hbar}{\tau_{\text{imp}}} \sim n_{\text{imp}} E_F
\]

\[
\frac{\hbar}{\tau_{\text{ep}}} \sim k_B T
\]

\[
\frac{\hbar}{\tau_C} \sim \frac{(k_B T)^2}{E_F}
\]

(67)

where \( n_{\text{imp}} \) is the concentration of impurity atoms. The three terms are additive contributions to the value of the total \( \tau \),

\[
\frac{1}{\tau} = \frac{1}{\tau_{\text{imp}}} + \frac{1}{\tau_{\text{ep}}} + \frac{1}{\tau_C}
\]

(68)

In the low temperature limit, there are negligible lattice vibrations to permit the propagation of phonons through the material so that electron phonon scattering is dominated by the Coulomb and impurity scattering. Because both these contributions depend on \( E_F \), the relative magnitude of the two depends on the system studied. It is assumed, however, that the variation in \( n \) and \( m_{\text{eff},k} \) with \( E_F \) and \( T \) is much more significant than variations in \( \tau \) under the low temperature, dilute donor conditions under which we are interested in estimating conductivity. The conductivities of Ta/Sn rutile alloys based on the results are discussed in sections (6.2.8) and that due to hydrogen defects is discussed in section (6.3.5).
5.3 Ab Initio Modelling of Defect Structures

The electrical conductivity of a semiconductor can be increased through doping to give a majority of $n$ or $p$-type carriers. This follows from the band description of solids and the fact that for a semiconductor the band gap between the valence band and conduction bands exists but is small enough for electronic excitations to take place with a probability $\sim e^{E_g/k_B T}$ where $E_g$ is the band gap. Introducing atoms that donate, or accept electrons can introduce energy levels in the band gap and affect its size and the position of the Fermi level $E_F$ within it.

With advancements in DFT calculations it has become possible to calculate the formation energies of defects from first-principles, and to obtain estimates of the doping limits and the corresponding bounds to the possible concentrations of charge carriers. This section introduces the methodology behind calculations on defect structures used to provide an idea of the feasibility of doping and discusses an informative description of the mechanisms by which defect states are incorporated and contribute to electron transport that can be obtained from thermodynamic models.

5.3.1 Defect energetics and scaling

The conservation of Gibbs Free energy before and after the creation of a defect implies the expression for the formation energy of a defect relative to the reference states of the constituent host and defect species [191]

$$\Delta G_f = G_{T}^{\text{def},q} - G_{T}^{\text{perf}} - \sum \Delta n_i \mu_i + x \mu_e$$

(69)

where $G_{T}^{\text{def},q}$ and $G_{T}^{\text{perf}}$ are respectively the total Gibbs free energies per formula unit of the system with the defect in charge state $q$, and the perfect system, $\mu_i$ is the chemical potential of the atom(s) removed or added to create the defect in number $\Delta n_i$, where $\Delta n_i$ is positive for addition and negative for removal, and $\mu_e$ is the electron chemical potential, where $x = \Delta n_e$. It is defined by the position of the Fermi level $E_F$ which takes values between the energy of the valence band maximum in the perfect crystal and the band gap $E_g$, so that $\mu_e = E_{\text{VB M}} + E_F$, [192], [196] where VBM stands for Valence Band Maximum. The electrons are thus generally considered to be taken from the valence band to $E_F$ in order to occupy the defect states.

In the production of a chemical compound, with or without impurities, a change will occur in the chemical potential $\Delta \mu$ of the constituent atoms from the values in their naturally occurring states. This change is determined by the experimental conditions under which the material is prepared. For a thermodynamically stable compound $A_xB_y$ produced from annealing a metal $A$ with another gaseous species $B$, the chemical potential of $A$ and $B$ cannot be higher in the crystal than in the elemental reference states, and therefore the inequalities must hold

$$\mu_B \leq \mu_{B(g)}(p_B, T), \quad \mu_A \leq \mu_{A(s)}$$

(70)

where $A(s)$ is the free energy per atom of the solid metal in a given crystal structure. The relevant experimental parameters are the partial pressure $p_B$ and $T$. The formation energy of
the crystal varies as a function of the pressure and temperature dependent chemical potentials of the reference states,

\[ \Delta G_{f}^{A_xB_y}(p_B, T) = \mu_{A_xB_y} - x\mu_{A(s)} - y\mu_{B(g)}(p_B, T) \] (71)

where it is assumed that there is negligible variation of \( \mu_{A_xB_y} \) with \( p \) and \( T \). Additionally, under equilibrium conditions in the compound formed we have that \( \mu_{A_xB_y} = x\mu_{A} + y\mu_{B} \). Combining (70) and (71) and applying conservation of Gibbs Free Energy leads to the inequalities [192]

\[ \frac{1}{x} \Delta G_{f}^{A_xB_y} \leq \mu_A - \mu_{A(s)} \leq 0 \]
\[ \frac{1}{y} \Delta G_{f}^{A_xB_y} \leq \mu_B - \mu_{B(g)}(p_B, T) \leq 0 \] (72)

giving a range of values for the chemical potentials of the constituent ions \( A^{c+} \) and \( B^{d-} \), assuming there is only one possibility for oxidation state \( c \) and anion charge state \( d \). This gives a range of the reference states \( \mu \) for the atoms in the defect calculation. However, the term involving \( \mu \) in (69) does not depend on the supercell size and therefore does not have a bearing on the scaling of the formation energies. On the basis of equation (69), the formation energy \( \Delta G_{f} \) per formula unit can therefore be calculated for a series of supercells of increasing size.

**Charged defects** For the results reported in the first part of this work on intrinsic defects and extrinsic doping of SnO\(_2\), where one has a localized source of ionic (point) charge, neutrality is ensured by redistributing electrons displaced from the introduction of the defect explicitly throughout the unit cell. In the final set of results on hydrogen as a dopant (impurity), an approach additionally used is to calculate the charged H in the presence of a compensating neutral (Jellium) background charge. As discussed in section (5.3.3), the formation energy of a charged defect is infinite in a formal sense. With explicit electron redistribution there is local charge neutrality within the unit cell, whereas the charged cell is strictly non neutral locally regardless of the placement of the defect. The difference between the total energy of the two systems can be expressed in terms of the alignment of the Fermi level. However, as explained in section (5.3.3), the charged defect of an isolated defect modelled by any periodic method represents a source of finite size error. The difference between the situations of charged defect + jellium background and redistributed charge may be of critical importance in the case of modelling an H\(^+\) defect. Regions of charge non-neutrality on the local scale depend on the occupation of the band gap states which are of differing conduction and valence band character (in general, in a fashion related to their position in the gap). The two modelling approaches could represent the local charge distribution differently, affecting the accuracy of the formation energy of the defect as a function of concentration (size of the system modelled).

### 5.3.2 The Supercell method

When a defect is created in a bulk system, there is an energetic cost associated with creating a ‘stress field’ due to the perturbation of the host atoms from the equilibrium positions they adopt
5.3 Ab Initio Modelling of Defect Structures

Matthew Worsdale

in the perfect crystal. Most DFT codes calculate the total energy of such a system in two stages; firstly, an energy calculation based on the perfect system, and secondly, the energy change in relaxing the structure with the defect to the equilibrium position. The main computational challenge that must be tackled is the loss of the periodicity of the perfect system due to the defect. Since DFT codes generally use periodic boundary conditions for crystals, the system that is studied is effectively an infinite one, however small or large the bulk system is modelled to be [209]. The bulk is approximated by the use of a supercell, essentially a larger region defined by a translation matrix operating on the lattice vectors of the unit cell, containing a single defect. Whatever region is used, it is essentially repeated by the basis functions over all space. For a charged defect, this has the unfortunate consequence that the defect interacts with copies of itself over long distances due to the Coulomb interaction. Various remedies exist to treat this problem, which are introduced below in section (5.3.3). From first principles considerations, the scaling of the defect formation energy $\Delta G_f$ should display some fundamental trends, irrespective of the type considered (charged or neutral, extrinsic or intrinsic):

- **Convergence** The defect formation energy should converge with the supercell size, whether it is a decreasing or increasing quantity. There is a simple physical reason for demanding this trend. A stress field can be defined representing the degree of mechanical stress due to the change of the equilibrium lattice, with a value associated to each spatial point. As defects are brought close together, their stress fields interact. But in the dilute limit, the magnitude of the field at a point far away from the defect becomes small and there is an increasingly small energy change associated with the displacement of atoms far from the defect site, so $\Delta E$ should be decreasing as the distance from defect $L \to \infty$. For the case of a charged defect, there will be electrostatic interactions between the defect and its images, which are created by the periodic functions used in DFT. These result in significant errors on the calculated values, and correction methods need to be employed to recover the correct physical trends (next section).

- **Range of defect formation energies** The range of values for the formation energies, from the value corresponding to the smallest supercell to the value corresponding to the bulk limit, cannot be too great relative to the size of the band gap for the material, calculated with the same DFT code and within the same approximation regime. As the supercell size approaches the maximum limit used in the computation, the method should become an accurate description of the true defect system, and hence changing the supercell size increasingly corresponds to changing the defect concentrations in the bulk. Hence the range of formation energies cannot be of the order of magnitude as the band gap.

**Supercells used for calculations** Following the method described in (5.3.2), supercells were generated by applying a transformation matrix to the lattice vectors. The most important consideration behind the choice of the cells was that they should be geometrically related to each other so that the differences in size between them are fairly regular. This naturally helps
the scaling analysis. For most of the cells two types of matrices, A and B were used,

\[
A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\]

(73)

A list of all the supercells used is given by table (3) and two different profiles in the crystal lattice are illustrated in figure (5.1). As the figures show, the supercells vary in the thickness in the z direction. To ensure that the defect considered in each supercell was equivalent to all the others, one of the O atoms on the high and low z-edges was chosen. The use of a translation of the lattice vectors guarantees that the atomic positions in supercells have the same symmetries. Since the total energy depends only on the absolute distance between atoms, it is therefore fixed by the unit cell and translations of it as long each resulting supercell contains an integer number of formula units of SnO$_2$.

5.3.3 Correction Methods

Errors in the scaling of defect formation energies arise from a number of sources and at this stage a complete account is not presented since it is expected to be possible to describe the errors accurately when a complete study of defects for a particular system has been undertaken, including all types of defects and a complete scaling analysis using the supercell method. At that stage it would be more appropriate to conduct a full review of sources of error in supercell calculations. However, to hopefully provide a context for the results obtained so far in section (5.7) here is briefly discussed some correction methods used to remove the most important sources of error.

**Finite size scaling** This is a scheme used to treat errors on defect properties that have become dependent on the supercell size $L$, which is a consequence of the limitations of applying DFT codes with periodic boundary conditions. Essentially, because the supercell is only an approximation to the bulk system, only in the limit of an infinite-sized supercell will the true defect system and its properties be recovered. As discussed above, the introduction of a defect forces the displacement of neighbouring atoms from their equilibrium positions in the perfect lattice. An error on the total energy after relaxation of the supercell appears that is related to supercell size, because it is related to the magnitude of displacement from the equilibrium crystal structure that the perfect supercells possess. For each supercell, the degree of perturbation of the perfect cell due to the displacements and subsequent relaxation is slightly different. This error is only removed in the limit of an isolated defect, when the relaxations, which take place over the whole system, are not truncated. However, since all supercells have the same symmetry we expect there to be a relationship between the size and the magnitude of error. This is the motivation behind so-called finite size scaling. With this approach, the defect formation energy is calculated in several different sized cells of identical symmetry and a fit performed as a function of inverse supercell size [210]. By extrapolating the fitted function to zero inverse lattice
Figure 5.1: Top: The $xy$-plane of the SnO$_2$ supercell as referred to in the text. The reciprocal lattice vector $K_2$ shown on figure (6.1) is out of the page. Bottom: View of the $z$-layering of the supercell layers.
constant we obtain the formation energy in the limit of a bulk material (isolated defect). For neutral defects this strategy should significantly reduce the finite size error.

**Electrostatic interactions.** For charged defects the situation is more complex. The physical reason for the relaxations is the long-range electrostatic interactions between ions, but because of the periodic boundary conditions used in the computation a charged defect also ‘sees’ images of itself throughout space, producing spurious self-interactions. This constitutes an additional source of error on the scaling of formation energies that has to be distinguished from the simple, the elastic relaxations described above. These, however, will be particularly affected by the defects and can be expected to scale with the electrostatic defect-defect interactions [192], since the defect charges, unlike the other ions, are not screened by electron charge to a significant extent. In addition to the self-interaction of the defect, its presence introduces an \( L \)-dependent variation in the charge density between different supercells [211]. The electrostatic potential at a reference point \( r' \) due to a charge distribution at \( r \) will have contributions from the different moments of the charge- monopole, dipole, quadrupole terms that scale as \( 1/L, 1/L^2, 1/L^4 \) and so on. Since we are particularly interested in the limit of the isolated defect, we note the Markov-Payne formula for the finite-size corrected total electrostatic energy for a periodic system with aperiodic charge density [211]

\[
E = E_0 - \frac{q^2 \alpha}{2L\epsilon} - \frac{2\pi qQ}{3L^3\epsilon} + O(L^{-5})
\]  

(74)

Here \( E_0 \) is the energy of the isolated defect. The Madelung constant \( \alpha \) is defined as

\[
\alpha = \sum_j \frac{z_j}{r_{ij}/r_0}
\]  

(75)
in terms of the distance $r_{ij}$ between an atom at $r_i$ and its $z_j$ nearest neighbours which can for a perfect crystal be expressed in terms of a constant distance $r_0$ \cite{102}. The subscript $i$ labels the constant associated with each species for a particular ion site $i$ such that the electrostatic potential at $r_i$ is periodic with $r_0$,

$$V_i = \frac{e}{4\pi\epsilon_0 r_0}$$

(76)

For a supercell with an infinite number of pairwise contributions $i, j$, the associated Madelung energy, the second term in (74), will become infinite. To counteract the isolated charge of the defect, the usual strategy is to make the system equivalent to a periodic array of point charges $q$ that are neutralised by a uniform counter charge, that have an effective dielectric constant $\epsilon$ so that in the large $L$-limit the system can be described approximately by the expansion given by (74). \cite{214}, \cite{210}. In this scenario the deviation from $E_0$ for a supercell of size $L$ is modelled as being caused by the polarization of charges in response to the aperiodicity of the charge density, resulting in the effective dielectric constant departing from the $L \to \infty$ limit where this quantity equals $\epsilon$. By fitting the formation energies as a function of $L$ to equation (74) the constant $\epsilon$ can be determined, assuming that $\alpha$ is known. This is a property of the space lattice and is calculated by the Ewald summation method \cite{214}. This method works by apportioning the aperiodic charge into shells containing amounts of the positive and negative ions such that the shell is neutral. The shells are chosen in order to make the sum over the pairwise interactions between ions, depending on the lattice constant $r_0$, absolutely convergent, in other words not dependent on the order of the summation of the terms in the expansion over the lattice vectors. This can be done by multiplying each term in the sum by a dimensionless factor \cite{212} to govern the relative magnitudes of the contributions from the different cells decaying as a function of distance.

The Makov-Payne method and extensions are intended for the limiting case of defects giving localised point-like charges. At the other limit, a defect creates a delocalised state with a defect charge distribution dispersed over the lattice. In this case, the aperiodic charge density is more appropriately treated by calculating the perturbed external potential $v_{ext}(r)$ directly. The simplest approach is to 'realign' the potential by calculating the difference between the charged defect and the perfect host, where $v$ is obtained by averaging a reference electrostatic potential averaged over atomic spheres, calculated at sites far from the defect \cite{213}. However, this method only suffices when the defect charge is confined to a single site. In a more ionic system such as a charged alloy, there is a deviation of the atomic pseudopotential from the bare Coulomb potential which causes a shift in the average potential over the volume $V$ \cite{199}

$$\Delta V_{PS} = \frac{1}{V_{cell}} \sum_i \int_V d^3r \left[ V_{PS}^i(r) + Z_i \right]$$

(77)

where $V_{cell}$ is the volume of the unit cell, $V_{PS}^i(r)$ is the local part of the atomic pseudopotential. The supercell energies need to be compensated by a corresponding energy correction $E = \Delta V_{PS} \sum_i Z_i$. This is approach works well for localised states in semiconductors, though is not so applicable to ionic insulators \cite{192}, but systems based on SnO$_2$ in this work are semiconductors (in the case of extrinsic $n$-type donors) or are metallic (in the case of hydrogen doping).
5.4 Charge Difference Maps

To gain an idea of how the defect impacts the electronic structure, it is useful to study the charge distribution process, which can be isolated from the formation of bands arising from the atomic relaxations. This separation can be achieved by studying the charge density through a charge difference map, and is useful in the dilute limit where the influence of a defect on the host can be isolated from the collective interactions in higher concentrations between dopants and the lattice. The difference map is introduced by calculating static atomic configurations that result after the defect has been placed, but with the final defect site restored to that of the perfect host,

$$\rho_{D+H} - \rho_{(H+D)-D}$$

where $\rho_{D+H}$ is the charge density of the final system with the defect after geometry optimisation and $\rho_{(H+D)-D}$ is the charge density that is calculated from a static run using the coordinates of $\rho_{D+H}$, but with the defect removed. It is of key importance that no coordinate positions or lattice parameters can change in the calculation of $\rho_{(H+D)-D}$, as any relaxation in the nuclear coordinates also impacts the total charge density. In this way, the charge migration that is linked to the defect can be decoupled from the total electrostatic configuration in an approximation. If the TBA can be assumed to hold when a defect is introduced, its AO will interact principally with those on the nearest neighbour atomic sites. However, in the calculation corresponding to $\rho_{(H+D)-D}$ the electronic relaxation in the charge difference map represents the interaction between the charge densities of electrons introduced by the defect and that of the host electrons, driven mainly by electrostatic considerations between a delocalised states from a superposition of AO. Hence, the difference map gives a reasonable interpretation of the charge migrations to and from the defect site. To calculate the difference map, the charge densities are initially calculated by CRYSTAL on a grid of evenly spaced points. For visualisation, the format of the density distributions was converted to be read by the ParaView program [197]. The resultant situation modelled is the perturbation to the charge density from the equilibrium due to the addition of the defect. Charge redistributes, without coupling to the nuclear coordinates, which therefore effectively excludes the interactions between AO and the contribution to band deformation that is due to the crystal structure change.

The significant remaining interactions that dictate the charge density distribution are electrostatic. Charge should move only from the depletion region constituted of the distorted host CBM and VBM, and the defect atom energy bands, which in the case of hydrogenic defects are hybridised with the host orbitals forming two bands, one resonant in the CB and the other in the VB. The bonding and anti-bonding states’ positions are assumed to be dependent on the bonds formed between H and the host anions and cations, yet with fixed atomic positions, the charge redistribution is not dependent upon the occupation of the states (as specified by DOSS). Charge redistribution with fixed nuclear coordinates can nevertheless exert an influence on the form of a band by changing the proportions of charge that are occupied by electrons in the same states. The re-alignment of the VBM and CBM with the calculation of a defect modelled by DFT is well known, an effect originating from the occupation of host states by electrons donated from the defect [192].
5.5 Bulk Phase Diagram Methodology

5.5.1 Mixtures and multi-phase behaviour- The Convex Hull

Up to this point it has been assumed that stable materials will be composed of a single compound formed in a given crystal structure. However, a ground state system will often be able to further minimise its energy by forming an aggregate of two or more ground state compounds that are lower in energy. This is illustrated graphically in figure (5.3). Here, ground states of the binary system $A_xB_{1-x}$ exist for $y, z$ in addition to the pure species A and B. However, $A_zB_{1-z}$ is not stable whereas $A_yB_{1-y}$ is. The stable material for $y < w < 1$ is the general mixture of the stochiometric compounds at $x=z, y$ and the element $B$. Any pure binary compound with composition $w$ is less stable than the aggregate of off-stoichiometric forms of $A_yB_{1-y}$ and $B$. The mixture will have properties derived from these two pure ground states and $B$, even though in the mixture they are not stochiometric and so may not have the crystal structure of the ground states. The existence of multi-compound systems guarantees that for any number of composition variables a minimum energy surface can be constructed, bounding all systems in the space of a set of generalised composition variables $X_1,...X_n$, that is convex throughout the whole space, the *convex hull*. In a system with a finite number of ground states, the properties

![Figure 5.3](image_url)
of a general compound that lies on the convex hull will vary between the pure ground states and intermediate properties of a number of them, since the convex hull also includes all the ground states. Another way to relate the pure compounds to the mixtures is that the latter are produced by disordering stochiometric compounds, with a slight variation in the composition of a compound, whose most stable form will be a number of given perfect or near perfect crystal structures, introducing an increase in thermodynamic stability and other important properties. Thus in searching for the desired materials, even a slight deviation from pure compounds to mixtures may substantially change the system properties.

In the simplest case \( X_i \) are pure elements, but since any combination of the variables may minimise the energy at a point in the space, we see that the phase diagram of an arbitrary system can become complex. For the case of a binary compound, illustrated in figure (5.3) we only have effectively one variable \( 0 < x < 1 \) governing the composition, i.e for a given \( x \) we can only have compounds of the form \( A_xB_{1-x} \). However, the compound for a given \( x \) is in general distinct from others with \( x' \), on account of having a particular crystal structure, and consequently a different energy for the ground state structure. For a ternary compound, discussed in the next paragraph, there are many more possibilities for a given energy. At this stage it is important to note that, since we are constructing phase diagrams on the basis of composition, an important step in calculations is to ensure that we know all structures for a given set of composition variables, and the range of Gibbs free energy of each.

When we have three constituents in the compound of interest, and the compound formula may be written as \( A_xB_yC_z \), a convenient graphical representation of the relative proportions of the three constituents is an equilateral triangle, each side representing the ratio of \( A \) to \( B \), \( B \) to \( C \) or \( C \) to \( A \). The relation between the three sides is easily inferred if the triangle is constructed on a cartesian \( x-y \) axis, with one side as the \( x \)-axis between the origin and unity, and a second side the line \( x = y \). Then the third side will be described by the equation \( y + x = 1 \). If the independent values of the the ratios of say \( A, B \) and \( B, C \) are known as \( x \) and \( y \), this relation then gives the set of available values for \( z \), the ratio of \( A \) and \( C \) at that composition. This time, the formation energies of compounds are graphed in the \( z \)-direction and the convex hull is the surface on which they lie, of course directed outwards at every point.

We restrict the explanation of phase diagram representation to the specific cases of binary and ternary compounds as compounds with more composition variables are less likely to be considered in our investigation. It suffices to say the \( n \) variable generalisation will need \( n - 1 \) spatial dimensions to be represented.

### 5.5.2 Thermodynamics of Pure compounds from DFT data

At \( T = 0K \), DFT is a good approximation to the ground state of the many electron system, and is the starting point for investigations into compound stability under general physical conditions when \( T > 0K \), and the chemical potentials of the atomic species and electrons depart from the equilibrium values corresponding to the formation of the compound. Stability under a given set of conditions is evaluated from considering all the *phases* that the system
might access as its thermodynamic potential varies, as a function of those conditions. The fundamental thermodynamic quantity of interest in determining compounds that form under given conditions is the formation energy $\Delta G_f$ with respect to the elemental reference states of chemical potentials $\mu_A, \mu_B, \mu_C$

$$\Delta G_f(A_xB_yC_z) = E_T(A_xB_yC_z) - x\mu_A - y\mu_B - z\mu_C$$

where the introduction of independent subscripts $x, y, z$ indicates that the full ternary phase space is being explored. However, the physical behaviour of different materials depends on the electron orbitals involved. For a pure metal, electrons are delocalised over the crystal lattice and the corresponding electronic states are expected to show more dispersion throughout the Brillouin zone. The electronic orbitals involved in this behaviour are $s$ or $p$-type and they generally describe the low-density valence electrons. In other solid state systems such as a semiconducting oxide, it is the more tightly confined inner shell electrons, generally described by $d$-type orbitals, that contribute to bonding. In addition, the frozen core or pseudopotential approximation applies to most metals used in DFT calculations, which forces the electronic orbitals in the lower shells, (up to the third shell for Sn and fourth shell for Ta with the use of CRYSTAL in this work) to be doubly occupied. For metals that have a partially filled $d$-band lying lower in energy than the $s$-type and $p$-type orbitals in the next highest shell, this means that some of the outermost electrons lie in $d$-bands. The result is that the description of solid state oxides requires $d$-type orbitals, and these are best described by a DFT functional that is constructed in terms of a basis of functions that are local to a relatively small domain of reciprocal space, such as the Gaussian Type Functions (GTF) that are employed by the CRYSTAL09 code. For a metallic system, on the other hand, the delocalised $s$-type electrons require a basis of periodic functions over a fuller domain of reciprocal space. The ab initio wavefunction description of the core and the valence electrons is for this reason often inconsistent, because the valence electrons are low density while the core electrons are very dense so the correlation and exchange effects have very different influences in the two regimes. Hence, the general practice of DFT within the pseudopotential approximation to represent the core electrons with only valence electrons treated explicitly.

Moreover, even if the basis function used in a particular DFT code does provide an accurate physical description of the material, the functional chosen will give different values for physical properties in a given problem. This means that to construct phase diagrams of compounds across different chemical spaces using more than one distinct DFT functional introduces more than one arbitrary scale of energy values, meaning that the formation energies cannot be calculated directly from equation (79) as the errors in energy values will not cancel between characteristically non-similar materials. The accuracy of practical DFT is not transferable enough between different chemical spaces to construct phase diagrams using a single approach. For these reasons, equation (79) cannot be applied for formation energy calculations and an alternative must be used that achieves some consistency between energy reference values for all the chemical species involved.
5.5.3 Virtual reference states

Two different methods can be pursued to arrive at formation energies for a set of compounds in the (Ta,Sn,O) ternary space.

The first approach is to define formation energies by reactions that do not involve the isolated metal elements $M_1$ and $M_2$, but only oxides and $O_2$. A set of simultaneous equations can be written, by setting equal two equations written in terms of the experimental and the unknown DFT energies of the compounds. In this way, the formation energies should in principle all be self consistent,

$$ \Delta G_{C_i}^{Exp} = f_1(C_1^{Exp}, C_2^{Exp}, ..., C_n^{Exp}) = \Delta G_{C_i}^{DFT} = f_1(C_1^{DFT}, C_2^{DFT}, ..., C_n^{DFT}) \quad (80) $$

For this method to work, the number of compounds (hence, the maximum number of terms) used must be less than or equal to the number of equations. It is difficult to construct a set of reaction equations where two or more of the equations are not similar, perhaps differing only in one coefficient multiplying the same $C_i$. This means that the equations are likely to be overdetermined. In addition, the solution will be very sensitive to the value of the oxygen reference energy, even though it does not appear explicitly, because some equations will contain a term with a large stochiometric coefficient of $O_2$. These difficulties result in a large time burden for finding suitable equations and checking if solutions are consistent.

As an alternative the form of equation (79) can be used with virtual reference states $\mu_1^{*}, \mu_2^{*}$ and $\mu_0^{*}$, which can be written as a matrix equation with $C$ the matrix of stochiometric coefficients. This set of equations is overdetermined as the dimensions of the matrix are 3 x $n$ whereas those of the vector of virtual reference states are 1 x 3, where $n > 3$. However, by multiplying the matrix equation with the transpose $C^T$, a square matrix is obtained constituting a normal equation,

$$ C^T C G_{DFT} = C^T G_{Exp} \quad (81) $$

and the solution is the vector of virtual reference states that minimises the residuals between the total DFT energies of compounds $G_{f}^{DFT}$ and $\Delta G_{f}^{Exp}$, such that variationally,

$$ \Delta G_{f}^{DFT} = \min_{\mu^*} \left| \Delta G_{f}^{Exp} - (G_{T}^{DFT} - C \cdot \mu^*) \right| \quad (82) $$

This is effectively a statement of the least squares problem in matrix form. The equation is used to construct a linear fit between $G_{f}^{Exp}$ and $G_{f}^{DFT}$, taking a number of values of $G_{f}^{DFT}$ as interpolation points. The calculated formation energies are then related to the DFT total energies through $C$,

$$ \Delta G_{f}^{DFT} = \Delta G_{T}^{DFT} - C \cdot \mu^* \quad (83) $$

5.6 Exploring phase space near stochiometric compounds

The convex hull is the locus of lowest energy materials, in the space of composition variables. Pure compounds will be vertices on this minimum-energy surface. Allowing for the possibility
of mixtures of arbitrarily complicated compounds of the composition variables, there can be coexistence of many phases at once. This ensures that the convex hull cannot be determined by consideration of just two variables at a time. One must use the stable compounds obtained from DFT calculations as a starting point and employ methods that 'scan' the compound space for minimum energy hybrids of these compounds. Two main computational methods are suitable for this task, cluster expansions and machine-learning approaches. The former is used in this work.

5.6.1 Cluster expansions (CE) - configurations

A rigorous method that accounts for all available compounds in a portion of the phase space by approximating their energies is the cluster expansion (CE), so named because it expands the energy of a system at a point of the composition space in terms of geometrical units containing a number of lattice points at which variations of the composition variables are considered. To quantify the contribution to the overall disordered energy of a given atom substitution appearing in the statistical ensemble, one must take into account the clusters formed by groups of nearest neighbours to a lattice point in that structure. The CE method follows of the logic of the Ising Model, whereby each lattice point has two possible occupations and interactions are between pairs of sites only. Occupation variables \( \sigma_i \) are introduced, having two distinct values representing occupation by a species A or species B. Then, in order to quantify the substitutions in a solid of increasing size, we write the total energy formed of a given structure \( m \) and the clusters belonging to it in an expansion of interactions \( J_i \) as

\[
E_m(\sigma_1, \ldots, \sigma_n; J_1, \ldots, J_n) = \sum_i J_i \Phi_i \tag{84}
\]

where

\[
\Phi_n = \frac{1}{N_n} \langle \sigma_1 \sigma_2 \ldots \sigma_n \rangle \tag{85}
\]

is the correlation function that calculates the probability of each independent configuration of A and B that can be obtained by making substitutions at the sites of the \( n \)-nearest neighbours. \( N_n \) is the number of such pairs, weighting the probabilities by how many changes of the \( n \)-nearest neighbours are possible [216]. The clusters containing the nearest neighbours take the forms of pairs, triplets and tetrahedra for the cases \( n = 1,2,3 \). Examples defined in an FCC lattice are shown in figure (5.4). The \( J_i \) are effective cluster interactions that parameterise the strength of interaction between atoms that are \( i \)-nearest neighbours. We can write a set of equations of the form (84), one for each of the \( m \) structures considered, in matrix form

\[
E_m^{DFT} = \Phi_{mn} J_n \tag{86}
\]

with the matrix of correlation functions \( \Phi_{mn} \) and the vector of cluster interactions \( J_n \). The energy of a structure \( m \) is chosen to be the value from DFT calculations, so a 'fit' to these theoretical values is thus obtained by solving (86) for \( J_n \),

\[
J_n = \Phi^{-1}_{mn} E^{DFT}_m \tag{87}
\]
If this equation has a solution, which means that the matrix $\Phi_{mn}$ is invertible, the $J_n$ allow one to determine the energies of, in principle any, compound of interest in an adopted structure that lies in the phase space, if the correlation functions can also be computed. However, as explained in the next section, for practical implementations of a CE it is favourable to use statistical fitting procedures in addition to direct inversion of equation (86) to maximise both the convergence and the scope of physics of interactions that can be described.

A large number of configurations of species A and B on the lattice corresponding to one composition can be obtained rapidly from Monte-Carlo methods (the random sampling of configurations produced by a lattice of binary states that have the same overall numbers in state 1 and 2.) An additional computation would be needed to get the correlation functions by calculating the products appropriate for each lattice. In this way local minima of the Gibbs Free Energy for the entire system could be found by finding the configuration with the lowest energy.

An invaluable asset of the CE method is its ability to predict ordered states. For example, in a system where an ion can have different valence states one observes a disordered phase characterised by mixed valency and an ordered phase where the valency of all ions is the same. By using the two valence states as the binary cluster expansion variables one can determine the phase transition between the two phases by finding the configurations of valency states that correspond to a minimum of the electrostatic potential energy [217].

The correlation function calculates all combinatorically different configurations of clusters up to order $n$ of the expansion, that can each have two states; in the case of an alloy, this means each configuration that can be produced from a fixed composition. The cluster expansion therefore provides all the configurations of a disordered lattice, but only in cases with a constant crystal structure, i.e the act of doping is unable to change the geometry of the lattice.
5.6 Exploring phase space near stochiometric compounds

5.6.2 Implementation of Cluster Expansion (CE) - linking CRYSTAL09 to custom code (ATAT)

Within the Alloy Theoretic Automated Toolkit (ATAT) [219]- [224], the method of determining the effective cluster interactions (ECI) (see section (6.2.1)) is based on a linear system of equations of the type of equation (84), which calculate the total energy of a structure $S_j$ as a function of the configuration of the sites at which substitutional defects are made, [225], [216]

$$E(S_j) = \sum_{i=1}^{n} m_i J_i \Phi_i(S_j)$$

where $\Phi_i$ denotes the average of the correlation function $\Phi$, which is calculated by considering all the geometrical clusters in a lattice and their possible occupations by Sn and Ta, given by all possible values of products of binary occupation variables, $\Phi_i = \langle \sigma_1 \sigma_2 ... \sigma_i \rangle$. The other parameters represented by the terms in (88) are the multiplicity of clusters for a given site $m_i$ and the effective cluster interactions $J_i$ which parameterise the strength of interactions of type point, pair, triplet, quadruplet and so on. To construct the equation set represented by (88), a set of $j$ structures are chosen, in each of which is considered $i$ terms in an expansion of clusters that are all defined on the lattice, making the object representing the cluster function average in equation (88) an $i \cdot j$ matrix, $\Phi_i(S_j) = \Phi_{ij}$ with each structure $S_j$ associated with a matrix represented in terms of an expansion in point, pair triplet and quadruplet clusters of the rutile lattice. As often implemented, each row $j$ of the matrix corresponds to a different configuration.

In the investigation documented here, the structures used in the CE are all derived from the rutile lattice via a linear transformation, but each transformation introduces a unique ordering of Ta and Sn with a unique set of configurations so that the averages of the cluster functions are different in each case $j$. The effective cluster interactions can be determined through matrix inversion of (88),

$$J_i = \left[ \sum_{i=1}^{n} m_i \Phi_i(S_j) \right]^{-1} E(S_j)$$

In principle this will yield a solution for the ECI if no two structures share all the same terms in the cluster function average, in other words two rows of the matrix $\Phi_{ij}$ are identical. However, the fitting procedure has the limitation that the number of structures available (in practice) is limited while exact behaviour of the system is only described when the number of fitting parameters, the ECIs, approaches infinity. Thus, the number of ECI should be large, but too many can introduce an effective linear dependence between the constants [226]. A higher number of ECIs also makes the matrix increasingly square- but a model that provides an exact solution via (89) will only describe a fraction of structures in the set under investigation, providing no information about those not included in the fit: the predictability of the model would become restricted by the optimisation of fit accuracy. Thus, the best compromise for constructing the cluster expansion is to have a number of ECI less than the number of structures, and a way of assessing the predictability of the model, how well it performs for a given set of $n$ structures.

The cross validation procedure (CV) uses the difference between the energy of a structure
from the first-principles calculation and the predicted energy, which is obtained from the (least squares) fit to all the other configurations used in the original fit,

$$(CV)^2 = n^{-1} \sum_{i=1}^{n} (E_i - E_i)^2$$

where $E_i$ energy of structure $i$ in $i = (0, \ldots, n)$ predicted from from the cluster expansion fitted to a least squares fit to energies of $n$ structures (as many as are available) and $E_{(i)}$ for $i = (1, \ldots, n)$ is the energy predicted using $n - 1$ structures, where the one energy excluded is for the same structure $i$ as calculated by $E_i$ [219]. The CV provides a more comprehensive indicator of the predictive capability of the CE by exploiting correlations across different structures. For a given cluster, correlations can be grouped so that each structure in the group shares a common vector of correlations as determined by the cluster expansion for the $n$ structures [218]. Each correlation is assigned to a structure whose energy is fitted as described. Repeating for the $n$ possible assignments then makes it possible to define the CV by (90).

The CE also requires accurate reference energies for the end points of the line in phase space that is being explored, pure Ta and Sn oxides. Assuming that the set of structures in the composition space are well chosen, the CE progresses by constructing a larger and larger set of equations based on (88), using the same set of ECI, until the cross validation score reaches an acceptable value. ATAT initially works with a minimal cluster expansion that includes only nearest neighbour pair interactions. As more structures are included it is necessary to include ECI terms that correspond to triplet and quadruplet interactions. These different degrees of interaction are slightly different in each unique cluster that is included. Hence, the ECI are a function of the diameter of the cluster and the number of terms increases as it is necessary to include more pair terms, more triplet terms and more quadruplet terms.

If the cross-validation score does not converge with increasing number of structures, this likely reflects anomalous low order ECIs, since each successive order is based on the those below it.

The operation of the ATAT code needs to be linked to that of the CRYSTAL program in order to obtain first-principles energies of the structures generated by the former. The procedure is illustrated schematically in figure (5.5). The CE considers the contribution of the Ta and Sn ordering to the microstates arising from the distribution of Ta and Sn in the lattice. The states are characterised additionally by orderings in both the spin of Ta$^{4+}$ between the lattice sites and the valence of Ta and Sn. As an approximation, spin-spin interactions (between spin up and spin down electrons) are excluded (their treatment would require functionals tailored for DFT+U), so that ordering of Ta$^{4+}$ and Ta$^{5+}$ is a degree of freedom implicitly representing localisation of an unpaired electron. These constraints require the ordering to be fixed by the valences, which can be 2+, 4+ for Sn and 4+, 5+ for Ta. This is ensured by calculating the minimum Ewald sum of all the configurations produced by the CE for each structure, as a measure of the most stable low energy ordering without spin. Because the initial structure produced by ATAT may be primitive, or be so after the ordering of Ta and Sn has been chosen by the Ewald attribution, the symmetry group for the structure is applied to it if necessary so that the full crystal structure can be recognised by CRYSTAL. As a final check, the structure is tested for charge neutrality. The final structure is supplied to CRYSTAL through an external
5.6 Exploring phase space near stochiometric compounds Matthew Worsdale

Figure 5.5: Flow diagram to illustrate procedure for preparing structures in the space Ta$_x$Sn$_{1-x}$O$_2$ calculated in the cluster expansion with ATAT
geometry input with the valences and the point symmetry operators which substitute for the symmetry information given by the space group.

The CE can be performed in two basic ways. In the constrained CE, the change of crystal structure during the geometry optimization stage of the total energy calculation must preserve the lattice symmetry, which is enforced by allowing the nuclear coordinates to change but not the lattice parameters. In the unconstrained case, both atomic coordinates and the cell parameters are allowed to change. The symmetry is assessed by CRYS
tAL, which requires the point group symmetry operators to be retained at each iteration of the geometry optimisation.

5.6.3 Multiphase co-existence, competition, metastability

Disordered structures from a cluster expansion may be in thermodynamic competition with other phases, which themselves may show disorder over a composition range. If the convex hull for the CE spans the whole of the composition range, it will intersect the free energy curves for any other phase, so that over a subsection of the range one of the phases will be energetically favoured and the ground state stable material will consist only of that phase. The interpretation of this is that one phase is more stable through competition. In this case, one is interested in the doping limits where the curves intersect; at this point, the energetic gain of doping the host used in the CE is less than that of the competing phase at the same composition. This arrangement is illustrated in figure (5.6).

On the other hand, if there are a number of free energy curves existing only over subspaces of the whole composition space, then the most stable material possible will be a phase coexistence. Whether competition or co-existence between the phases is the prevailing effect may depend on the degree of similarity (or difference) between them. If it is possible for the crystal structure associated with one of the phases to be continuously deformed into the other through the act of doping and the associated structural changes, then one would expect one of the phases to out-compete the other at a given composition, as doping tends to drive towards one structure. If however, there remains a significant energetic difference between the phases over much of the composition range, then doping will not have the effect of promoting one structure and the expected outcome is a two-phase co-existence.

5.7 Intrinsic defect calculations in SnO$_2$

Here is outlined the methodology for calculation of oxygen vacancy defects. This forms the initial part of a comprehensive investigation into extrinsic and intrinsic defects in the system that enables one to establish the maximum concentrations of charge carriers, leading ultimately to predictions of the conductivity trends as a function of the prevailing thermodynamic conditions and doping limits. For the consideration of the feasibility of the defects, the primary property
is the defect formation energy, recalled from the methodology in section (5.3.1),

\[
\Delta G_f = G^{\text{def}}_{T} - G^{\text{perf}}_{T} - \sum_{i} \Delta n_{i} \mu_{i} - x \mu_{e}
\]

where \( n_{i} \) for this defect system is only the oxygen chemical potential \( \mu_{O} \). For this quantity the DFT energy of an isolated \( O_2 \) molecule was computed with the same functional and basis set and the result halved to obtain an energy per O atom of \(-2046\,eV\). As discussed in (5.3.2), the true value used in preparation of a defect sample would lie within a range fixed by the chemical potentials of all the species involved in their naturally occurring states, as well as the experimental conditions.

When the O atom is removed, the two valence electrons that were shared with the Sn to form covalent bonds are left over and must be redistributed back to the Sn. This restores local charge neutrality (within a supercell) that is required for the DFT codes. In addition, to respect the fact that the electrons left behind in the vacancy creation were in oxygen-like orbitals, the basis functions which CRYSTAL09 uses to approximate the electron density are kept at the vacancy site (‘ghost’ functions) [136], before the defect system is created by the specification of the defect site and the redistribution of electrons, including that specified by the valence redistribution. Because no electrons are added to or removed from the overall system (\( x=0 \))

Figure 5.6: Generalised phases \( \alpha \) and \( \beta \) in a binary composition space with end limits A and B and Gibbs Free Energy as a function of the molar fraction \( x_{B} \). Phase \( \gamma \) exists in the same composition space but has only a narrow homogeneity range relative to \( \alpha \) and \( \beta \) and will compete with these phases.
the electron chemical potential $\mu_e$ does not appear in (91), which can therefore be rewritten,

$$\Delta G_f^{V_{ac}(O)} = G_{T}^{def,q} - G_{T}^{perf} + \mu_O$$

(92)

The use of the B3LYP functional was continued following its successful trial documented in section (6.1). A convergence criterium of $10^{-6}$ eV was imposed on all total energy values.

<table>
<thead>
<tr>
<th>Supercell</th>
<th>Translation Matrix</th>
<th>Number of formula units $x$ (atoms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>1 (6)</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>2 (12)</td>
</tr>
</tbody>
</table>
| 3         | \[
\begin{pmatrix}
1 & -1 & 0 \\
1 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\] | 4 (24)                               |
| 4         | 2 x A              | 8 (48)                               |
| 5         | 2 x B              | 16 (96)                              |
| 6         | 3 x A              | 27 (162)                             |
| 7         | 4 x A              | 64 (384)                             |
| 8         | \[
\begin{pmatrix}
4 & -4 & 0 \\
4 & 4 & 0 \\
0 & 0 & 3
\end{pmatrix}
\] | 96 (576)                             |

Table 3: Supercells used to examine the scaling of defect formation energy in SnO$_2$, obtained by the respective translation matrices applied to the unit cell.

5.8 Hydrogen defects in SnO$_2$

5.8.1 Hydrogen defects and calculation details

Hydrogen can form two basic types of defects in the host lattice; substitutional and interstitial defects. In the substitutional case H replaces a host oxygen; In the interstitial case it will bond either to a host Sn or O. In keeping with the discussion in section (4.6.1), given that both the substitutional and interstitial defects can form in different valencies and act as donors and acceptors, with the formation energies a function of the chemical potentials of O (in the substitutional case) and H, it is assumed that a relation between the two defect types can be made at a given value of $\mu_H$ and $\mu_O$, and hence the relative stability of the two defects will be an important factor behind their respective concentrations in the material. To determine this, the formation energies of the defects at different positions are calculated at all possible valencies. In all cases, the final energetically stable position of the defect (after structural optimisation) could vary significantly depending on the initialisation (placement). For every initial geometrical configuration of the host atoms + defect, global minima will exist where the same final position corresponds to more than one input position. For an interstitial defect,
finding phases that contain potential energy changes due to both the continuous rotational
degrees of freedom of OH groups (for interstitial defects) and discrete (configurational) degrees
degrees of freedom of substitutional defects would also require generalization of a continuous cluster
expansion (CE) technique [232]. Without these approaches and under time constraints, some
results may be inaccurate.

**Substitutional** This defect is treated with redistribution of the the valence electrons of the O
removed from the lattice evenly between cations in the system using the CHEMOD facility in
CRYSTAL09. As discussed below, it is possible that other placements (initialisations) of the
charge to be redistributed would more accurately reflect the bonding of H with the host, though
trialling the possibilities would be exhaustive.

**Interstitial** Two initialisations of H were considered. The first corresponds to the position for
interstitial Sn which was used in studied reported in the literature review of SnO$_2$ [162], [165].
To allow for any possible dependence of the final position on the initial coordinates for this
crystallographic position, the calculation was repeated with the H at a number of subdivided
intervals on the vector between adjacent O atoms on the corners of the 'square' bounding the
interstitial region. This position will be referred to as 'type A' in the discussion of the results.
The other position sees H take a position in the (110) plane, on a vector between Sn and the

Figure 5.7: Interstitial hydrogen defect types 'A' and 'B', as described in text, left and right
respectively

nearest O atom (the two are not bonded). This will be referred to as 'type B'. The two types are
illustrated in figure (5.7). All calculations were employed using the 2x2x2 (48 atom) supercell.

**Basis sets** For all calculations except the Jellium approximation used to study a charged H$^+$
defect, the use of electron redistribution to give charge neutrality favours the use of a single
basis set for Sn. When the CRYSTAL code takes the superposition shell charges to commence
the SCF calculation of total energy, the symmetry requirements force the same charge for all

93
atoms of a given species. In practice, the bonding between H and the host atoms will likely only involve a number of nearest neighbours, meaning that the redistribution of the valence charge from hydrogen between all the Sn using CHEMOD. In practice, it could be a more accurate alternative to use separate basis sets for Sn, one for those that do share valence electrons with H and another for those that don’t, but this would require the proportion of charges to be redistributed to the different species to be known 'a priori'.

Furthermore, Basis Set Superposition Errors (BSSE) can occur when different basis sets overlap, which will occur to a different extent when the energy of the system is minimised as a function of the geometry depending on the basis set chosen. This could make it more important to perform the calculation with mixed basis sets, where some of the orbitals are unoccupied, and compare with the uncorrected case [227].

The basis set used for hydrogen is given in the appendix.
6 RESULTS

6.1 Bulk Defect Calculations on rutile SnO$_2$

As discussed in section (4.2.2), much research activity into replacement catalyst supports for PEFCs has been directed onto metal oxides based on the idea that the metal element cannot be further oxidised as it is already in its highest oxidation state. Using the methodology outlined in section (5), defects of SnO$_2$ in the rutile structure are investigated. To assess the formulation of DFT most appropriate for use, fundamental properties were calculated using two different platforms, CRYSTAL09 and VASP (Vienna ab initio Simulation Package). On this basis, CRYSTAL09 was judged to deliver parameters in better agreement with experiment.

6.1.1 Bulk calculations in VASP and CRYSTAL09

Initially, two calculations of SnO$_2$ bandstructure were made using VASP, one under the LDA (Local Density Approximation) regime and the other under the GGA (Generalised Gradient Approximation) (see sections (4.3.5) and (4.3.6)). Although these are not the only two approximations to the exchange correlation functional, they are among the most widely used functionals in ab initio investigations of electronic structure generally, as well as the simplest in the hierarchy of specialisation thus making it important to have a result against which to judge the performance of more developed functionals.

**Study of Convergence with k-point mesh**

Before undertaking the bandstructure calculation the SnO$_2$ rutile unit cell was optimised, using a conjugate-gradient algorithm supplied with VASP [149] in which the ions are allowed to relax into an instantaneous ground state under the influence of the initial inter-ionic forces, where the total energy is recalculated and the forces are taken as the total gradient [148], [149]. As discussed in section (4.3.9) the total energy is determined from an integration in reciprocal space over points where the single-particle Kohn-Sham equations are solved. The accuracy of the relaxation, just as the total energy calculation, therefore depends on the sampling of the k-points in the Brillouin zone (BZ), and in addition the definition of the grid in reciprocal space over which the integration is taken. The density of k-points needs to be ‘optimal’, i.e. a number that will capture the $\tilde{k}$ dependence of the energy at different points in the BZ. This is dependent on the symmetry of the integration region (the BZ defined by the lattice vectors) and the unit cell size. To test k-point convergence, the structure was allowed to relax as described for a varying number of uniformly distributed k-points obtained by subdividing the BZ in each direction. A convergence criterium of $10^{-5}$ eV was set for the accuracy of the total energy and ionic relaxation was set to stop when the change in total energy between two relaxation steps $dE < 10^{-4}$ eV. The calculation was spin-polarised [148] in order to allow for the possibility of different occupancies of states by spin-up and spin-down electrons.

The optimisation saw convergence of the total energy to within the aforementioned convergence.
6.1 Bulk Defect Calculations on rutile SnO$_2$

Matthew Worsdale

Figure 6.1: The path in reciprocal space used to obtain the bandstructure for SnO$_2$ and defected structures throughout.

A criterion achieved with a $k$-point density of $72(\AA)^{-3}$ for GGA and $160(\AA)^{-3}$ for LDA. A clear convergence of the energy with increasing $k$-points can be seen on figure (6.2), suggesting that a rather dense $k$-point mesh is needed to ensure that total energy calculations are independent of $k$-point sampling.

**Bandstructure** The procedure for determining the bandstructure using VASP is in two main stages. Firstly, a self-consistent calculation of the energy is made with a high accuracy convergence criterion in order to determine the ground state charge density. Secondly, this charge density is used as the fixed charge density in a static run (fixed ionic positions) with the energy evaluated along high-symmetry directions in reciprocal space (a non self-consistent run).

Because symmetry properties of the Brillouin zone cannot be exploited, since the integrations are now confined at points on a line, VASP internally generates strings of high-symmetry points along user specified lines. There is some degree of freedom in how these paths are defined. The choice here is derived from the Wigner-Seitz cell of the reciprocal lattice, following Setyawan et al. [150]. Ten points were used along each of the three paths $\Gamma \rightarrow X$, $X \rightarrow M$ and $M \rightarrow \Gamma$ which are shown in Figure (6.1). The pseudopotentials used in the calculations were of the ultrasoft form, within the PAW approach. The configuration of the outermost electron states was $s^2p^2$ for Sn and $s^2p^4$ for O.

The geometry corresponding to a 24 $k$-point calculation was used for the bandstructure, which is shown in (6.7). A separate bandstructure is obtained by VASP for both electrons in a spin-up/spin-down pair in a spin-polarised calculation, but was found to be practically identical in
Figure 6.2: Total electronic energy after relaxation as a function of the number of k-points in the Brillouin Zone of the SnO$_2$ unit cell in the LDA and GGA (PW91) [186], [187] functionals as calculated in VASP. The two functionals give energies over different scales, which is arbitrary in DFT as discussed in section (5.5.2). In both cases a k-point density greater than 250 points ($\text{Å}^{-3}$) was needed for convergence.

Discussion The most significant difference between the LDA and the GGA concerns the band gap. The GGA calculation gives no band gap at all whereas the LDA bandgap is approximately 1 eV and direct. This is very inaccurate compared to the experimental value, but would agree with the widely reported trend that local functionals in general, whether LDA or GGA, underestimate the band gap as discussed in section (4.3.6). This is a known problem with semi-local functionals, originating from self-interaction errors in the Kohn-Sham method [228]. Other defining features of the bandstructure are in both cases in broad qualitative agreement with what is known from experiment. Occupied bands below the Fermi level are characterised by limited dispersion, indicative of the relatively localised nature of the oxygen states. Bands above the Fermi level are much more dispersed. In particular, the lowest lying unoccupied band shows almost parabolic dispersion close to the $\Gamma$ point, reminiscent of the free electron. On this basis, one can say that LDA would be better than GGA for undertaking the defect calculations documented in this report, yet it is nowhere near adequate because the band gap is far too small, hence the electronic structure in the region of the band gap, including the defect levels therein, will likely be incorrectly positioned with no possibility to distinguish them from the bulk electronic structure.

Correct determination of the electronic structure and upon it a successful model of the ele-
Figure 6.3: Bandstructure of SnO$_2$ as calculated in the GGA approximation (top) and LDA (bottom). The $\Gamma$, X, and M points are described by figure (6.1). The red line on both plots indicates the Fermi level $E_F$. 
tron physics of the solid state system is due to a combination of choosing the best functional, the choice of a plane wave or gaussian function description of the wavefunction, and the level of the theory at which these are applied as indicated in Figure (4.8). In SnO\textsubscript{2}, a functional that can treat non-local exchange is required to overcome the self-interaction error of the local functionals LDA and GGA. This gives motivation for repeating the preceding calculations in CRYS\textsc{tal}09, using the B3LYP hybrid functional which uses a mixture of the exact Fock exchange and GGA exchange correlation, with in addition a contribution from a third type of functional originally introduced by Becke \[151\], \[152\], \[127\], discussed in section (4.3.7). This functional is not computationally efficient in VASP due to the need to include plane wave components with an increasingly small \(\vec{k}\) coefficient to realise a complete set, whereas in GTF based programs the gaussian functions are locally optimised.

With the B3LYP functional employed to calculate rutile SnO\textsubscript{2}, the band gap is 4.08 eV, following a convergence study graphed in figure (6.5). This is much closer to the experimental result, though still an appreciable overestimate. The bandstructure is shown in figure (6.4). It is in good agreement qualitatively and quantitatively with an earlier almost equivalent calculation on SnO\textsubscript{2} (using B3LYP with CRYS\textsc{tal}06) \[172\]. Since the electronic structure must be as accurate as possible, and the band gap as calculated by B3LYP is a small overestimate it is unlikely that any band gap states will fall erroneously within the valence or conduction bands, so the system should be correctly described as insulating, semiconducting, or metallic.

<table>
<thead>
<tr>
<th>Platform</th>
<th>Functional (approximation)</th>
<th>Band gap (eV)</th>
<th>(a)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VASP</td>
<td>LDA</td>
<td>1</td>
<td>4.972</td>
<td>3.244</td>
</tr>
<tr>
<td>VASP</td>
<td>GGA</td>
<td>0</td>
<td>4.956</td>
<td>3.262</td>
</tr>
<tr>
<td>CRYS\textsc{tal}</td>
<td>B3LYP</td>
<td>4.08</td>
<td>4.725</td>
<td>3.162</td>
</tr>
<tr>
<td>Experimental</td>
<td></td>
<td>3.54</td>
<td>4.737</td>
<td>3.185</td>
</tr>
</tbody>
</table>

Table 4: Physical properties of rutile SnO\textsubscript{2} as calculated in different DFT codes and functionals, in comparison with experimental values. Experimental band gap measured using optical absorption spectroscopy at \(T = 80\)K \[144\]

6.1.2 Bulk oxygen vacancy

The defect formation energy is graphed as a function of the inverse supercell size in figure (6.6). Instead of converging to a maximum defined asymptotically in the dilute limit, a maximum in the formation energy of 3.47 eV occurs for a supercell of 48 atoms, with apparent decrease for larger supercells. This suggests slow convergence of the formation energy with supercell size, though limits on computational resources needed to calculate larger supercells keep this claim unsubstantiated. Although the supercells used were not all of the same shape, there is no aperiodic charge in the unit cell since the vacancy defect was modelled as neutral. Hence the behaviour for supercell sizes greater than 48 atoms cannot be an indication of unphysical behaviour caused by spurious interactions of any charge created by the periodic boundary conditions, since
Figure 6.4: Bandstructure of rutile SnO$_2$ as calculated with the CRYSTAL09 code (based on Gaussian Type Functions), using the B3LYP hybrid functional.
$\rho(r)$ is constant. It is possible that the non-converged energy represents the elastic contributions to the total energies of the defect supercell, due to the creation of a quasi-mechanical stress field by the vacancy. Such effects have been found to give a change in formation energy for neutral vacancy defects of the order of $0 - 0.1 \text{eV}$ for a change in supercell size from 72 atoms to 576 atoms [199], whereas here the largest cell calculated is only 384 atoms due to aforementioned computational limit constraints.

Of primary interest to judge whether the results is sensible in comparison to other reported results is the range of formation energy variation over a defined variation in the oxygen vacancy concentration, and the physical trend in the quantity described. While this defect is of much interest to computational and experimental investigation of conductivity, relatively few studies are found examining a bulk oxygen vacancy computationally in SnO$_2$, though results are more easily found for studies of the defect on given surfaces of rutile, or on nanofilms. One study by Godhino et al. [171] reports a result for an isolated defect (it should be cautioned that VASP was used with the known underestimation of total energies by GGA) with variation of the formation energy between the low and high limits of defect concentration from 2.14 eV to 5.24 eV, or a proportional increase of 2.4. The corresponding proportion in this case is about half that factor, while the absolute values are also of the same order of magnitude. Entirely like for like comparison is complicated by the concentration limits by Godhino et al being defined by oxygen chemical potential rather than supercell size, with relation of the limits in the two data sets therefore dependant on the absolute values of $\mu_0$ that are generally dependent on computational details. It gives confidence to be able state that the values are of the same order of magnitude, given that the properties of SnO$_2$ based materials are commonly reported to vary dramatically with small changes in thermochemistry.
The bandstructure for the dilute limit of the oxygen vacancy is shown in figure (6.7) (for

![Figure 6.6: Formation energy of neutral oxygen vacancy defect $\Delta G^{Vac(O)}$ obtained by removing an O atom and redistributing the valence electrons from the Sn–O with the spatial states of the electron preserved (see text), for unrelaxed geometry (green line) and relaxed geometry (blue line), Scaling is with the effective inverse lattice constant $1/a = x^3$. The supercells used correspond to numbers 1 to 6 listed sequentially in table (3). While the cells are not all of the same shape, they are equivalent from a computational point of view because the vacancy defect is modelled as neutral (see text).]($\text{Insert Figure 6.6}$)

supercell number 6 in table (3), having 162 atoms). The defect forms a deep donor level about 2.45 eV below the CBM. Such a defect level will make minimal contributions to $n$-type conductivity, since thermal excitation of electrons decreases exponentially with increasing donor depth by Maxwell-Boltzmann statistics. So, although the defects have feasible formation energies, it is unlikely that they are a basis for increasing the conductivity of SnO$_2$, particularly given that oxygen vacancies can become re-oxidised over time.

6.1.3 Conclusions

The effect of oxygen vacancy defects on the electronic structure of SnO$_2$ has been studied. Modelling oxygen vacancies as neutral defects, and using the supercell approximation to approach the dilute limit has been sufficient to quantify their contribution as $n$-type donors. The remainder of this thesis presents calculations on extrinsic defects in the SnO$_2$ system as an alternative to intrinsic defects.
Figure 6.7: Bandstructure for the Vac(O) defect in bulk rutile SnO$_2$ calculated in the dilute limit (162 atom supercell, number '6' in table (3)). Charge neutrality of the unit cell was enforced by redistributing the charge of two electrons left at the site of the oxygen vacancy equally between all tin atoms.
6.2 Ta-doped SnO₂

6.2.1 Motivation for study of Ta-doped SnO₂ and exploration of Ta-Sn-O ternary phase space

Extrinsic defects can be introduced into a host as substitutional impurities or interstitial impurities. Substitutional defects can be conveniently modelled as site substitutions. Suppose that it is possible to obtain the desired conductivity properties from a solid solution (M₁M₂)O₂ with a thermodynamically viable formation reaction from two binary oxides of the respective metals M₁ and M₂. The intermediate oxides then lie within a binary phase region that has the pure oxides M₁O₂ and M₂O₂ as the endpoints. In particular if the crystal structure is the same in both, then the sites occupied by M₁ or M₂ are equivalent throughout the composition range. In this case, the phase space of all the mixed compounds (M₁M₂)O₂ produced from the pure oxides can be studied as a binary subspace within a ternary phase space of M₁, M₂ and O, as in section (5.5.1), with the disorder of doping pure compounds studied in the context of substitutional doping. As each M site can then have two possible occupations, interactions can be approximated as between M₁/M₂ pairs, and all of the mixed compounds can be studied by the use of a cluster expansion (CE). As SnO₂ is a thermodynamically stable host under the required conditions, a second metal that forms the same crystal structure as a ground state is targeted. Nb and Ta can both form the rutile dioxide [140]. Moreover, they will both act as an n-type donor in a system with SnO₂ as they are both capable of forming an oxidation state of 5+ while Sn can only oxidise up to 4+. In order to explore the possibility of extrinsic doping it is necessary to study the ternary system (Sn,Ta,O) or (Sn,Nb,O) and obtain the phase diagram in order to know stable compounds and their crystal structures, since the binary subspace bounded by rutile dioxides may contain additional binary and ternary stochiometric compounds.

As a side note, the distribution of host and dopant atoms between sites could in principle have an important bearing on the system behaviour- two configurations corresponding to the same composition x, and even those that are equivalent by symmetry, may not be equivalent overall, because of interactions between the sites that are distance dependent. However, it is assumed that the physics of all such interactions are incorporated through the CE.

6.2.2 DFT Exploration of Ta-O-Sn ternary space

Experimental data [215] for the composition space reveals a number of commonly occurring binary oxides with little difference between the corresponding values for ΔG and ΔH. The computational details of the DFT calculations and the resulting energies are displayed in Table (5). Experimental thermochemistry data was only found for the five binary compounds. Within these, in addition, the available formation energies TaO and TaO₂ referred to the gas states. For this reason, the fitting based on the procedure described in section (5.5.3) was made using only the data for SnO, SnO₂ and Ta₂O₅. Interpolated formation energies ΔGᵢ^{Exp} = ΔGᵢ^{DFT} − C · μ
were therefore taken for TaO and TaO\(_2\), on the basis of the fitting procedure, to be compared with the true experimental values. These calculated values are offset by approximately the same energy from the corresponding interpolated values, suggesting that this offset approximately corresponds to the difference in energy between the solid and gas phase of these compounds. The calculated formation energies \(\Delta G_f^{DFT}\) are used as the basis of the ternary phase diagram.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Space Group(s)</th>
<th>Valencies</th>
<th>Platform</th>
<th>(\Delta G_f^{exp}) (eV/f.u)</th>
<th>(\Delta G_f^{DFT}) (eV/f.u)</th>
<th>Conductivity type</th>
</tr>
</thead>
<tbody>
<tr>
<td>SnO</td>
<td>C m c 21</td>
<td>Sn(^{2+})</td>
<td>CRYSTAL</td>
<td>-2.661183</td>
<td>-2.661183</td>
<td>Semiconductor</td>
</tr>
<tr>
<td>SnO(_2)</td>
<td>P 42 n m n</td>
<td>Sn(^{4+})</td>
<td>CRYSTAL</td>
<td>-5.389423</td>
<td>-5.389423</td>
<td>Semiconductor</td>
</tr>
<tr>
<td>TaO</td>
<td>F m -3 m</td>
<td>Ta(^{2+})</td>
<td>CRYSTAL</td>
<td>-1.694922 (g)</td>
<td>-5.122880</td>
<td>Metallic</td>
</tr>
<tr>
<td>TaO(_2)</td>
<td>P 42 n m n</td>
<td>Ta(^{4+})</td>
<td>CRYSTAL</td>
<td>-2.081478 (g)</td>
<td>-9.785044</td>
<td>Metallic</td>
</tr>
<tr>
<td>Ta(_2)O(_2)</td>
<td>P C C M</td>
<td>Ta(^{5+})</td>
<td>CRYSTAL</td>
<td>-19.806035</td>
<td>-19.806035</td>
<td>Metallic</td>
</tr>
<tr>
<td>SnTaO(_3)</td>
<td>P m -3 m</td>
<td>Sn(^{4+}) Ta(^{2+})</td>
<td>CRYSTAL</td>
<td>N/A</td>
<td>-10.157011</td>
<td>Metallic</td>
</tr>
<tr>
<td>SnTa(_2)O(_6)</td>
<td>C 1 2/c 1</td>
<td>Sn(^{2+}) Ta(^{4+})</td>
<td>CRYSTAL</td>
<td>N/A</td>
<td>-25.972354</td>
<td>N/A</td>
</tr>
<tr>
<td>SnTa(_2)O(_7)</td>
<td>C 1 2/c 1</td>
<td>Sn(^{2+}) Ta(^{5+})</td>
<td>CRYSTAL</td>
<td>N/A</td>
<td>-26.038042</td>
<td>N/A</td>
</tr>
<tr>
<td>Sn(_2)Ta(_2)O(_7)</td>
<td>F d -3 m</td>
<td>Sn(^{4+}) Ta(^{5+})</td>
<td>CRYSTAL</td>
<td>N/A</td>
<td>-25.80254</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 5: DFT calculations on the Ta-Sn-O system for the Ta-O-Sn composition space. The convex hull is the surface of minimum energy that bounds \(\Delta G_f^{DFT}\), as discussed in section (5.5.1).
Figure 6.8: Ternary phase diagram for Ta-O-Sn system obtained from DFT using all known stoichiometric compounds (left), and using only the set of compounds for which experimental formation energies were used from reference [215]. (right)
Figure (6.8) shows the resulting phase diagram. The phase diagram based on the five experimentally available compounds is also shown for comparison. For the DFT case, it is seen that TaO$_2$ and SnO$_2$ are both stable compounds in this phase space, whereas the binary monoxides SnO and TaO are not stable in either case. Moreover, one of the ternary compounds SnTa$_2$O$_6$ considered is stable on the tie line connecting SnO$_2$ and TaO$_2$, indicating that this compound is competitive thermodynamically with the disordered compounds of SnO$_2$ and TaO$_2$ on the tie line bound by the pure dioxides. The instability of TaO$_2$ in the experimental case is, as discussed, because the formation energy TaO$_2$ taken from the tables is for the gas phase (g) of the compound, so does not coexist with solid SnO$_2$ and Ta$_2$O$_5$, while the DFT value for the solid state in figure (6.8) makes the TaO$_2$ stable along with SnO$_2$ and the compound SnTa$_2$O$_6$. Since SnO$_2$ and TaO$_2$ are both stable in the rutile structure, the study into extrinsic doping in the Ta-Sn-O system focuses initially on the binary subspace bounded by the two compounds, explored by a cluster expansion. SnTa$_2$O$_6$ is a stoichiometric compound that will partake in two phase coexistence or competition with the disordered compounds of the cluster expansion.

6.2.3 Cluster Expansion- Enforcement of Rutile Symmetry

The CE was constructed with the original rutile lattice of the structures preserved. After a relatively long computation time and calculation of 57 configurations, the convex hull shown in figure (6.9) emerged. An inherent problem with any cluster expansion is the fact that structures with extreme ratios of the two metals are less common than other compositions and require the computation of many larger structures, two factors which play into long overall computation times. However, the more fundamental problem in this case is that the convex hull is shallow across the range of structures computed; the energetic gain in stability of the structure due to doping is maximised at just 0.09 eV. This would suggest that the possible increase in thermodynamic stability due to the introduction of disorder through substitutional defects in the SnO$_2$/TaO$_2$ system is small and hence the likelihood of this structure being present in a mixed compound would be low.

However, the ATAT code gives other useful information about the system, such as the plot of the predicted energies $E_{CE}$ versus the true DFT formation energies $E_{DFT}$. The correlation between the two is far from satisfactory, as shown by figure (6.10). This can be explained as follows.

The set of structures of a given Ta : Sn ratio are distinguished from each other by the symmetry changes possible within the constraint of fixed rutile lattice parameters. Symmetry is defined by a space group, and a number of point group symmetry operators (PGO). A PGO is an operation defined on the lattice belonging to the space group of rutile which maps a given point to another that is equivalent by the symmetry. For the perfect SnO$_2$, there are 18 PGO in total. The structures generated and calculated by the cluster expansion are all times derived from the rutile lattice, but changing the Sn : Ta ratio will invalidate the PGO applicable to the lattice points where substitutions are performed, so that the CE structures retain a varying number of the original PGO set. One might expect that the overall symmetry of a structure
6.2 Ta-doped SnO$_2$

Matthew Worsdale

![Formation Energy Chart](image)

**Figure 6.9:** Formation Energies of all structures in the cluster expansion (CE) performed with enforcement of the preservation of rutile lattice. Rutile TaO$_2$ and SnO$_2$ are the reference states. The minimum energy surface or ‘convex hull’ for the cluster expansion is indicated by the black line. At each discrete composition is a ground state (gs) whose corresponding structure lies on the convex hull. The greatest stabilisation by 0.09 eV occurs for a Ta loading of 66.6%.

correlates at least roughly with the number of PGO, despite the likelihood that not all PGO have the same consequence in a given configuration. To an extent, a correlation between the number of symmetry operators) and the composition is to be expected simply because for some compositions, for example the extreme ends of the range, there are fewer symmetrical ways to arrange many atoms of one metal with a low concentration of the other. However, there are few examples where arguments of this kind apply. The structures with compositions Ta:Sn = 0.25,0.75 generally have at least 2 symmetry operators, with only one exception. On the other hand, the composition of Ta:Sn = 0.3333 has only structures of the minimum symmetry. This highlights that there is an overall dependence of the symmetry on configuration and hence some dependence on the composition.

The varying number of PGO effectively constrains the relaxation of some configurations more than others. If there is a symmetry breaking relaxation, which is forbidden by the set of PGO for some configurations but allowed for others, an inconsistency of the sampled energy space would result. This could be an explanation for the limited correlation between CE and DFT.

To obtain a more informative picture of the influence of symmetry properties of the rutile structure, it is relatively simple to construct groups of structures that share a common set of the rutile point group symmetry operators, of which there are 18 in total. The groups are constructed by noting the PGO that each structure possess, which is the same before
Figure 6.10: The true formation energies of structures in the cluster expansion (CE) $E_{DFT}$ and the corresponding values as predicted by the code based on least squares fitting, $E_{CE}$.

and after the relaxation. To represent the data, a matrix plot is used, figure (6.11), with each structure plotted along one axis and the symmetry operators on the other, so that the groups of structures correspond roughly to the vertical columns as viewed. In addition, the colour coding represents the stability of the structure with the listing on the horizontal axis following the order of increasing stability. The plot shows that structures of stability within a range frequently possess a common group of symmetry operators, for example, the group of structures that are the most unstable in the set, shown clearly on the right hand side of the plot. This emphasises that correlation between composition, symmetry and stability needs to be described in terms of PGO that are associated with particular symmetry breaking operations and this would require an explicit representation of the transformation effected by each operator. However, the relaxation of the configurations is incompletely described by the PGO when the rutile lattice is kept, since the excluded relaxations have varying influence in each structure. More meaningful results are given when the cluster expansion is repeated with any symmetry change permissible.

### 6.2.4 Cluster Expansion, no enforcement of symmetry

The option to allow lattice parameters to change is implemented in CRYSTAL, so that at each iteration of the geometry optimisation the set of PGO attributed to the structure can change. The final structures need to be analysed and compared in terms of their lattice parameters and other characteristics pertaining to a common lattice. It is simplest to go with the original rutile
lattice, but this requires an additional step in addition to the methodology followed for the initial CE. In that case, it was known that all operators applied to the same crystal system, but this time the final relaxed structures will have operators that pertain to a different space groups. This arises because in the transformation from the rutile unit cell to the various structures of the CE, there necessarily exists a primitive cell that is transformed by a matrix \( T \) to that used by the CE, where the primitive cell is related to the original rutile by a matrix \( R \). The primitive cell contains the minimum lattice points for the space lattice involved. However, because the composition (Ta:Sn ratio) is varying continually across the set of all structures, the primitive cell in each case is a different size and therefore the primitive cell corresponding to the final structure (obtained by a transformation of \((RT)^{-1}\)) cannot be studied to deduce the relationship between different CE structures because the relaxations are size dependent in part. More specifically, a general cell distortion will be a combination of:

- Local relaxations about atom centres. These may be purely space filling, *e.g.* occur because of a change in ionic radius between Ta and Sn
- Symmetry breaking distortions

It is not possible to distinguish the two types of relaxation, so the only valid method of comparing CE structures is to transform from the final relaxed geometry, *back* to rutile, using the inverse
of the original transformation matrix \( T^{-1} \). The cell resulting from this operation must be enlarged, so that each atom from the original cell is mapped exactly onto an atomic position in the original rutile basis. For example, if the largest entry in the original matrix \( T \) is 3, the enlarging matrix must be a 3x multiple of the identity matrix. This procedure can generate rutile-like cells for comparison.

The bond lengths are studied, as these should give an indication of where electrons are being accommodated due to the extrinsic doping in the system facilitated by Ta. From the computed structures in the CE, one is trying to obtain information about

- **Electrical conductivity**: assuming bulk semiconductivity is maintained, the mechanism for that conductivity, governing how electrons are donated to the conduction bands from the valence bands, and what electronic states/orbitals are involved. With a Ta-Sn alloy system, from first principles it is predicted that Ta will donate electrons to the system as it will share bonding electrons with Sn, leaving one \( e^- \) donated if the valence is 4+. 

- **Thermodynamic stability**: based on the above limited observations of the rutile-only system, it is expected that on stability structures will be divided into groups, and there is a need to understand the origin of these groups.

![Figure 6.12: Minimum energy surface 'convex hull', with no restrictions on the symmetry. Meaning of the plotted quantities is the same as in figure (6.9). Dashed red line is 'convex hull' of the stochiometric compound SnTa\(_2\)O\(_6\) whose energy has been calculated with rutile TaO\(_2\) and SnO\(_2\) as reference states.](image)

The CE with unconstrained relaxation readily gives results and eliminates the apparent configuration-
dependent source of error that generated a poor fit between $E_{DFT}$ and $E_{CE}$ in the constrained case. After the computation of 61 structures the convex hull is as shown by figure (6.12). This time, the system is much more sensitive to the site substitution, with the minimum ground state formation energy an order of magnitude larger than was observed for the constrained rutile case. The alloying fraction for which it is the greatest is Ta:Sn = 3:1.

From the point of view of doping limits, of interest is the intersection of the convex hull for the CE structures with that for the stochiometric compounds in the system, which as far as we know is just SnTa$_2$O$_6$. The formation energy of this structure is indicated on figure (6.12), lying just lower in energy than the most stable CE structure. Accordingly, the effective doping limit where the convex hull for SnTa$_2$O$_6$ intersects the convex hull for the CE corresponds to a Ta:Sn composition of about 3:1 (75%). For Ta fractions above this limit only, the rutile alloy is stable. For 67% < Ta < 75% there exists a two-phase coexistence with SnTa$_2$O$_6$; and for Ta < 67%, a two phase coexistence between SnO$_2$ and the ternary exists (at 0K). Thus, substitutional doping in rutile alloys between Sn and Ta is feasible for Sn < 25% (i.e., Sn doping of TaO$_2$) but beyond this limit, a two-phase mixture formed of a rutile solid solution and SnTa$_2$O$_6$ thoreaulite persists. It should be noted that this result could change at a finite temperature, in which case the convex hull for the rutile structures would be replaced by curves of Gibbs Free Energy (one for both TaO$_2$ and SnO$_2$), giving potentially altered doping limit(s). Since the ternary compound is only slightly more stable than the mixed rutile compound at the same composition, the ternary phase only slightly out-competes the rutile phase at the Sn-rich limit of the composition space at 0K, making it possible that with configurational entropy the Gibbs Free Energy curve for the rutile phase would intersect that for the ternary phase in this limit, allowing some solubility of Ta in SnO$_2$.

**Bond relaxations.** Projections of the final structure in one of three planes (the $a\overline{c}$ plane) are considered for analysing lattice distortions. With this view the rutile octahedra are well visualised.

The ground state structure (GS) with Ta:Sn = 3:1 features alternating double layers of Ta and of mixed Sn and Ta in the $a\overline{c}$ plane, see Table (6). The overlapping polyhedra within such a double layer is referred to as 'A and 'B. In both the Ta and Sn polyhedra there is an observable readjustment of the bond lengths. The two types of bond- apical and equatorial, referring respectively to the longer two bonds, and the four shorter ones that are perpendicular, are affected in different ways depending on the layer. In the case of Sn, the equatorial and the apical bonds both shorten by around the same proportion. For Ta, the equatorial bonds compress while the apical bonds lengthen, without exception. Particularly significant for the Ta polyhedra is the difference in magnitude of bond length changes between the two pairs of equatorial bonds that are offset in the $\overline{c}$ direction, by one order of magnitude in all cases. The lengths of both the shorter and longer apical bonds are periodic, with approximately the same length of each given bond in two octahedra that are separated by six layers in $\overline{c}$. The Ta apical bonds are in all cases shorter than in the initial structure, also by roughly one order of magnitude, but there is little variation between layers.

The most significant observations to draw from this analysis are that both the Sn-O and Ta-O bond lengths are affected by the doping and that there is a strong axial dependence to the
Table 6: Lengths of all Sn-O and Ta-O bonds throughout 6 layers in $\bar{a}b$ for the GS. Each layer is the direction of decreasing $c$. All lengths in Angstrom.
relaxations for the Ta polyhedra.

**Electronic Structure.** The band gaps for all the ground state structures of the CE were calculated. In all cases, localised states are present inside the band gap. The density of states (DOS) projected onto the atomic species (Ta,Sn,O) and the electronic orbital type (sp hybridised and d orbitals) demonstrate that these are d states belonging to Ta. This is shown for the structure with Ta concentration = 50% in figure (6.13), chosen for presentation because as will be explained, it represents the case where the effect of the doping on the electronic structure is the strongest. It could also be argued that the state in the band gap, Figure (6.13) represents one of the Ta–O bonds, since there is some projection of the DOS onto O. However, from the preceding analysis of the crystal structure, there is electron localisation on Ta causing bond deformations throughout the lattice, suggesting that electrons are donated to the system by the introduction of Ta with corresponding defect states in the band gap.

From the literature review of crystal field effects in section (4.5.3) the position of d electron states in the electronic structure depends on CFS, related to the crystallographic parameters (primarily the coordination, see section (4.5.3)), but also potentially on crystal geometry changes through JT. The optimisation of the geometry of the structure by which the space group symmetry and the volume can change, thereby permitting any possible reduction in the potential energy facilitated by Ta doping, can be switched off and the wavefunction associated with just...
Ta substitution without structural relaxation can be analysed. The bandstructure, again for ground state with Ta : Sn = 1 : 1 is presented, shown in figure (6.14). There remain states clearly identifiable as donor levels but they are shifted closer to the bottom of the conduction band (CB) compared to the case where the geometry is free to change. In addition, the bands show appreciably more dispersion for the unrelaxed case, and accordingly the Fermi level lies within the states, rather than corresponding to the top of them while the effective band gap is significantly reduced. Since the structure in question has a 1:1 composition of Ta and Sn, the density of states was calculated to examine the contribution of both atomic types to the projection. The states are mostly derived from Ta, to be expected since it is $d$ electrons that are affected by JT distortion.

### 6.2.5 Spin density and electron localisation

The physical influences driving changes in property of the disordered alloy are expected to be determined significantly by the interactions of lone electrons that can be more finely deduced.
by analysis of the spin density, or equivalently the difference in the density of $\alpha$ and $\beta$ electrons calculated by CRYSTAL. Important to reiterate is that interactions between spin-up and spin-down electrons are not treated explicitly; instead, the quantity analysed is the charge density of singly occupied orbitals that occur where the $\text{Ta}^{4+}$ oxidation state is adopted. To visualise this, the density $|\alpha - \beta|$ is calculated on a 3D grid of 50 points spaced evenly along each lattice vector. The format of this data is converted so that it can interpreted and displayed by visualisation software ParaView.

From the spin density, a 3D isosurface is determined in ParaView. The isosurface is tuned to focus on spin localised within the radius of single Ta sites. One observes in figure (6.15) with good clarity the $d$ crystalline orbitals $d_{xy}, d_{xz}, d_{yz}$ ($t_{2g}$ states), which have their conventional spatial representations. Also important are the points of localisation and any directionality to the orbitals relative to reciprocal lattice vectors, which govern coordinated interactions between $d$ orbitals. In analysing these geometrical aspects one needs to distinguish between the interrelated crystal field and Jahn-Teller effects.

**Figure 6.15**: **Left**: Spin density for the CE structure in the Ta lean limit with composition $\text{Ta} : \text{Sn} = 1 : 5$ with view parallel to the (111) plane of the untransformed rutile lattice, with view chosen to give an orthogonal projection. **Right**: For the structure in the Ta-rich limit with composition $\text{Ta} : \text{Sn} = 3 : 1$, as viewed in the (00$n$) plane ($n = \frac{c}{2}$) with an orthogonal projection.

Figure (6.15) shows spin density in both the Ta ‘lean (doping concentration of 16.667%) and Ta ‘rich limits (alloying fraction of 75%). Studying first the low-Ta (lean) limit, the view is parallel to the (111) plane of rutile. The cations seen in this view thus lie in the (110) plane, with spin localisation on the Ta towards the centre of the plane. On closer inspection, the orbital is directed between the Ta-O ligands signifying that it occupies one of the three $t_{2g}$ states ($d_{xy}, d_{xz}, d_{yz}$) that are reduced in energy due to crystal field splitting, as discussed in section (4.5.3). Figure (6.16) shows that there is pronounced distortion of the apical bond $a_1$. In addition, the orbital at the could be argued to have undergone a rotation in the equatorial plane, associated with a distortion of one equatorial bond $c_1$. These are both characteristic of small continuous distortions of orbitals. This suggests further that distortions involve numer-
6.2 Ta-doped SnO$_2$

Matthew Worsdale

Figure 6.16: Spin density around the Ta-O polyhedra shown in the Ta-lean limit on figure (6.15). **Left:** Indicated are the apical bond $a_1$ and equatorial bond $c_1$ that both appear to have distorted; **Right:** The same equatorial bond, with apical bond out of the page.

ous modes. It is not immediately clear how this makes it easier to localise the electron at the site compared to the undistorted octahedral geometry. One interpretation to draw is that it is preferable to have full occupation of one $t_g$ state rather than mixed occupancy across the three $t_g$-states, suggesting that JT is the dominant effect in the low-Ta limit, promoting electron localisation. In this regard, the difference between the Ta-lean and Ta-rich limits should be the degree of JT distortion, by which one of the three $t_g$ states becomes increasingly favoured for occupation by lone electrons compared to the other two. Identifying the distortion modes requires further information about the crystal normal modes across the structures of the cluster expansion.

In the rich limit there are also more possibilities for occupancies of the individual $d^n$ levels, which could make for multiple JT distortions. The appearance of a $t_g$ state is observed again, however the alignment of the spin varies between parallel to the $yz$-plane and a direction around 45° with respect to it. The reason behind the spin localisation at the particular Ta sites is not clear from this result alone, and could become clearer in theory by studying a larger supercell with the same average composition and symmetry though the ordering of Sn and Ta would likely change making comparison difficult in practice. It is clear that the total spin density is increasing with the Ta content of the structure, which supports that the number of Ta$^{4+}$ sites is increasing, implying JT distortion rising in magnitude concurrent with an increase in the donor activation energy. This is consistent with figure (6.17) showing the activation energy as a function of Ta loading, discussed in the next section. As Ta sites increase, the proportion of those that are JT 'active' and their distribution is also continuously changing, which determines the alloy formation energy and donor activation energy. However, the symmetric distribution of the active sites must give a self-reinforcing JT distortion. This condition is given by the alignment of the spin density in the crystal and the Sn – Ta ordering within that reference frame. The spin density data alone cannot show this, which would require analysis of the normal modes of rutile SnO$_2$ and TaO$_2$ [176]. These concluding remarks are developed in the outlook.

117
6.2.6 Donor activation as a function of concentration

The energy difference between the donor level induced by Ta and the CBM is plotted as a function of the Ta doping concentration in figure (6.17). The lowest concentration of Ta of the doped structures in the CE is only 16.67%, as can be seen in Figure (6.12), which makes it important to investigate the low-Ta limit. As the composition $x$ in $\text{Ta}_x\text{Sn}_{1-x}\text{O}_2$ approaches the smallest values, the system modelled increasingly approximates the limit of an isolated Ta defect in bulk $\text{SnO}_2$. In this regime, it becomes desirable to calculate the system by generating a supercell from the original rutile and manually choosing a site for the single Ta atom introduced—because with the CE the probability of generating a structure with extremely low Ta concentration becomes increasingly small, requiring greater computation time. A lattice constrained geometry optimization is employed, as was used in the first CE of which the results are shown in Figure (6.9), because the supercell is periodic and consequently the DFT code will introduce spurious interactions of Ta with its images throughout space. Consequently the interactions between the Ta $d$ states and O ligands that are the root cause of CFS and JT effects could introduce an erroneous cooperative effect that distorts the lattice substantially, if JT is influenced by changes in crystal geometry and particularly if the effect scales substantially with changes in the volume of the lattice. The 'true' isolated Ta in $\text{SnO}_2$ bulk could still feature CFS and JT but these may not be appropriately modelled by supercell based DFT calculations; by constraining the geometry of the unit cell, most relaxations that are driven by long distance effects are truncated, minimising propagation of error due to interactions between periodic images.

The supercell calculations were undertaken by substituting one Ta atom in supercells constructed as $n \times n \times n$ enlargements of the $\text{SnO}_2$ unit cell for $n = 2, 3, 4$. The depth of the defect state in the band gap was obtained as the function of supercell size approaching the dilute limit, for which the extrapolated donor depth found is 1.925 eV. In all cases, the uppermost state in the band gap corresponds to the Fermi level. The value is 1.83 eV for $n = 2$, rising as supercell size increases and converging, with the energy of 1.92 eV for Ta concentration of 0.78% being a good approximation to the value of $|E_C - E_D|$ in the dilute limit. If the dilute limit could be taken as a continuation of the structures in the CE, then the energy would be decreasing to the minimum value of 1.925 eV over the range of the supercell approximation. In fact, it is increasing throughout the range. However, the dilute limit with the supercell approximation is not equivalent to the cluster expansion because in the former, the original supercell lattice parameters cannot change, including volume changes that preserve the parameters. Because of this fix, intended to prevent mutually reinforcing distortions of the lattice by images of Ta, the local relaxation of the lattice around the isolated Ta (which effectively constitutes a point defect) will also be excluded. This defect can distort the lattice by a few different mechanisms through the defect-defect interaction including short range electronic interactions, and indirect effects such as lattice distortion (JT can contribute in both ways). As the supercell size increases the finite sized defect-defect interactions become less significant, and hence the convergence towards the approximate dilute limit energy that would be the same in both the cluster expansion and the (correct) supercell approximation. [h] The variation of the energy difference between the highest defect state and the CB as a function of Ta concentration is shown in figure (6.17). The
Figure 6.17: Depth of the defect state $E_C - E_D$ introduced by Ta in the band gap which may act as a donor state, as a function of Ta concentration. The range from 0-16.67% has been calculated in an approximation of the dilute limit (isolated point defect in bulk rutile SnO$_2$, with preservation of rutile symmetry enforced). The range with Ta > 16.67% has been calculated with the cluster expansion (CE). One can see that the CE can be directly extrapolated to the dilute limit, the point where the depth of the level has converged within the supercell approximation of that limit. The energy of the SnO$_2$ intrinsic oxygen vacancy defect Vac(O) of 2.5 eV is also indicated; for $|E_C - E_D| < $ Vac(O) Ta doping will not be feasible.
6.2 Ta-doped SnO$_2$

Matthew Worsdale

Figure 6.18: Depth of the defect state as calculated in the supercell approximation to the dilute limit. Second-order polynomial fit was used to fit the points for the three low Ta concentrations calculated using 2 x 2 x 2, 3 x 3 x 3 and 4 x 4 x 4 supercells.

The reduced energy difference between the highest defect state and the CB is found to decrease as one gets closer to the limit of isolated Ta in bulk SnO$_2$. However, the values of this quantity for the structures in the supercell approximation are lower than for the structure with the lowest Ta concentration in the CE, as can be seen in figure (6.18).

The reduced energy difference $|E_C - E_D|$ in the dilute limit means a lower activation energy, which favours higher electronic conductivity (see section (6.2.8)).

6.2.7 Carrier Statistics of extrinsic doping

The conductivity of a sample is proportional to the concentration of charge carriers. At higher doping concentrations, corresponding to alloying compositions of Sn and dopant, the distributions of the dopants will be inhomogeneous and the physics of cation interactions is continually changing with the substitutional doping. The concentration of ionised defects needs to be calculated, which will obey an expression of the form

$$N_{D^+} = N_{D^0} e^{\left(-\frac{(E_C - E_D)}{k_B T}\right)}$$

where $N_{D^0}$ is the concentration (cm$^{-3}$) of Ta in SnO$_2$, $E_C$ is the conduction band minimum and $E_D$ is the donor level (at the highest point of the band). This expression is modified to express the variation of the ionised donor concentration $N_{D^+}$ as a function of Ta concentration $x$, where $x \leq 1$,

$$N_{D^+}(x) = N_{D^0}(x) e^{\left(-\frac{E_A(x)}{k_B T}\right)}$$

120
6.2 Ta-doped SnO$_2$

Matthew Worsdale

Here, the activation energy $E_A(x)$ is obtained by a linear fit of activation energy $|E_C - E_D|$ to Ta concentration in the supercell approximated dilute limit and $N_{D^0}(x)$ is similarly obtained by a linear fit of the concentrations of Ta in the supercells, which are of the order of $10^{20}$ cm$^{-3}$. The concentration of donated conduction band electrons at room temperature as a function of $x$ was calculated using (94) with the result shown in Figure (6.19). For doping concentrations $x \leq 0.014$ the linear increase in donor concentration dominates in Equation (93) whereas for $x > 0.014$ the statistical likelihood of excitation of donated Ta decreases exponentially due to the increasing activation energy $E_A$. The assumption that these are the only two factors to take into consideration is motivated by the high activation energy which limits the donation of carriers from an isolated defect, with the peak occurring for a Ta concentration of close to 1 %. Only reductions in $E_A$ that also maintain or increase the dilute Ta concentration have the potential to increase electron concentrations in the conduction band beyond $10^{20}$ cm$^{-3}$, within the orders of magnitude needed for extrinsic $n$-type dopants as shown in table (2). The activation energy might decrease in the high Ta (low Sn alloying fraction), which is implied by the extrapolation of the result for $E_A(x)$ across the full range of $x$ shown in figure (6.17). Insufficient structures were found and computed in the high Ta limit to complete the plot, and would need computation of isolated Sn in bulk TaO$_2$.

To summarise, a higher carrier concentration could result from a combination of lower activation energy and higher Ta concentration, and appears to be limited by JT activity of the Ta$^{4+}$ state. The charge transport ultimately resulting from the activation of the band gap donor state also depends on the states in the conduction band that accept the promoted electrons, the position of which has not been precisely defined in this work (it is assumed instead that charge carriers are donated from the highest occupied states, the donor levels, to the CBM). Because the B3LYP
6.2 Ta-doped SnO

The functional has correctly predicted the position of donor levels within the band gap, there is no need to consider effects of electron concentration on band filling such as Moss-Burstein effects.

6.2.8 Conductivity

Conductivity can be estimated from the carrier concentration by following the assumptions of the Drude model explained in section (5.2). Combining (62), (68) and using the expression (94) for the concentration of electrons in the conduction band \( n_e(x) \) leads to the expression for the conductivity of the Ta/Sn alloy as a function of \( T \) and alloying fraction \( x = [\text{Ta}]/[\text{Sn}] \) in the dilute limit \(([\text{Ta}]/[\text{Sn}] << 1)\),

\[
\sigma_T(x, T) = e^2 \tau \frac{n_e(x)}{m_{\text{eff},k}(x)} = e^2 \tau \frac{N_{D_0}(x)}{m_{\text{eff},k}(x)} e^{\left(\frac{-E_A(x)}{k_B T}\right)}
\]  

(95)

The effective mass \( m_{\text{eff},k}(x) \) is taken to be dependent on \( x \) through changes induced in the dispersion of only the conduction band minimum, as a first approximation. For the dilute \( n \)-type Ta-Sn alloys, the value of \( E_F \) (at low temperatures) is independent of \( T \) and takes a position between the donor level and the CB, decreasing linearly with the number of impurities \( n_{\text{imp}} = N_{D_0}(x) \) due to the linear decrease in the donor level, \( (E_C - E_D(x))/2 \), as shown by figure (6.17). Hence the value of \( \hbar^2 n_{\text{imp}} \) (equation (67)) is independent of the number of Ta impurity atoms. Also for the low temperature limit, the value \( E_F(x, T) \) is independent of \( T \) and varies linearly with \( x \) (assuming Fermi-Dirac statistics and no change to the energy of the CBM \( E_C \)). Therefore, the electron-phonon and Coulomb contributions to the scattering will be the most significant for the alloys, limiting the conductivity proportionally with temperature. The conditions that are of interest hence correspond to low Ta/Sn fractions since for high values of \( x \), thermal excitation of electrons to the conduction band is suppressed by the exponential factor in equation (95). The rapid decrease in the position of the donor level in the band gap \( E_D(x) \) with \( N_{D_0}(x) \) means that even for temperatures hundreds of degrees \( ^\circ \text{C} \) higher than the working temperatures of the PEFC of \( T \leq 350 \text{K} \), the carrier concentration in the conduction band is of the order of \( 10^{20} \text{cm}^{-3} \), while the conductivity overall is lessened at increasing temperature by the scattering contributions. Thus despite not having estimated the effective mass \( m_{\text{eff}}(x) \), it can be stated with high confidence that sufficient levels of conductivity could not be expected for the electrocatalyst Ta\(_x\)Sn\(_{1-x}\)O\(_2\).

6.2.9 Discussion- Effect of symmetry restriction in cluster expansions

There is a first-principles argument for the constraining of symmetry being responsible for the poor fit between \( E_{DFT} \) and \( E_{CE} \). Figure (6.11) identifies the particular PGO of each structure
in the CE. It can be seen that there are several groups of structures that all have the same PGO. However, the structures of course are also distinguished by the composition Ta:Sn. When the structure optimises its geometry in the DFT energy calculation, the atomic positions are free to change, under the combined influence of the interactions between Ta and Sn, and the symmetry of the system. Forcing the structure to obey rutile symmetry places constraints on the way in which its geometry may change. A higher number of PGO in the original structure effectively reduces the degrees of freedom in relaxation since there are more constraints on the final atomic positions. From structure to structure, there is a lot of variation in the number of ways the geometry can change while maintaining a set of PGOs. There will be groups of structures that have one or more PGO in common, which one can think of as the relationship between structures in the CE that changes as they optimise geometry. For some groups, the structures may be randomly distributed across the composition space, in other cases, the majority occur for one composition. At the same time, if the constraint of rutile symmetry were removed, the initial structure may be able to relax into a final geometry that is more stable than the best it can do if rutile symmetry must be kept.

The value of $E_{CE}$ is based on a fit of energies of a sample set of CE structures drawn from the whole set, as explained in section (5.6.2). Different configurations of Ta and Sn are impacted differently by the symmetry constraint, and those structures whose true energy $E_{DFT}$ corresponds to a fitted energy $E_{CE}$ of the other $n-1$ structures of its sample group, that have more symmetries in common, will likely be an outlier for that group, so the correlation between $E_{CE}$ and $E_{DFT}$ will be poor. In other cases, the $n-1$ fitting structures will have a more uniform distribution of energies so $E_{DFT}$ is less likely to be an outlier in the group. The poor representation of the relaxation of certain structures is an inherent problem of the symmetry constrained CE, and so it must be replaced by one where both the atomic positions and the periodic unit cell is allowed to change.

### 6.2.10 Crystal symmetry and Jahn-Teller effects

**Overview.** Following the discussion in (4.5.3), energy gaps may be introduced by CFS between those $d$ orbitals of Ta having a spatial distribution aligned geometrically with the Ta–O ligands. The resultant level structure of the $d$ states depends on the octahedral coordination of Ta with O, in other words, the energies where the two sub-levels created lie compared to the original $5d$ states. The JT effect on the other hand has the consequence of breaking the degeneracy of the states lying at the two sub-levels by distorting the lattice or electron orbitals. The effect depends on the degrees of freedom of the relaxation of the structure, governed by the point symmetry operators possessed by the structure within the relevant space group, properties which are unconstrained in the results documented in sections following (6.2.4).

In the cluster expansion, doping with Ta will have two basic consequences: the addition of $d$ states, which may donate electrons to the system, and the change in the ordering of Sn and Ta, hence the PGO of the structure. Additional consequences may also follow from a JT distortion. An important question is whether any of these factors should change with the number of $d$
6.2 Ta-doped SnO₂

Matthew Worsdale

electrons introduced into the system. This may depend on the electronic structure and the position of the \( d \) bands relative to the conduction band and the position of \( E_F \) as a function of electron concentration. Equally, the structural optimisation that goes with doping can be expected to influence the JT effect, which acts on \( d \) states that are made degenerate by CFS. The effect of CFS on the \( d \) state level structure and its electron configuration will not change in alloying as modelled in this work. \(^2\) A lesser number of PGO potentially allows more lattice distortions related to local point group symmetries of the \( d \) orbitals \([210]\), in the case that the number of operators changes within one space group. Also, the extent of JT could depend on the dispersion of \( d \) bands which is a consequence of their localisation and mobility throughout the lattice.

In summary, a higher Ta loading provides more \( d \) states in the band gap, and the JT distortion is determined by the position of \( E_F \) and the structural considerations (Ta-Sn ordering and symmetry).

Electronic structure. When the Ta is added to the Sn based oxide, the Ta and Sn are both forming covalent bonds with O, and there is a significant interaction between the orbitals of the two cations, which results in the dispersion of both of the states. The process of stabilising Ta energetically through the geometry relaxation both reduces the dispersion and splits these two \( t_g \) states with one of them promoted slightly in energy to around the CBM, and the other reducing in energy. This state, from being a perturbed host state (PHS) \([199]\) apparently derived from the CB, becomes a deep donor level bound by the \( t_g \) electrons whose periodic components increasingly grow in magnitude. The crystal field energy (CFSE) that separates the \( t_g \) and \( e_{2g} \) states \(D_q\) can change only due to coordination of the cations in the oxide, or their ionic radius. Since the former is reasoned to be a constant within this problem, and the ionic radii of Sn and Ta are very similar, the effect of CFS probably does not change significantly with either the relaxation of the structure or the Ta concentration. Preventing relaxations, as was done to obtain figure (6.14), should therefore distinguish the JT distortion. In this regard the difference between Figures (6.13) and (6.14) forms a sufficient explanation of the origin and disappearance of active donor levels with Ta doping in SnO₂.

The JT effect could be only a local, or alternatively a cooperative, extensive phenomenon in the system studied. The JT lattice distortion may cause an (unidentified) symmetry breaking with a reduction in the number of states that are equivalent by symmetry, hence splitting of degenerate states. The symmetry breaking is a step change effect, in which the whole structure changes its geometry. However, JT can occur through any geometry change that provides some reduction in degeneracy. Therefore continuous changes in bond lengths, or continuous distortions of the electron orbitals, may occur without changes to the PGOs of the crystal.

In the ground state structure, the predominant character of the bond relaxations is alternating compressions and elongations in the \( \hat{c} \) direction. The density of the unpaired electrons in the

\(^2\)This would require the coordination of Ta with O to change which would only happen in the case of phase transition from rutile, for example to the competing ternary phase SnTa₂O₆, though this is not permitted in the cluster expansion \([232]\) so that what disordering between Sn and Ta that is coincident with the transition to the ternary in reality, potentially represents to an extent the constraints arising from the inhibited relaxations (by similar reasoning to that given in section (6.2.9) for the symmetry restricted CE) if the phase transition is possible.
lattice, figure (6.15) suggests that the localisation and alignment at particular sites could drive lattice distortions, possibly introducing a directionality to distortion resulting from JT over a certain length scale. The incorporation of the Ta into the structure gives rise to the effective donor state in the band gap. The position of $E_F$ determines the occupation of the donor state if it is disperse enough that the position is within it, as in the example shown by figure (6.14), but at zero temperature the position is taken to be halfway between $E_D$ and $E_C$. Aside from changing occupation of bulk states across the doping concentration range, the important consideration behind the potential conductivity due to donated electrons is therefore the energy difference between the conduction band minimum and the donor state, $|E_C - E_D|$. The efficacy of Ta as a donor depends on two possibly interrelated factors,

- The position of the Fermi level $E_F$ and band gap edges $E_C$ and $E_V$,
- The crystal field splitting and Jahn-Teller changes to the $d$ band positions.

Across the composition range of the CE, the value of $|E_C - E_D|$ increases up to the concentration of Ta=50% but then decreases at higher concentrations. However, increasing formation energy of higher-Ta alloys corresponds to the increasing structural distortion, which depends on electrons localised on sites that are Ta$^{4+}$. This could be a consequence of the local bonding environment of Ta that may change substantially through the progressive impact of distorting the lattice from the substitution of Sn with Ta.

Another possibility for the dominant effect is the interrelated changes of structure and Ta - Sn ordering. The doping can be thought of as continuously changing the environment of all $d$ states through states determined by the crystal PBC, seen in the bandstructure, while changes to the lattice geometry will affect some $d$ states more than others. Then there is the symmetry of the structure which is also changing continually with more Ta and subsequent lattice relaxations, which have no constraints on them. When the structures optimise their geometry, the way they distort initially is governed by the PGO (of rutile), then the PGO of other space groups become involved. At some concentrations, there will be more variation within the group of structures of the number of point group symmetry operators (PGO) they possess, than for other concentrations, which may have more bearing on the size of $|E_C - E_D|$ than the total number of JT active sites, hence the observation that the donor state position in the band gap actually rises (rather than continuing to fall) beyond a Ta concentration of 50% (note that this is not the same thing as the formation energy of the structure Ta$_x$Sn$_{1-x}$O$_2$, which has its maximum at Ta=75%). It is possible that this limit represents the peak JT activity. An alternative explanation is the nature of the defect state formed by the density of lone $d$ electrons, as in figure (6.15). At low concentrations the additional electrons on isolated Ta behave as a localised isolated defect. With increasing Ta concentration their coordinated interaction increases, but past a given point the well dispersed orbitals effectively form a delocalised state across the lattice, with lower efficacy for bringing a distortion.
6.2.11 Conclusions

At higher Ta concentrations, the Jahn-Teller effect increases through collective lattice distortions, which actually impedes electron delocalisation and prevents an increase in mobile conduction band electrons due to an increase in ordering over increasingly long length scales, despite the formal increase in disorder due to the swapping of Sn and Ta. With the advance of these effects, the JT drives the lowering of donor states in the band gap, and the spatial distribution of single unpaired Ta\(^{4+}\) has increasing possibilities and differing consequences in terms of potential minimisation. A further implication is differing possibilities for the localisation of spin, on account of the deformation of spin isosurfaces. This could be further studied in the low-Ta structure to give further insights into the determination of the \(d\) orbital geometry from the crystal structure and vice versa, in the limiting case where interactions between Ta become coordinated. It is emphasised that we have not considered spin and orbital ordering effects. In the latter, \(d\) state that is adopted (amongst \(d_{xz}\), \(d_{yz}\) and \(d_{xy}\)) has its lobes aligned with crystal directions, which could offer some explanation the active distortion modes.

Even for low Ta alloying fractions, JT effects drive electron localisation that promotes stability at the expense of conductivity. The higher the Ta content of the structure, the more likely it becomes that electrons localise on the states created in the band gap that are lowered in energy by CFS and JT, with Ta adopting a \(4^+\) state as a result. There is not evidence from the calculations, however, of the alternative \(5^+\) oxidation state enabling more electron donation due to electron delocalisation. If Ta is in the \(4^+\) state, its’ highest occupied state is one of the states in the band gap, and an electron can be promoted from this state through ionisation, with a probability that will increase with temperature. The drive to ionisation competes with the JT that lowers the energy of the structure through its effect on the \(d\)-orbitals only. The structures where this effect is less pronounced (more Ta\(^{5+}\)) have less energy associated with the disorder introduced by the CE and are expected to be less resistant to chemical attack. Without completing a Monte Carlo averaging over the CE structures, to calculate the curves of Gibbs Free Energy of the disordered systems, we can expect that increasing temperature will allow a structure having an arbitrary concentration of Ta to access configurations with more \(4^+\) states. In summary, there is a basic problem that the structures with the best electronic conductivity will be less persistent than those that have a lower conductivity (and are also less oxidation resistant).

The precise interactions between the \(d\) electrons that are responsible for the lowering of band gap donor states has to be identified in future work. The existence of JT active sites explains the stability and defect activation of the disordered alloys. It has been shown that these two fundamentally important properties are linked through JT, i.e increasing stability coincides with falling conductivity throughout most of the doping range. This neatly explains the maximum of conductivity at doping concentrations in the range of 0-5 % according to experimental work as discussed in section (4.5.2) for Ta and oxides of other transition metals that have similar electron chemistry.

A further recent experimental study reports conductivity trends in Ta doped SnO\(_2\) in precise agreement with these results [231]. Thin film SnO\(_2\) deposited with Ta by DC magnetron sput-
tering has maximum conductivity at around 1% Ta as found here. There is a gradual decline of around one order of magnitude over heavier doping of 1-4% Ta, which is consistent with the progressive increase in defect activation energy due to JT distortion seen in figure (6.17). In the range of 4-8% Ta, there is a sharp fall in conductivity down to levels more characteristic of undoped SnO$_2$. This can also be consistent with the results here, as it is plausible that over that range of doping and at the temperatures of 400°C used in the annealing of the films, the (small) solubility limit of Ta in rutile SnO$_2$ due to entropic effects, is exceeded, making the competing ternary oxide SnTa$_2$O$_6$ thermodynamically favourable. This would suggest that ease of formation of the ternary represents another constraint on producing electrically active SnO$_2$ through Ta doping in practice.

Even though the nature of the collective JT distortion, in terms of symmetry properties over varying length scales and/or the local electronic interactions confined to a Ta bonding environment, has not been explained fully in this work, theoretical works focusing on the JT effect propose feasible routes for extending the methodology used here to quantify it explicitly with analytical models and the outlook discusses these with brevity and their possible coupling to a DFT based methodology for defect calculations. The results, therefore, suggest that extrinsic conductivity in $n$-type oxides from group V elements, $M$, is not feasible because the coordination of M and O produces a subset of degenerate $d$-states that are frozen out due to the JT effect.

In a real sample of Ta$_x$Sn$_{1-x}$O$_2$, there could be the additional effects of mixed valencies of Sn and Ta and intrinsic defects. These could lead to effective over or under coordination of Ta and O at some sites with symmetry changes not considered here. By the methodology followed for the CE, valence states are pre-selected according to the minimal electrostatic potential energy configuration determined by the Ewald sum, and may then change in the course of the DFT energy calculation. However, Ta$^{4+}$/Ta$^{5+}$ valence changes in the cluster expansion independent of the site occupation are not considered given that combinatorically, the resulting configuration space would be very large.

In as much as the chemistry and electronic structure of Ta, Sn and O in a given valence and composition can be decoupled from the crystal structure in which it occurs, it would be desirable to have control over both the electronic conductivity and the thermodynamic stability by producing structures that have a given symmetry but not the electronic properties that lead to JT by preventing distortions at certain positions using extrinsic and/or intrinsic defects. The possibility would need to be evaluated in future work treating these possibilities explicitly.
6.3 Hydrogen defects in bulk rutile SnO₂

6.3.1 Stability of substitutional and interstitial defects

Two types of hydrogen defect are considered in this work: interstitial H, which bonds to O to form an OH group, and substitutional defects replacing an O by H. An important consideration is which defect type is more thermodynamically stable based on formation energy as a function of the constituent reference states \( \mu_H \) and \( \mu_O \) calculated by DFT at zero temperature. Because the interstitial defect bonds to an existing O whereas the substitutional defect depends on the removal of an O, only one of them can be stable in the bulk. To determine the phase that is stable, one starts with the formation energy written in normal form to easily identify the plane coefficients,

\[
\Delta N_O \mu_O + \Delta N_H \mu_H + qE_F - G_T^{def,q} + N_S G_T^{per,f,q} = -\Delta G_f(\mu_H, \mu_O)
\]

where the coefficients \( \Delta N_H = 1 \) and \( \Delta N_O = 0 \) for an interstitial and substitutional defect, respectively, \( qE_F \) is the contribution from the Fermi energy. \( G_T^{def,q} \) is the energy of the defect supercell as calculated, \( G_T^{per,f,q} \) is the energy of a stochiometric unit of SnO₂ with \( N_S \) the multiplicity of the unit cell chosen to make the undoped supercell commensurate with the defect cell. This equation comprises four unknowns, if one includes the \( E_F \). However because both defects are modelled as neutral, the term in \( E_F \) vanishes and the system is hence determined by two planes, one for both the substitutional and interstitial configurations. Six bounding planes are defined in the same space by limits on \( \mu_H \) and \( \mu_O \) normalised to the DFT energies of \( \frac{1}{2}H_2 \) and \( \frac{1}{2}O_2 \) respectively, while \( \Delta G_f \) takes negative values. The three-dimensional convex hull (or polytrope) that spans the space is calculated.

The polytrope is shown in figure (6.20) for the input values \( \frac{1}{2}\mu_{O_2}^{DFT} \leq \mu_O \leq (-\frac{1}{2}\mu_{O_2}^{DFT} + 5) \) and \((-\frac{1}{2}\mu_{H_2}^{DFT} + 18 \text{eV}) \leq \mu_H \leq 0 \text{eV} \) to show clearly all the phases. The interstitial defect, accordingly, is stable for 14.1 eV \( \leq \mu_H \leq 33.9 \text{eV} \), the lower limit being just slightly less than the energy needed to promote a hydrogen atom from its ground state as calculated by DFT. The upper limit corresponds to the disappearance of \( H_2 \) as a stable phase, making the three-dimensional minimum energy surface non-convex at that end point. Over part of the the stability range of the interstitial defect, the substitutional defect ('sub(H)') also forms a phase and can be more stable, depending on the values of both \( \mu_O \) and \( \mu_H \).

The lower bound of stability of the substitutional defect is \( \mu_O = -2050 \text{eV} \), just 4 eV lower than \( \frac{1}{2}\mu_{O_2}^{DFT} \). Below this limit, the phases are undefined with respect to each other. Increasing \( \mu_H \) sees the substitutional defect remain stable compared to the interstitial, up to the point where the phase boundary meets the \( \Delta G_f(O_2) \) plane if this is before it meets the \( H_2 \) plane, which is the case for \( \mu_O - \frac{1}{2}\mu_{O_2}^{DFT} \leq 3 \text{eV} \). Therefore, for \(-4 \leq \mu_O \leq 3 \text{eV} \) and \( \mu_H \mid G_f(sub(H)) = G_f(O_2) \leq \mu_H \leq \mu_H^{MAX} \), the substitutional phase vanishes, making the interstitial defect stable. If, however, \( 3 \leq \mu_O \leq 18 \text{eV} \), the substitutional phase again becomes stable. For increases of hydrogen and oxygen chemical from reference states of up to around 10 eV, then, smaller increases of up to around 3 eV of \( \mu_O \) will give one region of sub(H) stability for \( 0 \leq \mu_H \leq \mu_H \mid G_f(sub(H)) = G_f(O_2) \) and another region of OH-group stability for \( \mu_H \mid G_f(sub(H)) = G_f(O_2) \leq \mu_H \leq \mu_H^{MAX} \). For \( \mu_O \geq 3 \text{eV} \),
6.3 Hydrogen defects in bulk rutile SnO$_2$

Matthew Worsdale

Figure 6.20: Convex hull for the interstitial (OH group) and substitutional (‘sub(H)’, O replaced by H) defects as a function of chemical potentials $\mu_H$ and $\mu_O$. The polytrope is plotted over the ranges (eV) $\frac{1}{2}\mu_{DFT}^{\text{H}_2} + 18 \leq \mu_H \leq 0$, $\frac{1}{2}\mu_{DFT}^{\text{O}_2} + 3 \leq \mu_O$. The elemental chemical potential references, based on the DFT energies of H$_2$ and O$_2$ are respectively $\frac{1}{2}\mu_{DFT}^{\text{H}_2} = -15.9$ eV, $\frac{1}{2}\mu_{DFT}^{\text{O}_2} = -2046$ eV.

the substitutional defect is stable irrespective of the value of $\mu_H$.

Table (7) shows the formation energies of the defects described in the methodology for the fixed oxygen and hydrogen chemical potentials. Figure (6.21) shows the formation energy of the neutral and formally charged interstitial defect as a function of the Fermi energy, assuming a linear dependence of the formation energy on Fermi energy.

6.3.2 Defect lattices and bandstructures

Figure (6.22) shows the final lattice geometry in the vicinity of the substitutional and interstitial defect, respectively. In the interstitial case, the formation of a bond between H and O appears to cause corresponding distortions to the four nearest neighbour Sn – O bonds via the movement of O closer to the H to which it bonds. The distortions are comparable for the bonds in directions parallel and perpendicular to the Sn – O adjacent to the final H position, since the bonding O moves in the x and y directions to get closer to H. The resultant lattice distortion appears to be limited in spatial extent, perhaps because the stress associated with the bond shifts is distributed evenly between the normal modes along the apical and equatorial bonds.

For the substitutional defect, the distortions are also comparable for the directions parallel and perpendicular to the plane in which the H defect is substituted.
### Table 7: Calculated formation energies of interstitial and substitutional H defects in different charge state initialisations (formally neutral, except $H_{jel}^+$ which is a formally charged defect modelled with jellium). In these calculations, the reference chemical potentials correspond to the DFT energies of the dimers, $\frac{1}{2}\mu_{O_{2}}^{DFT} = -2046$ eV and $\frac{1}{2}\mu_{H_{2}}^{DFT} = -15.9$ eV.

#### 6.3.3 Electronic structure

For both substitutional and interstitial defect there is noticeable change to the band gap region, with partially occupied bands amongst a disperse manifold in the vicinity of the CBM, giving a metallic state at $T = 0K$ - rather than a semiconductor as in the case of intrinsic O vacancies and Ta doping. The bands that form the CBM are so disperse that at first glance one could guess that the Fermi level $E_F$ must intersect the band whatever the status of the factors involved concerning the H defect - including the defect concentration, the chemical potentials $\mu_H$ and $\mu_O$ and the (perhaps position dependent) local environment of the defect. However, the bandstructure of the formally charged $H_{jel}^+$ interstitial defect, shown in figure (6.24), represents the same electronic structure except that the CBM is not occupied making the system insulating. Thus the conductivity of the system appears to be related to electron localisation, which motivates the analysis of charge difference maps in section (6.3.4).

Closer inspection reveals subtle differences in the two cases. For the interstitial case, it can be seen that in addition to metallic conductivity due to the position of $E_F$ at $T = 0K$ crossing two bands, these bands overlap (in $k$-space). The parabolic character of the CBM highlights that for the $H^0$ interstitial case, the electron in this state will behave as a Nearly Free Electron (NFE). In the (formally charged) $H_{jel}^+$ interstitial case, the only significant difference is the occupation of the conduction band, suggesting that it is the behaviour of the additional electron modelled in the $H^0$ case that makes the system conductive and not the bond it forms with a host atom.

For the substitutional bandstructure in figure (6.25), the system is again metallic with $E_F$ intersecting a band that is near to the top of the band gap. However, this band does not appear to be part of the CBM as in the interstitial case, since on close inspection the band does not overlap with the one above it at any point in the BZ. From this perspective it is not clear whether it is derived from the conduction band of the host, or instead is a donor level just below the conduction band edge. There is a close similarity of the appearance of the conduction band
Figure 6.21: Formation energy of the neutral interstitial defect and the formally charged (jel-
lium) interstitial defect as a function of Fermi energy (referred to the VBM). This is calculated
based on equation (69) and assumes linear dependence of the formation energy on $E_F$, which
is potentially a deficient assumption given the electronic structure predicting $E_F$ inside the
conduction band. Implications of this are discussed in section (6.3.7)

dge in this case to that of the oxygen vacancy defect (figure (6.7)), in which the top of the
valence band is instead a deep donor level in the band gap. The main differences between the
two cases are that for substitutional H, the deep donor from the oxygen vacancy is not present,
and the state at the CBM is partially occupied to give a metal or semimetal. The position of
this state is clearly lower on an absolute energy scale in addition by approximately 1 eV. This
gives a reduction in the size of the band gap as in the interstitial case. The electronic structure
difference can be explained in terms of the addition of H to the oxygen vacancy. When H is
placed at the vacancy site, its single 1s orbital interacts with the electrons that remain from
the Sn-O bonds broken by creation of the vacancy. This gives a bonding state in the VB and
an anti-bonding state in the CB. Two electrons at the vacancy site form bonds with Sn (which
occupies the original deep donor level) while the remaining third electron (from H$^0$) is free to
occupy the H anti-bonding state in the CB. The combination (hybridisation) of the 1s orbital
and the host orbitals is generally assumed to give a resonance of the host orbitals rather than a
distinct electronic state, with the result that the lowest of CB states becomes partially occupied.
By this reasoning, the substitutional and interstitial-H doped SnO$_2$ are both metallic essentially
because there is an additional electron from H$^0$ that simply changes the occupation of bands
without substantially modifying the host electronic structure. However, particular details of the
 electronic structure still need to be discussed, and some have already been noted, for example,
6.3 Hydrogen defects in bulk rutile SnO$_2$

Matthew Worsdale

Figure 6.22: Nearest neighbour bond lengths for the substitutional H defect (left, figure (6.22 a)) and interstitial defect (right, figure 6.22 b). For comparison, the undistorted apical bond length is 2.14303Å, and undistorted equatorial bond length is 2.12706Å.

the overlapping of first two bands of the CB in the interstitial case.

The relation of states in the low conduction band to the orbitals of H can become clearer by studying the density of states, which are plotted against the bandstructures. In terms of the hydrogen projection, this is predominantly coincident in the interstitial case with the Sn conduction band projection, whereas in the substitutional case, it largely coincides with the O valence band projection. Only a small proportion of the total density of states exists around $E_F$, and the projection onto the H itself is minimal in both substitutional and interstitial cases. The DOSS suggests that the formation of these states is a question of the distribution of the disperse bands in the vicinity of the conduction band minimum.

The interstitial H is bonded to O as seen from figure (6.22), but the bowed appearance of the CBM in the bandstructure suggests strong polarisation of the bond characteristic of an ionic bond. The band gaps for the charged H$^+_{\text{jet}}$ and neutral H$^0$ interstitial defects are 3.1 eV and 2.8 eV, considerably smaller in both cases than the 4.08 eV for undoped SnO$_2$, which can be explained by bonding of H$^+$ to O$^{2-}$ to form an OH$^-$ group, reducing the negative charge of the effective anion and the electrostatic attraction to Sn. If hydrogen is modelled as a neutral atom, the resulting state forming the CBM is partially occupied by the surplus electron that is not required to participate in the bond between Sn and O. If hydrogen is modelled as H$^+_{\text{jet}}$, there is no surplus electron but the singly negative charged anion resulting from H$^+$ and O$^{2-}$ does not change. The slightly greater band gap in this case (3.1 eV as opposed to 2.8 eV) indicates that the electrostatic interaction between the O,H combination and Sn compared to just O is not diminished as much as the neutral case, likely because the localisation of the electron in the neutral case enhances the polarisation of the bond. The deepening of the conduction
Figure 6.23: **Right**: Bandstructure for bulk interstitial H\(^0\) defect in rutile SnO\(_2\) with atomic state shell initialisation. DFT calculation performed for one hydrogen atom in a 48 atom cell of rutile SnO\(_2\) assuming defect position ‘A’ defined in section (5.8.1). **Left**: Total and hydrogen projected density of states (arbitrary units).
Figure 6.24: Bandstructure for interstitial $\text{H}_{\text{jel}}^+$ defect in rutile $\text{SnO}_2$, modelled as formally charged with a jellium background. Unit cell neutrality was enforced by redistributing the charge of the single hydrogen electron equally amongst all other tin atoms, with all other computational details the same. Density of states (not shown) is identical to that in figure (6.23).
6.3 Hydrogen defects in bulk rutile SnO$_2$

Matthew Worsdale

Figure 6.25: **Right:** Bandstructure for H$^0$ substitutional defect. DFT calculation performed for a 48 atom cell of rutile SnO$_2$ with one of the 32 oxygen atoms substituted by one hydrogen. Unit cell neutrality was maintained by redistributing the charge of the two electrons left at the site of the replaced oxygen atom equally amongst all tin atoms. **Left:** Total and hydrogen projected density of states (arbitrary units).
band minimum state and consequential reduction in the band gap is also apparent for the substitutional defect if one compares it closely to figure (6.7). By the same logic, substitutional H is forming a bond of ionic character with an Sn instead of an O which needs to be substantiated by examining the charge density in the vicinity of the defect.

6.3.4 Electron distribution- charge difference maps

Because the bonding between H and the lattice potentially affects the position of the bands that are related to H, the charge difference map is useful for visualising the localisation of the charge density. The maps are obtained, as discussed in section (4.3.10), as the difference between the final relaxed lattice with the H defect, and the same lattice with the H defect removed but the final geometry kept the same. The first difference map shown in figure (6.26) is for the substitutional defect. For this defect, analysis is made difficult by the fact that the substitution of H involves two stages, the removal of O, followed by the implantation of H. The large and mostly spherical isosurface surrounding H clearly corresponds to the core 6sp electrons of O (negative difference in charge density). The other movements of charge are to the Sn nearest to H on one side and a distribution on the opposite side of uncertain interpretation. The charge transfer to Sn indicates its bond with H. The localisation of the additional electron that is not needed for the bond is not immediately apparent.

The difference map for neutral interstitial hydrogen $H^0$ is shown in figure (6.27). The resolution of the isosurface was selected to focus only on the position of the H and its immediate vicinity. There is net movement of charge density to the site of the H (cyan isosurface). The shell of charge which surrounds the H site itself represents an sp-orbital produced from the hybridisation between O and H, while the additional charge shell that has moved towards the H but lies around the neighbouring O is most likely signifying the d-type shell that is used for polarization of the O orbital shell in the basis set, while at the other side of the O is (isosurface in blue) a contribution of charge that has moved away from the vicinity, formed as some superposition of atomic orbitals. The elongated appearance of this contribution signifies the polarisation of charge on O and H due to the largely ionic character of the bond.

In itself, the presence of charge at the hydrogen site is inconsistent with the DOSS for interstitial $H^0$, figure (6.23), which shows the H states to be above $E_F$ and hence unoccupied. However, representing the charge density in the lattice using the difference map makes it difficult to distinguish charge movements at the hydrogen from those in the lattice over a given scale. For a ‘neutral’ defect, H is initialised with a single s-electron \(^3\) to be equivalent to its atomic shell configuration, and the occupation of this orbital is free to change to give H\(^+\) and H\(^-\) charge states, though only within the constraint of global charge neutrality (to be distinguished from the case of a formally charged defect neutralised with jellium H\(^+\)\(_{jel}\)). When modelling interstitial $H^0$, therefore, the electron from H will redistribute in addition to the changes in electron density in the locality of the atom. The difference map of a formally charged H\(^+\)\(_{jel}\) was also calculated,

\(^3\)A 3p GTF is included in the basis set to give a degree of freedom to polarize the state as described in section (8.1)
6.3 Hydrogen defects in bulk rutile SnO\textsubscript{2}

Matthew Worsdale

Figure 6.26: Electron density difference map plotted as isosurface (xz plane) for the substitutional defect in the initialised neutral charge state H\textsuperscript{0}. Blue isosurface components correspond to negative density difference, cyan to positive density difference. The right side image shows the immediate environment of the substitutional H site (white circle).

figure (6.28) with the same procedure, for comparison with H\textsuperscript{0}.

Finally, since the final position of the intersitial H is commensurate with an OH\textsuperscript{−} group, a difference map for the molecular group was computed as the difference between O\textsuperscript{2−} and OH\textsuperscript{−} initially placed 1 Å apart and then fully relaxed. The resulting contour shown in Figure (6.29) closely resembles that of the difference maps for the removal of H from the lattice, suggesting that these are reflecting the formation of OH\textsuperscript{−} from a bulk H defect.

6.3.5 Conductivity

For the case of interstitial hydrogen (OH groups) in SnO\textsubscript{2}, it is clear from figure (6.23) that the electrons in the conduction band, which is occupied at zero temperature, behave as "nearly free" electrons due to the disperse nature of the CBM. This means that the system can be treated as a metal, with the concentration of carriers in the conduction band as a function of the energy E of the highest occupied state in the conduction band. Assuming again that the Drude model is valid, an expression of the same form as equation (95) estimates the conductivity of the hydrogen doped system \( \sigma_H \),

\[
\sigma_H(x, T) = \epsilon^2 \tau \frac{n_{e^{-}}(E)}{m_{eff,k}(x)}
\]
The occupation of states in the conduction band is a function of the energy of the state and the density of states $g_{DOS}(E)$ and the appropriate expression for the number of electrons $n_e^-(E)$ in the conduction band is of the form

$$n_e^- = \int_{E_1}^{E_2} f_{FD}(E) g_{DOS}(E)$$

and where $f_{FD}(E)$ is the Fermi-Dirac distribution function,

$$f_{FD} = \frac{1}{e^{\frac{E - \mu}{k_B T}} + 1}$$

Looking at the contributions to the scattering in equation (67), at low temperatures the scattering of the conduction band electrons by phonons and by Coulomb scattering is small. In addition, the parabolic shape of the conduction band at the CBM will give a low value of the electron effective mass $m_{e\text{ff},k}$, though the variation as a function of $x$ cannot be determined since the calculation of all hydrogen defects was limited by time constraints to just one supercell size. The conduction band concentration $n_{e^-}$ will increase with temperature, as the occupancy of states increases as a function of $T$ for $E > \mu$ with $\mu$ the Fermi energy at zero temperature. At higher temperatures, the Coulomb and phonon scattering become significant, and if the concentration of hydrogen atoms is additionally increased to raise the concentration of conduction
band electrons in the material, increased impurity scattering also must be taken into account. The behaviour of the system overall at higher temperatures depends on whether the heat capacity of the doped SnO$_2$ is dominated by phonon contributions of the rutile crystal structure, as the temperature will increasingly activate the phonon modes of the semiconductor leading to increased phonon scattering, while the significance of the Coulomb scattering depends on the relative magnitude of $T^2$ and $E_F$. The conductivity of the system could therefore be limited even though the electron concentration continues to increase due to the position of $E_F$ in the conduction band.

### 6.3.6 Discussion-Defect Stability

With variation of $\mu_O$ and $\mu_H$ the substitutional and interstitial defects are in competition with one another so a sample would be expected to have populations of both under varying thermodynamic conditions. The substitutional defect has the lower formation energy over most of the variation of $\mu_O$ over a range of $0 - 10$ eV (relative to the DFT energy of O$_2$) with a window in which the phase vanishes and the interstitial defect is made stable over a range of $\mu_H$, as explained in section (6.3.1). However, this outcome could also depend on the difference in energy between a neutral and a charged interstitial $H^{+}_{jel}$. In fact, $H^{+}_{jel}$ is predicted to form spontaneously based on the bare total energy value from the DFT calculation. However, this value excludes the corrections to the total energy that should be included when modelling a formally charged defect, amongst which the largest contribution is the Madelung energy (c.f equation (75)) which is normally negative [192]. Uncertainty remains also in respect of the
potential reference that should be used to calculate the formation energy, because $E_F$ lies in the conduction band, making the electron reference energy ambiguous (in addition to the reference energy of the CBM, as discussed in section (6.2.10)). For a localised impurity such as H its mobility and ease of migration are likewise of importance for stability but not calculated here.

It is worth noting a few details that are forthright in the particular case of the H defect system modelled in this work. The formation energy of the defects can be compared with another DFT study on H and intrinsic defects in SnO$_2$, by Singh et al [178]. In this case, the energy barrier to migration, taken as the binding energy of charged hydrogen atom with the O vacancy, is 1.67 eV - which corresponds to the energy barrier against migration between the interstitial and substitutional site. Acknowledging that these quantities may vary with the charge of the defects, this value is less than the difference in formation energies between substitutional and interstitial defect calculated in this work by several eV implying that a substitutional defect is stable against the migration of hydrogen. Moreover, the convex hull plotted in figure (6.20) shows that, if the oxygen chemical potential is only the order of a few eV above the elemental reference value, then there is a range of hydrogen chemical potential relative to the reference value of 18.9 eV over which the interstitial defect will be stable. It can therefore be inferreded that the interstitial defect could appear in higher concentrations in a sample if $\mu_H$ was controlled to be high under relatively oxidising conditions. The significance is there would appear to be some scope experimentally to favour the formation of the interstitial and compensate for its migration at elevated temperatures.

The relative stability of the defect types may also depend on the stability and concentration of oxygen vacancies that enable to some extent the formation of the substitutional type. This can
be analysed in terms of donor levels brought about by Vac(O). The work cited above would likely not be suitable for this task since the GGA+U functional was used, resulting in a band gap of only 1.65 eV, a severe underestimate which means that the position of any energy bands that the formation of the defects entrains and accordingly their energetics discussed by the authors may well be incorrect because levels that should be in the band gap are modelled as inside the conduction band. With the use of the B3LYP functional in this work, the band gap for undoped SnO\textsubscript{2} is large enough that the donor level for the oxygen vacancy defect is deep in the gap without H. Yet the interstitial defect causes the CBM to descend deeper in the band gap, which brings about the possibility of enabling this deep level to donate. This could have the consequence of making the formation of substitutional defects less preferable as the extra electrons from the original vacancy defect would become delocalised instead of staying at the O sites to form bonds with H, depending on the position of \(E_F\). Alternatively, this could also compensate the vacancy leading to the disappearance of the original donor level. This is the case for the substitutional defect, so that the formation of this defect is not expected to activate oxygen vacancies. Therefore, under conditions of greater electron transport, it is not necessarily the case that conductivity from expected defects in SnO\textsubscript{2} would be expected to increase or decrease linearly; in fact, the outcome is ambiguous. In conclusion, the promise of the interstitial H is probably greater than that recorded by previous DFT work due to a favourable energetic resilience to migration of the defect.

### 6.3.7 Electronic structure and conductivity

Looking at electronic structure, the interstitial bandstructures resemble more closely that of undoped SnO\textsubscript{2}, with a parabolic CBM. The substitutional bandstructure on the other hand may be derived from the oxygen vacancy case, which in itself is reasonable since introducing a substitutional H first requires an O vacancy to be created. As discussed above in section (6.3.3), the H substituted at an O vacancy site removes the deep donor level from the vacancy that is seen in figure (6.7), instead giving a partially occupied state at the CBM. Formally, this phenomenon is an example of a shallow donor state, where the host state at the edge of the CBM moves downwards into the band gap to form a donor level near the top of band gap that in this case is partially occupied at \(T = 0K\). Because this state is (high) in the band gap, it will have a mixed VB/CB character. In this situation H can behave amphoterically. While the bonding of the H to Sn evident from the charge difference map figure (6.26) would imply the state acting as an acceptor (H\textsuperscript{+}) if it were fully occupied \([203], [204]\), it is in fact partially occupied at zero temperature which means that the formation of both H\textsuperscript{+} and H\textsuperscript{−} would need to be considered as a function of \(E_F\) and hence the defect contributing to both \(n\)-type and \(p\)-type conductivity.

In the case of interstitial H\textsuperscript{0}, on the other hand, there is no state in the band gap so the position of \(E_F\) can only give zero or partial occupation of the conduction band - and indeed, it is partially filled. Hence, interstitial H\textsuperscript{0} acts exclusively as a donor. The relative stability of the defect in
the \(\text{H}^+\) and \(\text{H}^-\) charge states is not been calculated here since the potential energy reference for the electrons is not well defined, but with \(E_F\) well inside the band gap at \(T = 0\text{K}\), this should not matter.

Theoretical studies have, to the best knowledge, all proposed that any conductivity arises due to an H defect forming a shallow donor state. However, generally these studies have used DFT functionals such as LDA and GGA which strongly underestimate the band gap, so that the donor state cannot be distinguished from the states in the CB. An experimental study by King et al [202] sought to observe the donor level directly, described in section (4.5.5). A (very) shallow donor state was observed, yet this is not seen in the bandstructure for the interstitial defect figure (6.23).

However, King et al justify the shallow donor argument another way, centred around the notion of the CNL discussed in section (4.6.1). Since H is very localised, its AO interacts more or less evenly with AO across the BZ, and hence with both the cation and anion. Consequently, there can be a local charge excess or deficit as a function of small distances away from the H site, yet when averaged across the BZ the approximate CNL lies close to the centre of the band gap. Moreover, in SnO\(_2\) (and other \(n\)-type semiconductors such as ZnO and In\(_2\)O\(_3\)) the CBM is parabolic and extending well into the band gap about the \(\Gamma\) point. For such materials, and indeed in the calculations presented here for SnO\(_2\), the depth can be approximately midway down the band gap. This places the CNL just above the low lying CBM about the \(\Gamma\) point. At, or close to \(T = 0\text{K}\), one would therefore expect the CNL and hence the conduction band to be partially occupied. Hence King et al propose a novel explanation as to why H is an \(n\)-type donor when it normally acts as an acceptor in an \(n\)-type material. This perspective can explain some of the discrepancies between differing explanations for the behaviour of H in semiconductors. In the picture of King et al, the donor level is a deep donor rather than what is usually referred to by the literature on ab initio defect calculations as a 'shallow donor', in the sense of being well inside the band gap of the material (and hence separated by a relatively large energy difference from the conduction band, except the CBM near the \(\Gamma\) point). This deep donor level has also been proposed to originate from the dangling bond state on the cation due to the bond formed between H and the anion, by van de Walle and Janotti [240], [236] (for the general case of a compound semiconductor). The donor level itself should be distinct, it is argued, and derived from a composition of host states across the BZ.

The electronic structure presented here for SnO\(_2\) provides some supporting evidence in favour of the 'shallow donor' detected experimentally by King et al, at the same time partially vindicating the hypothesis of van de Walle and Janotti (and previous authors), that favours a deep donor with an alternative physical origin. The first feature is the low-lying CBM that is a feature of both explanations. Neither author specifically predicts the reduction of the band gap due to the increased depth of the CBM, found here give a reduction by 1.22 eV (and a smaller reduction for the substitutional defect). This is despite the fact that a CBM lying around midway between the rest of the CB and the VB is seen as a pre-requisite to form a conductor for King et al.

Janotti and van de Walle, however, view a dangling bond state below the CB as a consequence of the bond formed between H and anion, and which forms the deep donor level they hypothesise. In this work, however, the effect of H (interstitial) is to lower the CBM rather than to create a
distinct electronic state.

The other important detail that distinguishes the cases is the band filling. In the bandstructure here, \(E_F\) lies around 2.0 eV inside the conduction band. This degree of occupation of the CB is much higher than reported in any known study so far, and therefore warrants comment. In the case of King et al, the position of the CNL (for which one would expect equal concentrations of \(\text{H}^+\) and \(\text{H}^-\) to form) depends on \(E_F\). In an investigation performed on CdO, that they cite in support of their reasoning for \(\text{SnO}_2\) on account of it also having a deep parabolic CBM, the position of \(E_F\) was found to relocate to the same energy as CNL after sufficient diffusion of H. Since CdO is a similar \(n\)-type semiconductor to \(\text{SnO}_2\) in the sense of having the low-lying parabolic CBM, it should be a fair assumption that \(E_F\) is generally found around the edge of CBM in these materials, \(i.e.,\) if not inside it then very close. Van de Walle and Janotti predict that \(E_F\) should be located midway between their dangling bond donor and acceptor levels (which are close to the CBM and VBM, respectively), where it takes positions close to the CNL. The position they predict is hence approximately at the band gap mid point. This represents a notable difference from King et al who propose that the CNL is to be found inside the band gap, which is also the position of \(E_F\) following sufficient diffusion of H within the semiconductor. This difference can be explained by the fact that the CBM extends to near the centre of the band gap, as per the prediction of King et al and vindicated by the result presented here. As a result, the CNL in their picture is influenced proportionally by both the CBM position and the rest of the CB, placing it slightly higher than the band gap centre on average and therefore above the CBM in their model. The bowed feature of the CBM is absent for Janotti and van de Walle, and it could go some way to explaining why other \(n\)-type materials like ZnO are doped \(n\)-type by hydrogen, as the CNL would rise further inside the band gap and possibly inside the CB as this work, and King et al, have found. However, that phenomenon has principally been viewed to date in terms of the CNL position as determined by the point of equality of the \(\text{H}^+\) and \(\text{H}^-\) formation energies in a theory where those quantities are a linear function of \(E_F\), as discussed in section (4.6.1). Here, however, the position of \(E_F\) is not pinned in the band gap, and the formation energies of all the defects will be sensitive to thermodynamic conditions.

That the result here finds a conductor with \(E_F\) in the conduction band is a sign that the potential of H to increase conductivity when forming interstitial defects is significantly greater than previously recorded, with the implication that the theory of H for \(n\)-type semiconductors may need to be revised. With a partially occupied CBM, the formation energies of all defects in the material can change, and \(E_F\) varies as these defects become electrically active in the material. Hence there is actually a continuum of formation energies of \(\text{H}^0\) and other defects as the position changes inside the conduction band, as opposed to a single set of defect formation energies for a particular value of \(E_F\), which studies to date have argued.

This important feature of the interstitial H result can also be seen in terms of the distinction between donor levels and a semiconductor that is simply turned metallic by doping. To act as a donor in the conventional sense, a concentration of \(\text{H}^+\) or \(\text{H}^-\) would be required to form, made electrically active by the transfer of electrons to and from the valence and conduction bands. Yet the position of \(E_F\) inside the conduction band produces a metal both due to the electron
6.3 Hydrogen defects in bulk rutile SnO$_2$

concentration, and to the overlap of the first two bands in the CB. Because the experimental investigation of King et. al finds a (very) shallow donor state, however, the next section looks for evidence for the donor state through the charge difference maps. The results so far indicate that the formation of donor states may not be necessary to increase the concentration of $n$-type carriers, yet the possibility is not excluded as the next section discusses.

6.3.8 Evidence for donor state - Electronic structure and charge density

Zunger and Kiliç [238], and Xiong et al, [239] also have reported H donor levels in SnO$_2$, although the former speak of ‘delocalised’ H states forming in the low conduction band. Only DOSS is published to this effect. Van de Walle and Janotti favoured localised, deep donor levels formed due to the H bonding to anion. Careful scrutiny of these papers taken together emphasises the conceptual debate around the nature of H bonding in semiconductors and defect states generally.

An explanation in terms of delocalised levels is more reminiscent of perturbed host states (PHS), where an electron from H relaxes, creating a series of resonances with dispersion in the region of the change in electron concentration, that is delocalised in this region as a result (corresponding to the $\Gamma$ point in the calculation here). Indeed, the earliest studies of hydrogen in semiconductors implied that it forms linear combinations of the host states (albeit in a single particle Hamiltonian, as detailed in section (4.5.5)). The result for the substitutional defect in this calculation, discussed in the previous section, is likely an example of this type of defect state. Experimental work (King et al.) suggests that a hydrogen defect forms a shallow donor just a few eV below the CBM at the $\Gamma$ point. The investigation does not permit to distinguish if the hydrogen detected is an interstitial or substitutional defect, which is an important caveat. However given that both types are found here to be majority $n$-type conducting at zero temperature (the conclusion for the substitutional could change at higher temperatures), the results presented in this work can be a point of comparison between delocalised and localised hydrogen induced states.

At first glance the charge difference maps presented here suggest that there cannot be a shallow donor state. The electron density maps for H$_{tel}^+$ and H$^0$ are almost identical, with no difference in electron density on H between the two charge states, so that the charge migration to H that is seen in figures (6.28) and (6.27) involves (only) the bonding O. This, however, is consistent with the bandstructures for the two cases, figures (6.23) and (6.24). In both charge states, the H atom gains charge of an $sp$-character amounting to a single electron donated from O. The ionic character of the bond comes about by O$^{2-}$ donating one of its electrons to the bare proton, which can be seen by the fact that the molecular charge difference map for O$^{2-}$ and OH$^-$ (figure (6.29)) is more or less identical to that for interstitial H in the semiconductor. The electrostatic attraction allows the H to move closer to O to the point that its 1s electron can behave as a free electron and hence partially occupy the electronic state at the CBM. The strong effect on electron concentration can be seen as the result of the reduced electrostatic repulsion at the Sn site, allowing an energetically favourable distribution of NFE conduction band
electrons between the first two states of the CB, rather than leaving a singly occupied donor electronic state associated with $\text{H}$, as earlier authors argued. The charge density movements resulting from the delocalised 1s electron are not visible in the charge difference maps because the resulting electronic state is so delocalised. The electron will leave $\text{H}$ at an ill defined point in the process of its incorporation into the lattice.

The finding of metallic H-doped SnO$_2$ can therefore be explained without the need for a donor state. That one is observed under the methodology of King et al. requires comment. It is important to state that donor levels are not incompatible with the above reasoning. The 1s AO of $\text{H}$ may hybridise additionally with the Sn AO, even though there does not appear to be any electron sharing between them of ionic or covalent character. This could produce an electronic state that is orthogonal to the host orbitals. Modelling the absorption of molecular $\text{H}$ into the bulk crystal may produce some degree of Basis Set Superposition Error (BSSE), caused by the overlap of the basis functions of $\text{H}$ with those of the host atoms as H approaches its absorption location. The wavefunction of $\text{H}$ will have a basis set encoding a varying description of the host electronic structure which could affect energy levels including that of any donor state. To assess the BSSE, the calculations for the defects presented here would be repeated with $\text{H}$ fixed at a set of positions between its initialisation and final position bonded to $\text{O}$.

Another important observation from the experiment is the paramagnetic signal detected (and assigned to) the shallow donor state. The donor state is therefore singly occupied, at least at the lowest temperatures. This could imply that some degree of spin polarization of the system, or spin ordering on the sites where $\text{Mu}^0$ forms, is responsible, this effect then diminishing for higher temperatures $T \geq 30$K. Alternatively, increasing temperature could lead to a change from single to double occupation of the donor level and hence the disappearance of the paramagnetic signal.

The lack of evidence of spin-polarization in the result here simply reflects the delocalised state that cannot adopt a spin-up or spin-down state on average. However, the concentrations of $\text{H}$ used in the DFT calculation (one $\text{H}$ atom added to a 48-atom supercell) are likely not quite representing the bulk limit. In this limit, spin polarization effects are expected to be more significant, making a state occupied by only $\alpha$ or $\beta$ electrons appear in the bandstructure at a sufficiently low concentration. Even not considering spin, another possibility depending on concentration is that the electron liberated in the transition from $\text{H}^0$ to $\text{H}^+$ is delocalised in the conduction band, with the transition probability of the electron between any two Sn approximately equal. The resulting state, the parabolic CBM, is well represented by the quadratic dispersion as a 'nearly free electron' in the reciprocal space. As the size of the system increases, the probability of any given transition decreases and the effective period of the 'wave' decreases as it becomes less dispersed. The state resembling a free electron would be replaced by the 'shallow donor' state suggested by King et al. that is more likely to be a deep donor that touches the CBM. To test this hypothesis, it would be necessary to perform calculations on the defect in a larger supercell than has been used to obtain these results.

The emergence of metallic SnO$_2$ upon H-doping due to the interstitial defect can be expected to occur readily according to our result. The efficacy of the dopant could be limited by the thermodynamic stability of the interstitial defect and depend also on other defect populations.
in the material. Indeed, it is actually the substitutional defect that is predicted to be more stable where it does form, governed by the oxygen chemical potential $\mu_0$ (figure (6.20). The interstitial hydrogen substantially distorts the Sn-O bonds as shown in figure (6.22 b), due to ionicity of the OH bond. As the loading of interstitial H is increased, this arrangement could become increasingly disfavoured energetically, accommodating fewer defects. The substitutional defect, despite requiring the removal of an O from the lattice, simply changes bond lengths with little localised distortion, and indeed oxygen vacancies are found to be present in modest concentrations in samples of SnO$_2$.

The experimental doping of SnO$_2$ might therefore face the difficulty of establishing a substantial population of interstitial defects. This is difficult to quantify without exploring temperatures closer to the target application working temperature, which affects both the relative stability of interstitial and substitutional defect and also the stability of the interstitial defect once formed, being highly mobile and sensitive to conditions during cooldown [233], [241], [242].

### 6.3.9 Conclusion

Hydrogen doping appears to make SnO$_2$ conductive without creating the ‘shallow donor’ state that the most up to date experimental evidence suggests [202]. However, it is more likely that the donor state not seen by the analysis here is actually the deep donor suggested by King et al. to theoretically explain their experimental result. The OH$^-$ group appears regardless of whether the H defect is modelled as formally charged, or neutral (the shell initialisation of H in the CRYSTAL program does not matter). Only in the neutral case does the system become conductive. This shows that the conductivity does indeed come from the extra e$^-$ of hydrogen, but through producing a state that is so delocalised that the electron does not show up at a well defined point on the lattice. In terms of the charge difference map that attempts to show it, the contribution to any isosurface cannot be distinguished from that of the rest of the system. The formation of the OH$^-$ group is found in agreement with earlier computational studies for interstitial H, giving an explanation additionally for the lowering of the CBM into the band gap, and the filling of this band.

The charge carrier statistics associated with the interstitial H will depend on both the position of $E_F$ and the position of the conduction band minimum at the $\Gamma$ point activating deep donor states produced from Ta and oxygen vacancies. Hence there are two potential sources of $n$-type conductivity due to interstitial H, while the result for substitutional H suggests that this will behave as an amphoteric donor, capable of contributing towards $n$-type or $p$-type conductivity. Previous work on H defects in SnO$_2$ proposing shallow, or deep, donors generally used LDA and GGA functionals. The band gap is usually severely underestimated with these functionals, making it very difficult to distinguish donor electronic states from other conduction band states. This highlights the importance of using a hybrid functional such as B3LYP for analysing the defect properties of wide band gap oxides.
7 OUTLOOK

7.1 Rigorous treatment of Jahn-Teller effect

The mechanism responsible for lowering the donor bands within the band gap is not fully elucidated. To complete the analysis, it needs to be proven that a JT effect occurs affecting specific $d$ orbitals. For this, one would need to calculate the projection of the assumed donor states onto the $d_{x^2-y^2}, d_{z^2}$ states. There are a number of possible routes to formulate JT distortions in terms of lattice properties. One of the chief ambiguities left unresolved in studying the Jahn-Teller induced distortions of the Ta–Sn rutile alloys is the close integration of electronic and structural degrees of freedom in the symmetry breaking of the coordinated Ta-O ligands, which are difficult to isolate partly because they are not encoded into the DFT (Kohn-Sham) Hamiltonian on a term by term basis. However it is possible to study the effects as coupled modes in terms of the field of the electronic state interacting directly with the lattice vibrational modes (that are not derived from true phonons but based on a semi-classical oscillator) as described by O’Brien and Chancey [176]. The problem then becomes a problem of energy minimisation in the space of coordinates that can be chosen conveniently to study a given space lattice and distortions (spatial transformations) within it. The framework is best built up by first constructing the Hamiltonian for the given crystal system and calculating its normal modes, and then choosing the couplings to different terms in the expression to reflect numerous electronic details such as the variation in the wavefunction (between ion sites, in the case at hand) quantum tunnelling between sites, and excited states.

Such an approach, which rests upon the analytical system of the harmonic oscillator, is a powerful way to build up increasingly complex distortions of the doped crystal and this gives it the potential to quantify the collective distortions which are driven by a combination of coulomb, orbital and defect-defect interactions over varying length scales, even though the nature of the true, fully interacting many electron system remains unknown to an arbitrary degree of complexity. As long as a scaled JT effect is the dominant effect in the system, this approach would have much utility for developing the method documented in this thesis in order to determine the particular crystal distortions, not forgetting that solving a system of oscillators would quickly become computationally challenging and without considering the compatibility with the mean-field picture assumed in DFT. However, the outcome of such a complete study could be very useful as knowledge of the distortions could enable one to predict sites where selective doping could limit the collective distortion of the crystal (described below).

7.1.1 Engineering of symmetry and electronic properties of structures in cluster expansion (CE)

By adding intrinsic defects to the CE structures it might be possible to influence separately the relaxations governed by the symmetry of the structure and the electronic properties to overcome
the problem of 'freezing out' of donor levels introduced by Ta. A possibility is the inclusion of neutral defects that impede or place constraints on the relaxations driving lowering of the d-states by JT. The inclusion of such defects would need to be integrated into the methodology of the cluster expansion, since to have the intended effect they would need to be distributed at the correct crystal sites, and as for the substitutional doping of Sn and Ta, the combinatorial possibilities of distributing the dopant Ta with other defects would have to be taken into account, since the collaborative JT effect is a function of the multiplicity of configurations. One approach may be to widen the definition of the sites where substitutional doping occurs, to include the sites of the other compensating defects, and then use a CE in which the occupation of each site can have three different values- Ta, Sn or compensating defect. This ternary parameterisation of the CE has been previously approached [225].

If the sought properties can be obtained from the CE, doping limits need to be determined from the intersection of the minimum energy surfaces of the disordered and ordered compounds, which are respectively the Gibbs free energy curves and the convex hull surface with the ternary compound SnTa₂O₆. Monte-Carlo thermal averaging over the CE states is needed to formulate one or more Gibbs Free Energy curves across the composition space (with the minima being the CE groundstates).

### Alternative extrinsic dopants

The focus has been on Ta due to the interest in the addition of donor d electrons to a system with SnO₂ as the host. Other dopant metals M, that form stable oxides at elevated temperatures in an acidic environment, can be explored. The valence implied by the relevant oxide can be taken as the basis for the assumed shell structure of the element, and whether d electrons can be expected to play a role. It would be expected that the highest occupied bands would coexist over the same energetic range as the oxygen valence bands. However, in a mixed oxide with another metal, the environment a given M atom experiences may, through effects such as hybridisation and the Jahn-Teller distortion described above, result in certain bands becoming higher or lower in energy than in pure oxide. Some of these effects are a result of disorder in the compound, which can only be assessed by performing the CE. However, calculating the bandstructures of oxides of the M' is a productive first step in order to know the band gaps and an indication of any likely donor states. Table (8) shows possible doping elements. By
calculating the bandstructures of the oxides in which they occur, it will be possible to deduce whether they could act as a host oxide or as a dopant, for \( p \)-type and \( n \)-type doping. If \( \text{SnO}_2 \) is the host, then the required valence states \( V \) for \( n \)-type doping must be \( V \geq 4 \) in order to have electrons donated to the conduction band. The preference towards the sharing of electrons by particular orbital types in a bond due to factors such as the orbital symmetry, hybridisation and diffusivity, may dictate the orbitals occupied and in which atomic species (\( \text{Sn M}' \), or \( \text{O} \)) affecting the energy of the defect state, its dispersion, and appearance in a particular segment of the BZ. This could influence the type of electrical conductivity. In the Ta-Sn-O system modelled in this work, the conclusion is that electrons are donated to the system by Ta due to the crystal field splitting. Yet the useful effect of states high in the band gap is undermined by collaborative Jahn-Teller effects.

On the other hand, another possibility in the consideration of alternative dopants are two elements that do not share the same crystal structure. This would be possible if, as in the case of \( \text{Ta} - \text{Sn} - \text{O} \), the phase diagram revealed a subspace bounded by a different stochiometric for each of the two metals. This could be treated in one of two ways: either by modelling one dopant as an extrinsic defect, or continuing with a cluster expansion based on two different parent lattices. This would introduce more complexity into analysis of the system behaviour, but the option to do it is provided by the MAPS code.

In summary, a short-term line of investigation for the future work is DFT calculations on all potential host and dopant metals, which will enable rough determination of the order stability of the metals to oxidation as defined by equations (102) and (103) in the next section.

### 7.2 Stability as the oxidation voltage at doping limit

The preceding analysis of the CE (symmetry unconstrained) assumed that a principal degradation mechanism of disordered Ta-Sn mixed oxides is oxidation. A definition of the oxidation potential can be derived by consideration of the doping limit for the creation of a disordered compound \( \text{Ta}_x \text{Sn}_{1-x} \text{O}_2 \) through the introduction of either Ta or Sn into either of the pure oxides \( \text{TaO}_2 \) or \( \text{SnO}_2 \), since it is expected that the final compound \( \text{Ta}_x \text{Sn}_{1-x} \text{O}_2 \) will be produced with pure oxides as hosts. Corresponding to the chemical potential \( \mu_M \) at which the doping concentration of Ta (Sn) is added to \( \text{SnO}_2 \) (\( \text{TaO}_2 \)) is an oxidation potential \( \phi(\delta) \) which is of course a function of the doping limit \( \delta \). Where the formation energy of the (disordered) compound at \( T > 0 \text{K} \) is equal to that of the \( T = 0 \text{K} \) compound, we have the condition

\[
\mu_M^{T=0 \text{K}} = \frac{\partial E_f^{T=0 \text{K}}}{\partial n_M}
\]

and we are looking for the associated value of \( n_M \), related to \( \delta \). A change in the external chemical potential will change the composition of the minimum energy compound, according to the Gibbs Free Energy curve, as a linear change in the composition corresponds to a non-linear increase in formation energy. However, in the limit of small concentration changes, \( \frac{\partial E_f^{T=0 \text{K}}}{\partial n_M} = \frac{\Delta E}{\Delta n_M} \) and
the energy corresponds to the dissociation of Sn$_{1-\delta}$M$_{\delta}$O$_2$,

$$\text{Sn}_{1-\delta}\text{M}_\delta\text{O}_2 \rightarrow (1-\delta)\text{SnO}_2 + \mu_{\text{O}_2} + \delta M^{+\gamma} + ye^- \quad (101)$$

for the corresponding valence $y$ in the limit of the minimal electron transfer processes. We are interested in this small energetic change where (100) holds, within the limits of validity of the model on which the cluster expansion is based. Applying the condition (100) in terms of formation energies gives

$$\frac{\Delta G_f(\text{Sn}_{1-\delta}\text{M}_\delta\text{O}_2)}{\Delta n_M} = \frac{(1-\delta)\Delta G_f(\text{SnO}_2) + \delta(\Delta G_f(M^{+\gamma}) + \phi y) + \mu_{\text{O}_2}}{\delta} \quad (102)$$

where $\phi$ is the chemical potential of electrons transferred in the process. The formation energy can also be written in terms of the chemical potential of the dopant metal at concentration $\delta$, the potential needed to form a stochiometric compound with the pure oxide SnO$_2$,

$$\Delta G_f(\text{Sn}_{1-\delta}\text{M}_\delta\text{O}_2) = \Delta G_f(\text{SnO}_2) (1-\delta) + \delta \mu(\text{M}_\delta) + \delta \mu_{\text{O}_2} \quad (103)$$

where the notation in the chemical potential $\mu(M_\delta)$ denotes that the concentration is held fixed and only the chemical potential of the metal species is varying. Equating (102) and (103) gives the electron chemical potential as a function of the concentration $\phi(\delta)$,

$$\phi(\delta) = \frac{1}{y} \left( \frac{(1-\delta)^2 \Delta G_f(\text{SnO}_2) + (\delta - 1)\mu_{\text{O}_2}}{\delta} + \delta \mu(\text{M}_\delta) - \Delta G_f(M^{+\gamma}) \right) \quad (104)$$

In the above, the free energy of formation of the metal ion of species M, $\Delta G_f(M^{+\gamma})$ is to be taken from the data of known M, taking care that the definition of the electron chemical potential, or Fermi level, $\mu_e$ is consistent with that used in the specification of the reference values.

With a result for the doping limit oxidation potential, the quantity can be estimated from the calculations on compounds in the Ta–Sn–O space. The formation energy $\Delta G_f(\text{Sn}_{1-\delta}\text{M}_\delta\text{O}_2)$ can be calculated in terms of the oxides of M and SnO$_2$, for example if we are interested in the compound that is $1/8\text{M}$,

$$\Delta G_f(\text{Sn}_7\text{M}_1\text{O}_2) = \Delta E_8^1 = \frac{1}{20} E_T(\text{Sn}_7\text{M}_1\text{O}_{16}) - \frac{1}{160} (7\Delta G_f(\text{SnO}_2) + \Delta G_f(\text{MO}_2)) \quad (105)$$

For other concentrations where $\delta << 1$, the same applies if a supercell formed of the Sn$_{1-\delta}$M$_\delta$O$_2$ compound is considered. This approach allows us to get the formation energy of many different compounds from the same constituent oxides. If we calculate the energies of two compounds having the same $\delta$ in this way, we obtain a relationship that gives an estimate of $\mu(M)$ at fixed $\delta$, and we can then calculate the corresponding electron chemical potential $\phi(\delta)$ based on equation (104). Repeating the procedure for many different compounds of the form Sn$_{1-\delta}$M$_\delta$O$_2$ gives $\mu(M_\delta)$ as a function of $\delta$ and by repeating this procedure across a number of compounds with different M we can get a basic picture of the relative stability of systems with SnO$_2$ as host and M as dopant metal.
7.3 Final Concluding Remarks

This work has investigated a number of promising potential sources of conductivity of rutile SnO$_2$ using a suitable DFT parameterisation. Since SnO$_2$ is stable in the intended application and can be doped with a descriptive simulation by an ab-inito approach, the focus on it is justified as a good use of computational resources. The principal findings addressing the original research statement are,

- identification of H-doped SnO$_2$ as metallic, subject to sufficient concentration and diffusion of the H interstitial defect,
- stability of interstitial H depends on both oxygen and hydrogen chemical potential,
- both $n$ and $p$ type conductivity are possible for substitutional H defect,
- extrinsic doping (particularly in alloying concentrations) of SnO$_2$ by Ta and other group V elements to increase conductivity is not feasible, placed on a firm theoretical basis by identification of collaborative Jahn-Teller (JT) effects

The results propose a shift away from extrinsic dopants as primary candidates for SnO$_2$ based catalyst support materials, which can hopefully help to prevent future fruitless attempts to activate wide band gap semiconductors and save earth resources.

The most important extension to this work would be to determine the conditions for formation and stability of the H defects, with particular interest in the interstitial, under more realistic conditions of higher temperature. Quantifying this is essential in order to determine the true potential of conductivity due to H incorporation. The work on Ta-doped SnO$_2$ proposes an extension to explain the operation of JT in crystals in terms of the normal modes of the bulk structure, which would enable one to identify the distortions of a given crystal structure explicitly by first principles methods, potentially opening up the possibilities for optimising coordination of composition and crystal structure to control conductivity and moving potentially towards ab-initio materials design for feasible applications.

Calculations in this work could be amended by moving to a higher level of DFT theory, to facilitate consideration of excited states and varying thermodynamic conditions. While the level of description facilitated by the DFT approach used here is sufficient to explain conductivity mechanisms, there is less of an explanation of the mechanisms behind stability of an arbitrary doped material, even though the ATAT code does permit an assessment of phase stability, which has not been attempted. The framework developed here would allow fairly rapid development of a method to treat perturbations to the $T = 0$K approximation of simple DFT, such as the nudged elastic band method for computing metastable configurations of H, which are important for computing the structural changes with a heavy loading of the impurity that might be required to determine the feasibility of interstitial H as discussed in section (6.3.8). This could also be coupled with a finite temperature approximations such as Monte-Carlo techniques that are described in section (5.6.1).
Together, the work on Ta-doping and H defects demonstrates the importance of the functional used in DFT. The B3LYP functional properly treats interactions between involving the $d$-electrons, to the extent that the localisation of the electron of the interstitial $H^0$ defect was well modelled. This shows that the electronic structure can be adequately described in terms of defect-defect, orbital and electrostatic interactions which are those important at the level of DFT used here. B3LYP also comfortably estimates the wide band gap of $\text{SnO}_2$ which means that it has not been necessary to apply band gap corrections as the donor states and others near the top of the band gap do not move into the CBM due to the defects. Hence, the defect properties can unambiguously be shown through the effects on the bulk crystal and electronic structure. It is worth emphasising therefore that these results indicate DFT parameterised by hybrid functionals provides a more accurate description of $\text{SnO}_2$ and its defect chemistry than the local and semi-local functionals often employed to study the material.
8 APPENDIX - CRYSTAL09 Basis Sets

8.1 Hydrogen

For Hydrogen a three shell basis set [229] was employed, two s type shells and one p type shell, with one of the s– shells constructed using 3 primitive GTF and the other with a single GTF and equally for the p type shell.

<table>
<thead>
<tr>
<th>Shell-type</th>
<th>$\alpha_j$</th>
<th>$c_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1s</td>
<td>$s$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18.73113696</td>
<td>0.033349460434</td>
</tr>
<tr>
<td></td>
<td>2.825394365</td>
<td>0.2347269535</td>
</tr>
<tr>
<td></td>
<td>0.640121692</td>
<td>0.8137573262</td>
</tr>
<tr>
<td>2s</td>
<td>$s$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1612777588</td>
<td>1.00000000</td>
</tr>
<tr>
<td>3p</td>
<td>$p$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.100000000</td>
<td>1.000000000</td>
</tr>
</tbody>
</table>

8.2 Oxygen - $\text{Ta}_x\text{Sn}_{1-x}\text{O}_2$

The basis set for O employs a 8-411d(1) contraction, using one s, three sp and one d shell used to control polarization of the 3sp shell.

<table>
<thead>
<tr>
<th>Shell-type</th>
<th>$\alpha_j$</th>
<th>$c_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1s</td>
<td>$s$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5484.671660</td>
<td>0.001831074430</td>
</tr>
<tr>
<td></td>
<td>825.2349460</td>
<td>0.0139501722</td>
</tr>
<tr>
<td></td>
<td>188.0469580</td>
<td>0.06844507810</td>
</tr>
<tr>
<td></td>
<td>52.99645000</td>
<td>0.2327143360</td>
</tr>
<tr>
<td></td>
<td>16.89757040</td>
<td>0.4701928980</td>
</tr>
<tr>
<td></td>
<td>5.799635340</td>
<td>0.3585208530</td>
</tr>
<tr>
<td>2sp</td>
<td>$s$</td>
<td>$p$</td>
</tr>
<tr>
<td></td>
<td>15.53961625</td>
<td>-0.110777549</td>
</tr>
<tr>
<td></td>
<td>3.599933586</td>
<td>-0.1480262620</td>
</tr>
<tr>
<td></td>
<td>1.013761750</td>
<td>1.130767010</td>
</tr>
<tr>
<td>3sp</td>
<td>$s$</td>
<td>$p$</td>
</tr>
<tr>
<td></td>
<td>0.2700058226</td>
<td>1.00000000</td>
</tr>
<tr>
<td>3d</td>
<td>$d$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8000000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>
This basis set is an example of a 6-31d1 contraction, using six GTF for the 1s shell and two sp shells using one and three GTF, and one d shell for polarization.

<table>
<thead>
<tr>
<th>Shell-type</th>
<th>$\alpha_j$</th>
<th>$c_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1s</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8020.0</td>
<td>0.00108</td>
</tr>
<tr>
<td></td>
<td>1338.0</td>
<td>0.00804</td>
</tr>
<tr>
<td></td>
<td>255.4</td>
<td>0.05324</td>
</tr>
<tr>
<td></td>
<td>69.22</td>
<td>0.1681</td>
</tr>
<tr>
<td></td>
<td>23.90</td>
<td>0.3581</td>
</tr>
<tr>
<td></td>
<td>9.264</td>
<td>0.3855</td>
</tr>
<tr>
<td></td>
<td>3.851</td>
<td>0.1468</td>
</tr>
<tr>
<td></td>
<td>1.212</td>
<td>0.0728</td>
</tr>
<tr>
<td>2sp</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49.43</td>
<td>-0.00883</td>
</tr>
<tr>
<td></td>
<td>10.47</td>
<td>-0.0915</td>
</tr>
<tr>
<td></td>
<td>3.235</td>
<td>-0.0402</td>
</tr>
<tr>
<td></td>
<td>1.217</td>
<td>0.379</td>
</tr>
<tr>
<td>3sp</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4520495</td>
<td>1.0</td>
</tr>
<tr>
<td>4sp</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1678880</td>
<td>1.0</td>
</tr>
<tr>
<td>3d</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4509895</td>
<td>1.0</td>
</tr>
</tbody>
</table>
8.4 Tin - Oxygen Vacancy and Hydrogen Defect calculations

The basis set was of the DURAND-21G* variety. A Durand and Barthelat ECP was used for core electrons. Valence electrons were represented using two GTF for the first \( sp \) shell, one GTF for the second, and one \( d \)-type GTF for polarization.

<table>
<thead>
<tr>
<th>Shell-type</th>
<th>( \alpha_j )</th>
<th>( c_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4( sp )</td>
<td>s</td>
<td>p</td>
</tr>
<tr>
<td></td>
<td>0.664045</td>
<td>-0.531536 -0.094875</td>
</tr>
<tr>
<td></td>
<td>0.236154</td>
<td>0.858798 0.474750</td>
</tr>
<tr>
<td>5( sp )</td>
<td>s</td>
<td>p</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>1.0</td>
</tr>
<tr>
<td>5( d )</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.10</td>
</tr>
</tbody>
</table>

8.5 Tin - \( \text{Ta}_x \text{Sn}_{1-x} \text{O}_2 \)

The basis set was of the DURAND-21G* variety. A Durand and Barthelat ECP [243]- [246] was used for core electrons. Valence electrons were represented using two GTF for the first \( sp \) shell, one GTF for the second \( sp \) shell.

<table>
<thead>
<tr>
<th>Shell-type</th>
<th>( \alpha_j )</th>
<th>( c_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4( sp )</td>
<td>s</td>
<td>p</td>
</tr>
<tr>
<td></td>
<td>0.67428</td>
<td>-0.51718 -0.090973</td>
</tr>
<tr>
<td></td>
<td>0.23445</td>
<td>0.83855 0.46425</td>
</tr>
<tr>
<td>5( sp )</td>
<td>s</td>
<td>p</td>
</tr>
<tr>
<td></td>
<td>0.120</td>
<td>1.0</td>
</tr>
</tbody>
</table>

8.6 Tantalum

The basis set for Ta employed a user defined ECP for the core electrons published on the CRYSTAL website, with single GTF to represent the \( sp \) shells, three GTF for the first \( d \) shell and one GTF for the second.
### Core Pseudopotential

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$\alpha_{kl}$</th>
<th>$C_{kl}$</th>
<th>$n_{kl}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>438.9654138</td>
<td>-0.1418323</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>112.8894389</td>
<td>-44.6105104</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>39.0300509</td>
<td>-230.6825998</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10.5663985</td>
<td>-100.8703657</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3.1450008</td>
<td>-24.4967372</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.9962317</td>
<td>-3.9862338</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>29.9605388</td>
<td>63.8012063</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>10.1122075</td>
<td>154.0916202</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2.9034296</td>
<td>42.8001245</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.6314816</td>
<td>18.2668021</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.5171641</td>
<td>-7.6317151</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>82.8488496</td>
<td>1.9022608</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>26.2097320</td>
<td>55.8292082</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>8.2687456</td>
<td>121.9975284</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2.1411005</td>
<td>28.5395818</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.5300236</td>
<td>7.5403200</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>80.2483832</td>
<td>2.9426083</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>28.8958219</td>
<td>52.5314003</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>10.2668935</td>
<td>127.4437340</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2.7589509</td>
<td>39.0662845</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1.6942558</td>
<td>-4.8954047</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.3198755</td>
<td>-0.2876030</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

### Valence Electrons

<table>
<thead>
<tr>
<th>Shell-type</th>
<th>$\alpha_j$</th>
<th>$c_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5$sp$</td>
<td>0.5942</td>
<td>1.00000</td>
</tr>
<tr>
<td>6$sp$</td>
<td>0.200</td>
<td>1.00000</td>
</tr>
<tr>
<td>4$d$</td>
<td>21.3176</td>
<td>-0.0138</td>
</tr>
<tr>
<td></td>
<td>1.2386</td>
<td>2.8724</td>
</tr>
<tr>
<td></td>
<td>0.4222</td>
<td>4.3807</td>
</tr>
<tr>
<td>5$d$</td>
<td>0.2060</td>
<td>1.00000</td>
</tr>
</tbody>
</table>
Glossary of key abbreviated terms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADT</td>
<td>Accelerated Durability Test</td>
</tr>
<tr>
<td>BET</td>
<td>Brunauer-Emmett-Teller electrode surface area</td>
</tr>
<tr>
<td>B3LYP</td>
<td>Becke-Lee-Yang-Parr hybrid functional</td>
</tr>
<tr>
<td>BZ</td>
<td>Brillouin Zone</td>
</tr>
<tr>
<td>CB</td>
<td>Carbon Black</td>
</tr>
<tr>
<td>CBM</td>
<td>Conduction Band Minimum</td>
</tr>
<tr>
<td>CE</td>
<td>Cluster Expansion</td>
</tr>
<tr>
<td>CFS</td>
<td>Crystal Field Splitting</td>
</tr>
<tr>
<td>CNF</td>
<td>Carbon Nanofibre</td>
</tr>
<tr>
<td>CNT</td>
<td>Carbon Nanotube</td>
</tr>
<tr>
<td>CNL</td>
<td>Charge Neutrality Level</td>
</tr>
<tr>
<td>COR</td>
<td>Carbon Oxidation Reaction</td>
</tr>
<tr>
<td>DFT</td>
<td>Density Functional Theory</td>
</tr>
<tr>
<td>ECSA</td>
<td>Electrochemical Surface Area</td>
</tr>
<tr>
<td>EPR</td>
<td>Electro-paramagnetic spin resonance</td>
</tr>
<tr>
<td>FFP</td>
<td>Flow Field Plates</td>
</tr>
<tr>
<td>GCB</td>
<td>Graphitised Carbon Black</td>
</tr>
<tr>
<td>GDL</td>
<td>Gas Diffusion Layer</td>
</tr>
<tr>
<td>GGA</td>
<td>Generalised Gradient Approximation</td>
</tr>
<tr>
<td>GTF</td>
<td>Gaussian Type Function</td>
</tr>
<tr>
<td>HF</td>
<td>Hartree-Fock</td>
</tr>
<tr>
<td>HOR</td>
<td>Hydrogen Oxidation Reaction</td>
</tr>
<tr>
<td>JT</td>
<td>Jahn-Teller effect</td>
</tr>
<tr>
<td>KS</td>
<td>Kohn-Sham equation/energy</td>
</tr>
<tr>
<td>LCAO</td>
<td>Linear combination of Atomic Orbitals</td>
</tr>
<tr>
<td>LTHB</td>
<td>Low temperature hydrogen bubbling</td>
</tr>
<tr>
<td>LDA</td>
<td>Local Density Approximation</td>
</tr>
<tr>
<td>OER</td>
<td>Oxygen Evolution Reaction</td>
</tr>
<tr>
<td>OH</td>
<td>OH group - interstitial hydrogen defect</td>
</tr>
<tr>
<td>ORR</td>
<td>Oxygen Reduction Reaction</td>
</tr>
<tr>
<td>PBC</td>
<td>Periodic Boundary Condition</td>
</tr>
<tr>
<td>PEM</td>
<td>Polymer Exchange Membrane</td>
</tr>
<tr>
<td>PEFC</td>
<td>Polymer Electrolyte Fuel Cell</td>
</tr>
<tr>
<td>PHS</td>
<td>Perturbed Host State</td>
</tr>
<tr>
<td>PTIS</td>
<td>Photothermal Ionization Spectroscopy</td>
</tr>
<tr>
<td>PGO</td>
<td>Point Group Symmetry Operator</td>
</tr>
<tr>
<td>RDS</td>
<td>Rate Determining Step</td>
</tr>
<tr>
<td>SCF</td>
<td>Self Consistent Field</td>
</tr>
</tbody>
</table>

*Continued on next page*
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHE</td>
<td>Standard Hydrogen Electrode</td>
</tr>
<tr>
<td>sub(H)</td>
<td>Substitutional hydrogen defect</td>
</tr>
<tr>
<td>TBA</td>
<td>Tight Binding Approximation</td>
</tr>
<tr>
<td>TCO</td>
<td>Transparent Conducting Oxide</td>
</tr>
<tr>
<td>TEM</td>
<td>Tunnelling Electron Microscopy</td>
</tr>
<tr>
<td>TGA</td>
<td>Thermo-Gravitometric Analysis</td>
</tr>
<tr>
<td>$\mu$SR</td>
<td>Muon spin rotation and relaxation spectroscopy</td>
</tr>
<tr>
<td>Vac(O)</td>
<td>Oxygen vacancy in SnO$_2$</td>
</tr>
<tr>
<td>VBM</td>
<td>Valence Band Maximum</td>
</tr>
</tbody>
</table>
List of key notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>Rutile SnO$_2$ apical bond</td>
</tr>
<tr>
<td>$c$</td>
<td>Rutile SnO$_2$ equatorial bond</td>
</tr>
<tr>
<td>$F$</td>
<td>Faraday constant (96.4853415 Cmol$^{-1}$)</td>
</tr>
<tr>
<td>$z$</td>
<td>Number of electrons transferred in reaction</td>
</tr>
<tr>
<td>$N_A$</td>
<td>Avogadro constant (6.022140857 x 10$^{23}$mol$^{-1}$)</td>
</tr>
<tr>
<td>$R$</td>
<td>Gas constant (8.3144598J mol$^{-1}$K$^{-1}$)</td>
</tr>
<tr>
<td>$k_B$</td>
<td>Boltzmann constant (1.3806503 x 10$^{23}$m$^2$kgs$^{-1}$K$^{-1}$)</td>
</tr>
<tr>
<td>$E_{cell}$</td>
<td>PEFC cell operating potential (V)</td>
</tr>
<tr>
<td>$i_0$</td>
<td>Exchange current density (mA cm$^{-2}$)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Charge transfer coefficient symmetry factor</td>
</tr>
<tr>
<td>$[ox]$</td>
<td>Oxidised reactant concentration</td>
</tr>
<tr>
<td>$[red]$</td>
<td>Reduced reactant concentration</td>
</tr>
<tr>
<td>$p_0$</td>
<td>Standard pressure (atm)</td>
</tr>
<tr>
<td>$p_{H_2}$</td>
<td>Hydrogen gas partial pressure</td>
</tr>
<tr>
<td>$p_{O_2}$</td>
<td>Oxygen gas partial pressure</td>
</tr>
<tr>
<td>$p_{H_2O}$</td>
<td>Water vapour partial pressure</td>
</tr>
<tr>
<td>$a_{H_2}$</td>
<td>Activity of molar hydrogen in gas phase</td>
</tr>
<tr>
<td>$p_{O_2}$</td>
<td>Activity of molar oxygen in gas phase</td>
</tr>
<tr>
<td>$a_{H_2O}$</td>
<td>Activity of molar water in vapour phase</td>
</tr>
<tr>
<td>$\eta_k$</td>
<td>Kinetic cell overpotential (V)</td>
</tr>
<tr>
<td>$R_{elec}$</td>
<td>PEFC electrode ohmic resistance (V)</td>
</tr>
<tr>
<td>$\eta_{mt}$</td>
<td>Mass transport cell overpotential (V)</td>
</tr>
<tr>
<td>$\eta_0$</td>
<td>Ohmic cell overpotential (V)</td>
</tr>
<tr>
<td>$R_{elec}^{H+}$</td>
<td>PEFC electrode proton resistance (Ω)</td>
</tr>
<tr>
<td>$R_{elec}^{-1}$</td>
<td>PEM proton resistance (Ω)</td>
</tr>
<tr>
<td>$\eta_{membrane}^{H+}$</td>
<td>PEM proton resistance overpotential (V)</td>
</tr>
<tr>
<td>$E_{ORR}$</td>
<td>Oxygen Reduction Reaction thermodynamic equilibrium potential (V)</td>
</tr>
<tr>
<td>$E_{SHE}$</td>
<td>Standard Hydrogen Electrode potential (V)</td>
</tr>
<tr>
<td>$E_{COR}$</td>
<td>Carbon Oxidation Reaction thermodynamic potential (V)</td>
</tr>
<tr>
<td>$E_{KIN}$</td>
<td>Limiting cell potential for kinetic stability of carbon (V)</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Permitivity of medium</td>
</tr>
<tr>
<td>$\epsilon_0$</td>
<td>Permitivity of free space (8.854187817 x 10$^{-12}$Fm$^{-1}$)</td>
</tr>
<tr>
<td>$m_e$</td>
<td>Electron mass (kg)</td>
</tr>
<tr>
<td>$Z$</td>
<td>Nuclear charge</td>
</tr>
<tr>
<td>$V_{Ze}$</td>
<td>Nuclear-electron interaction potential</td>
</tr>
<tr>
<td>$V_{ee}$</td>
<td>Electron-electron interaction potential</td>
</tr>
<tr>
<td>$H_{KIN}$</td>
<td>Kinetic energy Hamiltonian operator</td>
</tr>
</tbody>
</table>
Table 10 – Continued on previous page

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Psi_{HF}$</td>
<td>Hartree Fock wavefunction</td>
</tr>
<tr>
<td>$V_{ext}$</td>
<td>External (nuclear) potential experienced by electrons</td>
</tr>
<tr>
<td>$\psi_e(\vec{r},t)$</td>
<td>Time-dependent many electron wavefunction</td>
</tr>
<tr>
<td>$\psi_e(\vec{r})$</td>
<td>Stationary many electron wavefunction</td>
</tr>
<tr>
<td>$E_x^{FOCK}$</td>
<td>Fock exchange energy</td>
</tr>
<tr>
<td>$E_v[\tilde{n}]$</td>
<td>Total energy functional of non-interacting charge density</td>
</tr>
<tr>
<td>$n(r)$</td>
<td>Charge density of non-interacting electrons</td>
</tr>
<tr>
<td>$E_v[n]$</td>
<td>Total electrostatic energy functional of interacting charge density</td>
</tr>
<tr>
<td>$E_{xc}[n]$</td>
<td>Exchange-Correlation Functional of charge density $n(r)$</td>
</tr>
<tr>
<td>$E_x[n]$</td>
<td>Exchange Energy</td>
</tr>
<tr>
<td>$E_c[n]$</td>
<td>Correlation Energy</td>
</tr>
<tr>
<td>$U[n]$</td>
<td>Hartree potential energy</td>
</tr>
<tr>
<td>$T[n]$</td>
<td>(Non-interacting) kinetic energy</td>
</tr>
<tr>
<td>$\tilde{n}_0(r)$</td>
<td>Minimum ground state charge density</td>
</tr>
<tr>
<td>$\nu_{xc}(r)$</td>
<td>Exchange correlation local potential</td>
</tr>
<tr>
<td>$A(N)$</td>
<td>Normalised coefficient for kinetic energy as function of number of electron gas states</td>
</tr>
<tr>
<td>$B(N)$</td>
<td>Normalised coefficient for electron interaction energy as function of number of electron gas states</td>
</tr>
<tr>
<td>$E_{x}^{\text{homo}}[n(r)]$</td>
<td>Homogeneous exchange potential energy</td>
</tr>
<tr>
<td>$u_{\vec{k},i}$</td>
<td>Bloch function</td>
</tr>
<tr>
<td>$C_{i,K}^k$</td>
<td>Wavefunction plane wave expansion coefficient</td>
</tr>
<tr>
<td>$\chi(\alpha_j;\vec{r})$</td>
<td>Gaussian type function (GTF) of diffusivity exponent $\alpha_j$</td>
</tr>
<tr>
<td>$d_i^m$</td>
<td>Atomic orbital (AO) contraction coefficient</td>
</tr>
<tr>
<td>$C_{j,m}$</td>
<td>AO Shell $m, l$ normalization constant</td>
</tr>
<tr>
<td>$X_{j}^{m}(\vec{r})$</td>
<td>Spherical harmonic $ml$</td>
</tr>
<tr>
<td>$G(\alpha_j;\vec{r} - \vec{A})$</td>
<td>Gaussian function centroid $\vec{A}$</td>
</tr>
<tr>
<td>$a_{\mu,i}(\vec{k})$</td>
<td>Bloch expansion coefficient</td>
</tr>
<tr>
<td>$\phi_i(\vec{r},\vec{k})$</td>
<td>Bloch function</td>
</tr>
<tr>
<td>$t_{2g}$</td>
<td>Threefold degenerate $d$ state in electronic structure</td>
</tr>
<tr>
<td>$e_g$</td>
<td>Twofold degenerate $d$ state in electronic structure</td>
</tr>
<tr>
<td>$D_q$</td>
<td>Crystal Field Stabilisation energy (eV)</td>
</tr>
<tr>
<td>$U$</td>
<td>Pairing energy</td>
</tr>
<tr>
<td>$H_{\text{Si-H}}^{\text{Ge}}$</td>
<td>Hamiltonian of Si-H acceptor complex in Ge</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Photon frequency ($s^{-1}$)</td>
</tr>
<tr>
<td>$E_g$</td>
<td>Band gap (eV)</td>
</tr>
<tr>
<td>$E_F$</td>
<td>Fermi energy (eV)</td>
</tr>
<tr>
<td>$E_C$</td>
<td>Conduction band minimum energy (eV)</td>
</tr>
<tr>
<td>$E_V$</td>
<td>Valence band maximum energy (eV)</td>
</tr>
</tbody>
</table>

Continued on next page
Table 10 – Continued on previous page

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{k}_c$</td>
<td>Conduction band minimum wavevector (Å$^{-1}$)</td>
</tr>
<tr>
<td>$\tilde{k}_v$</td>
<td>Valence band maximum wavevector (Å$^{-1}$)</td>
</tr>
<tr>
<td>$N_{D^+}$</td>
<td>Total number of ionised donor atoms</td>
</tr>
<tr>
<td>$N_h$</td>
<td>Total number of holes</td>
</tr>
<tr>
<td>$N_e$</td>
<td>Total number of electrons</td>
</tr>
<tr>
<td>$N_{D^0}$</td>
<td>Total number of unionised donor atoms</td>
</tr>
<tr>
<td>$V_R$</td>
<td>Long range perturbing Coulomb potential (eV)</td>
</tr>
<tr>
<td>$V_R^p$</td>
<td>Short range perturbing potential (eV)</td>
</tr>
<tr>
<td>$V_R^T$</td>
<td>Periodic component of $V_R$ (eV)</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>Multiplicity of defect configuration (binary system only)</td>
</tr>
<tr>
<td>$\epsilon_k$</td>
<td>Quasiparticle energy state of eigenvalue $k$ (eV)</td>
</tr>
<tr>
<td>$j$</td>
<td>Bulk electron current (A)</td>
</tr>
<tr>
<td>$E$</td>
<td>Electric Field (Vm$^{-1}$)</td>
</tr>
<tr>
<td>$V_{\text{sample}}$</td>
<td>Total sample volume (m$^{-3}$)</td>
</tr>
<tr>
<td>$v_k$</td>
<td>Group Velocity (ms$^{-1}$)</td>
</tr>
<tr>
<td>$F(k)$</td>
<td>Occupation function of quasiparticle states $k$</td>
</tr>
<tr>
<td>$\mu_{e,h}$</td>
<td>Electron/hole mobility</td>
</tr>
<tr>
<td>$m_{\text{eff},k}$</td>
<td>Quasiparticle effective mass (kg)</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Total scattering time (s)</td>
</tr>
<tr>
<td>$\tau_{\text{imp}}$</td>
<td>Impurity scattering time (s)</td>
</tr>
<tr>
<td>$\tau_{\text{ep}}$</td>
<td>Electron phonon scattering time (s)</td>
</tr>
<tr>
<td>$\tau_C$</td>
<td>Coulomb scattering time (s)</td>
</tr>
<tr>
<td>$n_{\text{imp}}$</td>
<td>Concentration of impurity atoms in sample (m$^{-3}$)</td>
</tr>
<tr>
<td>$\Delta G_f^{\text{DFT}}$</td>
<td>Gibbs free energy of formation (eV per formula unit) (DFT)</td>
</tr>
<tr>
<td>$G_{\text{perf},q}$</td>
<td>Total energy (eV per formula unit) of perfect host material</td>
</tr>
<tr>
<td>$G_{f,q}$</td>
<td>Total energy (eV per formula unit) of defect in charge state $q$</td>
</tr>
<tr>
<td>$\Delta n_i$</td>
<td>Number of atoms of species $i$</td>
</tr>
<tr>
<td>$E_{\text{VB}M}$</td>
<td>Energy of valence band maximum in perfect host (eV)</td>
</tr>
<tr>
<td>$\mu_e$</td>
<td>Electron chemical potential (eV)</td>
</tr>
<tr>
<td>$\rho_{\mu,\nu}^n$</td>
<td>Density Matrix Element for atomic orbitals $\mu,\nu$</td>
</tr>
<tr>
<td>$H_K^K$</td>
<td>Nuclear potential and kinetic energy density matrix in $K$ space</td>
</tr>
<tr>
<td>$F_K^{\mu,\nu}$</td>
<td>Fock and Coulomb exchange energy density matrix in $K$ space</td>
</tr>
<tr>
<td>$Q$</td>
<td>Quadrupole moment of aperiodic charge density</td>
</tr>
<tr>
<td>$q$</td>
<td>Monopole moment of periodic charge density</td>
</tr>
<tr>
<td>$\Delta V_{PS}$</td>
<td>Pseudopotential perturbation potential (eV)</td>
</tr>
<tr>
<td>$\rho_{D+H}$</td>
<td>Charge density of defect + host with optimised geometry</td>
</tr>
<tr>
<td>$\rho(D+H)-D$</td>
<td>Charge density of final host geometry without defect calculated by static SCF cycle</td>
</tr>
<tr>
<td>$C$</td>
<td>Stochiometric coefficient matrix</td>
</tr>
<tr>
<td>$J_i$</td>
<td>Effective cluster interaction</td>
</tr>
</tbody>
</table>
Table 10 – *Continued on previous page*

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi_{mn}$</td>
<td>$n$-nearest neighbour correlation function of crystal structure $m$</td>
</tr>
<tr>
<td>$\sigma_i$</td>
<td>Cluster expansion occupation variable</td>
</tr>
<tr>
<td>$CV$</td>
<td>Cross Validation score</td>
</tr>
<tr>
<td>$\Delta G_{Vac(O)}^{\text{f}}$</td>
<td>Oxygen vacancy formation (eV per formula unit)</td>
</tr>
<tr>
<td>$E_{CE}$</td>
<td>Cluster Expansion predicted formation energy (eV per formula unit)</td>
</tr>
<tr>
<td>$\sigma_{Ta}$</td>
<td>Conductivity of Ta/Sn disordered rutile alloy</td>
</tr>
<tr>
<td>$\sigma_H$</td>
<td>Conductivity of interstitial hydrogen doped SnO$_2$</td>
</tr>
<tr>
<td>$n_{e^{-}}$</td>
<td>Concentration of conduction band electrons m$^{-3}$ of interstitial H doped SnO$_2$</td>
</tr>
<tr>
<td>$H_{\text{f}}^{+}$</td>
<td>Formally charged interstitial hydrogen defect treated with jellium</td>
</tr>
<tr>
<td>$\alpha_{kl}, C_{kl}$</td>
<td>GTF based pseudopotential coefficients in CRYSTAL09</td>
</tr>
</tbody>
</table>
References


REFERENCES


[34] Santos E., Catalysts in Electrochemistry: From Fundamental Aspects to Strategies for Fuel Cell Development, Wiley 2011


164
REFERENCES


[45] T.J. Schmidt, C. Coperet, A. Neels, M. Lennartz and D. Kramer ”Highly Durable Oxide-based Catalysts for Polymer Electrolyte Fuel Cells (DuraCAT)”, Competence Center Energy and Mobility, Switzerland (2011)


[60] J.K. Kim et al, Carbon 77, 525537


REFERENCES

[68] Shao M., J. Power Sources 196, 2433-2444 (2011)
[82] M.S. Saha et al, international journal of hydrogen energy 37, 4633-4638 (2012)
[87] S.-Y. Huang et al. / Applied Catalysis B: Environmental 102, 7177 (2011)
REFERENCES

[94] Y. Ju et al, Advances in Condensed Matter Physics, Volume 2013, Article ID 365475
[100] Highly Durable Oxide-based catalysts for Polymer Electrolyte Fuel Cells, CCEM projects- Input for the Annual Activity Report
REFERENCES

Dordrecht 1987)


REFERENCES


[146] Earle M.D., Phys. Rev. 61, 56 (1941)


REFERENCES


REFERENCES


[243] Barthelat J.C. Private communication