

Ab-initio investigation of the antimony- vacancy complex and related defects in germanium.

Presented in partial fulfillment of the requirements of the degree MSc (Physics)

University of Pretoria

Geoffrey Webb

In the Faculty of Natural & Agricultural Sciences

University of Pretoria

2015

Supervisor

Prof. W.E. Meyer

Co-Supervisor

Dr. R.C. Andrew

Declaration

I, Geoffrey Webb, student number 04351045, declare that this dissertation, which I hereby submit for the degree of MSc (Physics) at the University of Pretoria, is my own work and has not been submitted to any other tertiary institution.

Signed:

Geoffrey Webb

On this day:

Acknowledgments

I would like to make mention of the following persons for their contributions towards the completion of this body of work:

- Firstly and most importantly a special thanks to my supervisor Prof Walter Meyer, who stood by me and believed in me, guided and advised me for many years. I thank you for your time, your patience and unwavering support since my first year as an undergraduate to completion of this dissertation.
- To the NRF for their financial support throughout the duration of my MSc. studies for the awarding to me of the Innovation Masters Scholarship. Without which this task would've been insurmountable.
- Prof D. Auret for being a mentor and an outstanding role model for me to look up to throughout my studies.
- To Dr. S.M.M. Coelho for his endless patience in the lab, teaching me the finer points of semiconductor physics.
- To my colleagues C. Ouma and E. Igumbor for their help and guidance throughout this dissertation.
- To Dr. J. Pretorius for setup and maintenance of the computational facilities.
- To my family and friends and those I love for their dedication to me and the endless encouragement to complete this piece of work.
- Finally Prof Chris Theron for believing in me and granting me the chance to achieve.

Abstract

Ab-initio investigation of the antimony-vacancy complex and related defects in germanium.

Submitted in partial fulfilment of the degree MSc (Physics) in the Faculty of Natural & Agricultural Sciences, University of Pretoria

Supervisor: Prof. W.E. Meyer
Co- Supervisor: Dr. R. Andrew

Recent advances in computational technology and algorithms have made it feasible to accurately model the electronic structures of solids by means of density functional theory. The development of hybrid functionals have improved the accuracy of band gap calculations and made it possible to make qualitative predictions regarding the charge transition energy levels of defects in semiconductors.

The Sb-V defect (also known as the *E*-center) in germanium is a well-known defect, which have been the subject of many experimental and some theoretical studies. It has been found to have interesting annealing properties and the aim of this study is to investigate the electronic properties of the Sb-V defect theoretically. The vacancy defect in germanium (V_{Ge}), the antimony substitutional (Sb_{Ge}) defect in germanium and the defect complex (Sb-V) arising from the combination of these two defects is explored in great detail and how they interact in proximity to one another is presented here. In addition, this work can be seen as a test for the effectiveness of the technique to model defects in semiconductors correctly.

The *E*-center defect was investigated using the HSE06 hybrid functional as implemented in the VASP code. A positive binding energy of 1.5 eV, 1.02 eV and 0.88 eV was found for the first, second and third nearest neighbor configurations respectively, between the Sb and the vacancy was predicted. No metastability was detected and the nearest-neighbor configuration had the lowest energy for all charge states. Four transition levels in the band gap were predicted, with energy level relative to the valence band maximum, lying at 0.52 eV (-2/-1), 0.40 eV (-1/0), 0.44 eV (0/+1) and 0.02 eV (+1/+2). The two mid-gap levels (-1/0) and (0/+1) had negative-*U* ordering with $U = -0.04$ eV.

These findings were consistent with the current experimental model of the Sb-V complex in germanium whereby no metastability has been observed experimentally. The energy level of the (-2/-1) corresponded well with the experimental DLTS level in n-type material at 0.37 eV, though the correspondence for the other levels was not as good. Experimentally, no negative- U behavior was observed, but the predicted negative- U behavior was rather small and no deliberate experiments have been performed to investigate the presence of negative- U behavior in the Sb-V complex.

Table of contents

Ab-initio investigation of the antimony-vacancy complex and related defects in germanium.....	i
Declaration.....	i
Acknowledgments.....	ii
Abstract	iii
Table of contents	v
Table of figures	ix
Table of tables	xi
List of acronyms.....	xii
Prelude.....	xiv
Chapter 1 Ab-initio techniques.....	1
1.1 The many-body Hamiltonian	1
1.2 The Born-Oppenheimer approximation.....	2
1.3 Mean-field ab-initio methods	2
1.3.1 The Hartree approximation: no exchange, averaged correlation.....	3
1.3.2 The Hartree-Fock approximation, explicit exchange, averaged correlation	4
1.4 Density functional theory (DFT).....	6
1.4.1 The Hohenberg-Kohn theorems, a formal derivation	8
1.4.2 The Kohn-Sham auxiliary system.....	12
1.4.3 The Kohn-Sham variational equations	15
1.4.4 Solving the Kohn-Sham equation: The self-consistent Kohn-Sham equations	17
1.5 DFT functionals	21

1.5.1	The LSDA approximation.....	22
1.5.2	Generalized gradient approximation (GGA).....	23
1.5.3	Hybrid functionals	23
1.6	Basis sets.....	25
1.6.1	The plane-wave basis set.....	25
1.7	Brillouin zone sampling.....	27
1.8	Pseudopotentials.....	27
Chapter 2 The Science of semiconductors.....		29
2.1	Introduction	29
2.2	Doping of semiconductors	29
2.3	Structural properties of semiconductor defects.....	31
2.3.1	Primary and secondary defects.....	31
2.4	Electronic properties of semiconductor defects.....	33
2.5	Charge distribution in semiconductors	34
2.5.1	Schottky barriers	35
2.6	Electrical characterization of semiconductor defects	38
2.6.1	Capacitance–voltage technique (C-V).....	38
2.6.2	Current–voltage technique	39
2.6.3	Deep level transient spectroscopy	39
2.7	Thermodynamic properties of defects.....	40
2.7.1	Formation energy.....	40
2.7.2	Activation energy of defect migration	40
2.7.3	Binding energy of defect complexes	41
2.8	Annealing properties.....	41
2.9	The Poole-Frenkel effect	42
Chapter 3 Defect metastability.....		43

3.1	Introduction	43
3.2	Bistability and Multistable defects	44
3.3	Charge-state-controlled metastability	45
3.4	Negative-U defects.....	46
3.4.1	Mechanisms leading to negative-U behavior	47
3.4.2	Properties of negative-U defects	48
Chapter 4 Modern techniques applied to computational defect studies		49
4.1	Ab-initio techniques applied to bulk properties	49
4.2	Ab-initio techniques applied to electronic properties.....	50
4.2.1	Errors related to prediction of the band gap and correction methods.....	50
4.3	Defect properties.....	52
4.3.1	Formation energy.....	52
4.3.1.1	Setting of boundary conditions.....	53
4.3.1.2	Brillouin zone sampling	55
4.3.1.3	Charged defect states.....	56
4.3.1.4	Spin orbit coupling - spin polarization	57
4.3.2	Thermodynamic transition levels.....	57
4.3.3	Optical transition levels.....	58
4.3.4	Defect diffusion and propagation	58
Chapter 5 Computational conditions for this study		59
5.1	Summary of calculations and boundary conditions.	59
5.2	Convergence criteria	59
5.2.1	Unit cell	60
5.2.2	Supercell size.....	60
5.2.3	Plane-wave cut-off energy	61
5.2.4	Brillouin zone sampling	62

5.3	Using DFT to predict metastability.....	62
5.3.1	The minimum energy configuration	63
5.3.2	Formation energy difference and occupation.....	64
5.3.3	Application of electrical bias and relation to charge state controlled metastability	64
Chapter 6 The antimony-vacancy (Sb-V) complex, the E -center defect in germanium		66
6.1	The experimental results of Fage-Pederson and Larsen.....	66
6.1.1	Experimental details and results of the paper.....	66
6.1.2	The $E_{0.37}$ defect.....	69
6.1.3	Summary of findings of Larsen	73
6.2	The modelling results.....	73
6.2.1	The V_{Ge} defect.....	73
6.2.2	The Sb-substitutional defect.....	75
6.2.3	The Sb-V complex.....	77
6.3	Discussion of the Sb-V complex in germanium.....	80
6.3.1	The vacancy defect	80
6.3.2	The substitutional antimony defect.....	82
6.3.3	The Sb-V complex.....	83
Chapter 7 Conclusions		86
Works Cited.....		88

Table of figures

Figure 1.1: A representation of the Hohenberg-Kohn theorem.	9
Figure 1.2: A representation of the Kohn-Sham theorem.....	13
Figure 1.3: Schematic representation of the self-consistent loop for the solution of the Kohn-Sham equations.	18
Figure 1.4: Calculation of the density using Fourier transforms and grids.	20
Figure 1.5: Comparison of a wave function in the Coulomb potential of the nucleus (dotted) to the pseudopotential (solid), the real and the pseudo wave function and potentials match above a certain cut-off radius.	28
Figure 2.1: Schematic representation of a semiconductor doped with a donor.....	30
Figure 2.2: Schematic representation of a semiconductor doped with an acceptor.....	31
Figure 2.3: Point defects in semiconductor lattice.....	32
Figure 2.4: Secondary defect complexes in the semiconductor lattice.....	32
Figure 2.5: Schematic representation of an n-type and p-type Schottky barrier at equilibrium.....	36
Figure 2.6: Representation of a Schottky barrier under forward and reverse bias for an n-type semiconductor.	36
Figure 2.7: Representation of a Schottky barrier under forward and reverse bias for a p-type semiconductor.	37
Figure 2.8 Visual representation of the migration activation energy of a point defect in a lattice.	41
Figure 3.1: Configuration diagrams from usual semiconductor defects. Potential configuration diagrams for charge state dependent metastability	44
Figure 3.2: A configuration co-ordinate diagram showing charge state controlled metastability in two configurations	45
Figure 4.1: A representation of the pristine Germanium 64-atom unit cell.....	55

Figure 4.2: Schematic representation of the Brillouin zone, the location of the Γ -point and the various other points.....	56
Figure 5.1: Diagram of formation energy as the size of the supercell increases.....	60
Figure 5.2: The formation energy of a 8 atom cell as a function of the plane-wave cut-off.	61
Figure 5.3: The energy convergence of a 64 atom supercell with a cut-off energy of 400 eV	62
Figure 5.4: Formation energy as a function of Fermi level for varying charge states.....	65
Figure 6.1: DLTS spectra recorded on Sb1 and Sb2, three days after 4×10^{13} cm ⁻² electron irradiation.	67
Figure 6.2: Electronic signatures of all observed electron traps in Sb1 and Sb2.....	68
Figure 6.3: DLTS spectra from Sb2 as a comparison of electron and proton irradiation and annealed at 110°C for 15 minutes.....	68
Figure 6.4: E center depth profiles in two diodes on Sb2, which were annealed with either a 4V reverse bias or with no bias.	70
Figure 6.5: Formation energy as a function of the Fermi Level (eV) at various charge states for the vacancy defect in bulk germanium.	74
Figure 6.6: Formation energy as a function of the Fermi Level (eV) at various states for the Sb substitutional defect.....	76
Figure 6.7: Figure showing the positioning of the defect complex positions with respect to each other in bulk germanium.....	77
Figure 6.8: Formation energy of the Sb-V complex as a function of the Fermi Level (eV) at various charge states for the three defined configurations.	79
Figure 6.9: Comparison of the defect level found in different studies of the germanium vacancy. ...	80

Table of tables

Table 5.1: An example of HSE06 calculated formations energies of two different defect states the minimum energy is shown in bold in each case.	63
Table 6.1: The properties of relevant electron and hole traps observed by Larsen in the Irradiation study.....	67
Table 6.2: Ge was doped with either the indicated elements and irradiated with the indicated proton or electron source.	71
Table 6.3: Calculated standard energies of formation (eV) with the Fermi level at the VBM of the germanium vacancy defect in each of the explored charge states obtained by use of the HSE06 functional.	74
Table 6.4: Thermodynamic charge state transition levels (eV) relative to the valence band maximum (VBM) using HSE06 functionals for the vacancy defect in germanium.	75
Table 6.5: Calculated formation energies (eV) at the VBM of the SbGe Defect in germanium relative to the pristine bulk system of germanium in each of the explored charge states obtained by use of the HSE06 functional.....	75
Table 6.6: Thermodynamic charge state transition levels (eV) relative to the valence band maximum (VBM) using HSE06 functionals for the antimony substitutional defect in germanium.....	76
Table 6.7: Calculated formation energies (eV) of the Sb-Vcomplex in germanium relative to the pristine bulk system of germanium in each of the explored charge states obtained by use of the HSE06 functional, with respect to the configuration of the defect.	78
Table 6.8: The binding energies as calculated by making use of eq (6.2) for the Sb-V complex in germanium.	78
Table 6.9: Thermodynamic charge state transition levels (eV) relative to the valence band maximum (VBM) using HSE06 functionals for the Sb-V complex.....	79

List of acronyms

BOA	Born-Oppenheimer approximation
DFT	Density functional theory
E_A	Acceptor level
E_C	Conduction band edge
E_D	Donor level
E_F	Fermi level
E_V	Valence band edge
GGA	Generalized gradient approximation
HA	Hartree approximation
HF	Hartree-Fock
HK	Hohenberg and Kohn
KS	Kohn-Sham
LDA	Local density approximation
MBSE	Many-body Schrödinger equation
MEP	Minimum energy path
MP	Monkhorst-Pack
NEB	Nudged elastic band
PAW	Projector augmented wave
PES	Potential energy surface
QMC	Quantum Monte Carlo
RT	Room Temperature
TSS	Transition state search
VASP	Vienna Ab-initio Simulation Package
XC	Exchange-correlation

Prelude

Semiconductors have changed the world. They have single handedly been the driving force behind the digital revolution. One could argue that, in building the information society, they have been as important as the invention of structural steel, poured concrete and the steam engine for the industrial revolution. The history of semiconductors and the reasons why they have been researched so deeply in our lifetimes is a far longer topic than this dissertation could ever be. Since the effect of semi conductivity was first observed in 1822 by Michael Faraday, who noticed a decrease of resistance with temperature in silver sulphide, the world has changed more than anyone could have predicted. Fast forward 120 years of study and research and the first germanium transistor was built in 1947 by John Bardeen and Walter Brattain and so the transistor was born in the first p-n junction. The development of the bi-polar transistor and tunnel diodes, integrated circuits to eventually the advent of lasers have followed from that point in 1947 providing an almost endless supply of new avenues to pursue in the scientific community. Most notable is that 28 of 108 Nobel prizes awarded in history have been awarded in the field of condensed matter. The field of semiconductor research has been a hotbed of innovation and cutting edge science for nearly 200 years.

Since the first germanium transistor, germanium slowly gave way to the vastly more popular semiconductor: silicon. Silicon, being readily available, cheap and having very appropriate electrical and chemical properties rose to become the substance to change the world. Germanium on the other hand fell into the chapters of history only in recent years to be re-awakened by a new group of researchers to re-examine how best this material can be used to pioneer the 21st century in the use of more sensitive devices making use of the much smaller band gap that germanium has.

In October 2000 J. Fage-Pedersen and A. Nylandsted Larsen published an article examining the irradiation-induced defects in germanium by methods of transient spectroscopy (Fage-Pedersen & Nylandsted Larsen, 2000). This paper studied n-type germanium doped with antimony and characterized several majority-carrier traps and one minority carrier trap. They characterized the antimony-vacancy defect complex (*E*-center) $E_{0.37}$ but it was found to anneal in a way which was deemed fundamentally different to that in silicon. The interesting point was that the effect was observed in p-type doped silicon and in n-type doped germanium. This type of common similarity in annealing properties was not yet seen in materials which differed so largely in composition. This

puzzled the authors at the time and has remained to this day an unexplored property of germanium.

In 2005 N.R. Zangenberg and A. Nylandsted Larsen continued the research publishing investigations of vacancy related defects in low-temperature electron irradiated, boron-doped silicon (Zangenberg & Nylandsted Larsen, 2005) and it was found that there existed a low temperature peak corresponding to a shallow trap at $E_V + 0.105$ eV. Zangenberg et al. observed charged state controlled metastability in the B-V center and this led to speculation that a similar form of metastability could exist in the antimony doped germanium. This odd annealing trend became the key factor used to make the link between the existence of metastability and this unexplained annealing trend. Recently by making use of modern DFT techniques, the B-V complex in silicon has been successfully modelled showing charge state controlled metastability (Ouma & Meyer, currently under review).

In an investigative study the technique that was used to successfully characterize the B-V complex in silicon is used in an attempt to prove the existence of metastability of the Sb-V defect complex in germanium. It was applied in an attempt to explain the strange annealing patterns observed by Larsen and subsequent studies relating to this E-center defect (Coelho, et al., 2013) (Nyamhere, et al., 2011) (Petersen & Nylandsted Larsen, 2010) (Auret, et al., 2008). However unlike the B-V complex in Silicon, to the best of our knowledge little evidence is reported in modern published work of the existence of metastability in the Sb-V complex or related defects. These annealing characteristics and the result of charge-state controlled metastability in the B-V complex in silicon however warranted investigation and in this work the focus is set on the Sb-V complex as the goal to determine the existence of, or possible explanation of, metastability. The Sb substitutional defect and the vacancy defect in germanium are analyzed using the same techniques to create a more complete understanding of the process under which these defects are allowed to form and exist.

This dissertation is divided into a theoretical background of the DFT technique, a brief overview of semiconductor physics, metastability of semiconductors, the computational methodology of the technique used in this dissertation and the results and conclusions of the study

Chapter 1

Ab-initio techniques

1.1 The many-body Hamiltonian

In order to understand the properties of matter the first step to begin with is to understand the interactions of electrons and their nuclei. Using only the most basic of particle properties such as the relative charge, the atomic number Z , and mass m , the properties of materials can be determined by solving the Many-particle, time-independent Schrödinger equation. (Griffiths, 2005), (Patterson & Bailey, 2010), (Springborg, 2000):

$$\left[-\sum_{i=1}^N \frac{1}{2m_i} \nabla_{r_i}^2 + \sum_{i>j}^N \frac{Z_i Z_j}{|r_i - r_j|} \right] \Psi(\mathbf{R}) = E\Psi(\mathbf{R}) = \hat{H}\Psi(\mathbf{R})[\hat{T} + \hat{V}]\Psi(\mathbf{R}) \quad (1.1)$$

This is a $3N$ -dimensional eigen-problem, where \mathbf{R} is the collective co-ordinate for all N particles r_1, \dots, r_N . $\Psi(\mathbf{R})$ is the wave function of all particles, required to be anti-symmetric under the exchange of two electrons and obeying the Pauli exclusion principle. This equation is formulated in atomic units ($e = m_e = \hbar = 4\pi\epsilon_0 = 1$) and \hat{T} and \hat{V} are the kinetic and Coulomb potential energy respectively.

The Coulomb potential term causes the differential equation (1.1) to be inseparable for more than a two particle problem and thus the simplest technique of separation of variables is excluded. This implies that the form of the wave function is no longer a simple product of one-electron orbitals when extended to beyond a two body problem.

Yet all materials of interest contain a large number of interacting particles, which means approximations must be made in order to reduce the complexity of the resulting Schrödinger equation (SE), since once the eigen-states of the system are known, many properties may be calculated. However these approximations may significantly affect the accuracy of the predictions of the system under investigation.

The following sections discuss the principal approaches to approximating the solution of the SE for real materials. Starting off with orbital based methods which approximate $\Psi(\mathbf{R})$ as a Slater determinant of single particle orbitals (Hartree-Fock theory (HF)) and followed by an introduction to density functional theory (DFT), which is based fundamentally on the charge density of the system rather than a wave function.

1.2 The Born-Oppenheimer approximation

One of the first approximations made to simplify the description of the interactions of particles and is used in almost all of the methods discussed within this dissertation is the Born-Oppenheimer approximation (BOA) (Patterson & Bailey, 2010); (Springborg, 2000); (Born & Oppenheimer, 1927). In this approximation, the nuclei are held in fixed positions in order to separate the nuclear and electronic co-ordinates from one another. This approximation is made on the basis that the mass of the nuclei is much larger than that of the electrons. Thus the BOA posits that since the nuclear velocities are much slower compared to those of the electrons, the nuclei are assumed to be stationary. The Hamiltonian of the system thus can be written as:

$$\hat{H} = \left[-\sum_{i=1}^N \frac{1}{2} \nabla_{r_i}^2 + \sum_i^N \sum_a^K \frac{Z_a}{|r_i - R_a|} + \sum_{i>j}^N \frac{1}{|r_i - r_j|} + \frac{1}{2} \sum_{a>b}^K \frac{Z_a Z_b}{|R_a - R_b|} \right] \quad (1.2)$$

where the nuclear motion and electron motion have been decoupled. This BOA Hamiltonian is still hard to solve with no analytic solution existing for more than one electron.

1.3 Mean-field ab-initio methods

The evolution of the approach of considering the interactions of each electron with every other electron was the development of the mean-field approach, which replaces the effect of all other electrons with a mean field. This allows the many body Schrödinger equation to be decoupled into a number of single particle Schrödinger equations. In the following sections two approaches are considered.

1.3.1 The Hartree approximation: no exchange, averaged correlation

The Hartree approximation (HA) is a common ab-initio approach for solving the many-body SE. The main idea is to separate the many-electron SE into many simpler one-electron equations (Patterson & Bailey, 2010) (Springborg, 2000) (Hichliffe, 2000). The justification for this is that the behavior of each electron is described within the net field of all other electrons, that is the electron experiences a mean-field potential,

$$V^{el}(\mathbf{r}) = -e \int \frac{1}{|\mathbf{r}-\mathbf{r}'|} \rho(\mathbf{r}') d\mathbf{r} \quad (1.3)$$

where each one-electron equation will yield a single-electron wave function, ψ_i , called an orbital and a corresponding orbital energy. The total electronic charge density is given by,

$$\rho(\mathbf{r}) = -e \sum_i^{N_e} |\psi_i(\mathbf{r})|^2 \quad (1.4)$$

where the sum is over all occupied levels, with the ion potential as

$$V^{ion}(\mathbf{r}) = -\sum_{\mathbf{R}}^{N_i} \frac{Z}{r-R} \quad (1.5)$$

where \mathbf{R} is the nuclear position and the potential of the system is given simply as $V = V^{ion} + V^{el}$.

Since the electrons are assumed to be independent (non-interacting), the N -electron wave function is written as a product of one-electron wave functions:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \psi_1(\mathbf{r}_1) \psi_2(\mathbf{r}_2) \dots \psi_N(\mathbf{r}_N) \quad (1.6)$$

Making use of the variational principle and minimizing the expectation value of the Hamiltonian with respect to the wave function variations produces a set of one-electron equations, called the Hartree equations:

$$-\frac{1}{2} \nabla^2 \psi_i(\mathbf{r}) + V^{ion}(\mathbf{r}) \psi_i(\mathbf{r}) + \left[e^2 \sum_j \int \frac{1}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' |\psi_j(\mathbf{r}')|^2 \right] \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}) \quad (1.7)$$

The self-interaction error of the Hartree method

A subtle yet very important note to be made about the Hartree method is that the electron potential term Equation (1.3) includes a non-physical repulsive interaction between each electron and itself as the electron interacts with the mean potential field which is calculated from $|\psi_i(\mathbf{r})|^2$, which includes the average effect of itself. The error in the energy is called the self-interaction error. This error is an important source of error within the approximation of functionals.

1.3.2 The Hartree-Fock approximation, explicit exchange, averaged correlation

The Hartree equations were a step forward in the first attempt at solving the many-body SE, this was only possible due to the Hartree approximations. Yet the Hartree approximation inadequately describes important properties of electrons, such as quantum indistinguishability and explicit averaged Coulomb correlation. Quantum mechanics requires the wave function to be indiscriminant as to which electron is in which state as all electrons are seen to be identical. This gives rise to the two types of quantum particles: Fermions and Bosons. Electrons fall into the class of Fermions, who by definition of the Pauli exclusion principle, exhibit anti-symmetric wave function behavior under the exchange of two particles. The requirement of the Pauli exclusion principle leads to the exchange energy of electrons, which can be thought of as another means of minimizing the Coulomb energy. The correlation energy refers specifically to the electron-electron interaction and mean-field calculation methods approximate this as an average effect of the Coulomb repulsion.

The Hartree-Fock approximation (Dal-Pino, 1993) builds on the Hartree approximation to include the indistinguishability and exchange properties of electrons, but still keeps the same mean-field approach to the electron correlation in order to use the one-electron equations. The correlation term is often defined based on the amount of correlation Hartree-Fock overlooks:

$$E_{corr} = E_{exact, non-relativistic} - E_{Hartree-Fock} \quad (1.8)$$

The Pauli exclusion principle requires that the wave function now be anti-symmetric under exchange, such that anytime two electrons are interchanged it results in the wave function changing sign:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i, \mathbf{r}_j, \dots, \mathbf{r}_N) = -\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_j, \mathbf{r}_i, \dots, \mathbf{r}_N) \quad (1.9)$$

In the Hartree-Fock method, the indistinguishability and exchange properties of electrons are very elegantly included within the mathematics by representing the wave function as a Slater-determinant of one-electron wave functions instead of a simple product of wave functions as in the Hartree-method. The determinant quite conveniently results as an anti-symmetric function of all permutations of one-electron wave functions:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \begin{bmatrix} \psi_1(\mathbf{r}_1) & \psi_1(\mathbf{r}_1) & \dots & \psi_1(\mathbf{r}_1) \\ \psi_1(\mathbf{r}_2) & \psi_2(\mathbf{r}_2) & \dots & \psi_N(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{r}_N) & \psi_2(\mathbf{r}_N) & \dots & \psi_N(\mathbf{r}_N) \end{bmatrix} \quad (1.10)$$

The quantum spin variables have been left out for clarity, yet are easily included alongside the position dependences.

As with the Hartree method, minimizing the expectation value of \hat{H} with respect to the one-electron wave functions results in the one-electron, Hartree-Fock equations:

$$\begin{aligned} & -\frac{1}{2} \nabla^2 \psi_i(\mathbf{r}) + V^{ion}(\mathbf{r}) \psi_i(\mathbf{r}) + V^{el}(\mathbf{r}) \psi_i(\mathbf{r}) \\ & - \sum_j \delta_{s_i} \delta_{s_j} \int \frac{1}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \psi_j(\mathbf{r}') \psi_i^*(\mathbf{r}') \psi_j^*(\mathbf{r}') = \varepsilon_i \psi_i(\mathbf{r}) \end{aligned} \quad (1.11)$$

where s_j represents the spin states. The 4th term on the left is new, when compared to the Hartree equations, Equations (1.7) and known as the exchange term. The exchange term is only non-zero when considering like-spins and the result is that like-spin electrons appear to avoid each other. This exchange term adds considerable complexity to the one-electron equations, making the Hartree-Fock equations difficult to solve outside of special cases.

The self-interaction error exactly cancels in Hartree-Fock method

The Hartree-Fock equations have a Hartree potential (classical Coulomb) term that includes the physically incorrect self-interaction in the mean-field approximation. However, in the Hartree-Fock equations, the self-interactions energy is exactly cancelled by the exchange term.

1.4 Density functional theory (DFT)

The section which follows includes theory belonging to the Hohenberg-Kohn theorems and the Kohn-Sham equations. The theory now shown follows the formalism laid out in (Martin, 2004) closely as it was found to be one of the best descriptions in literature.

DFT is currently one of the most successful methods available for calculating the properties of real solids from first principles. By calculating the Hamiltonian by means of mathematical operations, called functionals, performed on the electron densities instead of on the wave functions of individual electrons, DFT allows for great simplification to be made to the process of solving the many body SE. The framework of DFT is often seen as an extended form of the mean-field theory and consists of 2 major parts. In principle, this technique is exact, but the functionals are approximate.

The first part is two theorems developed by Hohenberg and Kohn (Hohenberg & Kohn, 1964a), (Perdew & Wang, 1986) the first of which states that the total energy, E_{tot} of a system in its ground state is, except for a constant, a unique functional of the electron density, $\mathbf{n}(\mathbf{r})$. Furthermore, the functional $E_{tot}|\mathbf{n}(\mathbf{r})|$ is minimized for the ground state electron density $\mathbf{n}_{GS}(\mathbf{r})$. This property allows the ability to calculate electronic properties based on the electron density (3 spatial variables), instead of the $3N$ -variable many-body wave function.

The second part of the theory was developed by (Hohenberg & Kohn, 1964a) and it states that the effective potential V_{eff} is uniquely determined (up to a constant) for a given ground state charge density.

Firstly we write the total energy functional as:

$$E_{Tot}|\mathbf{n}(\mathbf{r})| = T|\mathbf{n}(\mathbf{r})| + E_{Hart}|\mathbf{n}(\mathbf{r})| + E_{xc}|\mathbf{n}(\mathbf{r})| + \int V_{ext}(\mathbf{r})\mathbf{n}(\mathbf{r})d\mathbf{r} \quad (1.12)$$

Where T is the kinetic energy of a non-interacting system, E_{Hart} is the Hartree energy, E_{xc} is the exchange-correlation energy and V_{ext} represents an external potential, including the ions. By minimization of the total energy functional with respect to variations in the electron density, with the constraint that the number of electrons remains fixed, one can show for a one-particle equation that:

$$\left\{-\frac{\hbar^2}{2m}\nabla^2 + V_{eff}|n(\mathbf{r})|\right\}\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}), \quad (1.13)$$

where the effective potential V_{eff} is given by:

$$V_{eff}|n(\mathbf{r})| = V_{ext}|n(\mathbf{r})| + V_{Hart}|n(\mathbf{r})| + V_{xc}|n(\mathbf{r})| \quad (1.14)$$

and V_{Hart} is the Hartree potential,

$$V_{Hart}|n(\mathbf{r})| = -e \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \quad (1.15)$$

and

$$V_{xc}|n(\mathbf{r})| = \frac{\delta E_{xc}|n(\mathbf{r})|}{\delta|n(\mathbf{r})|} \quad (1.16)$$

Equations (1.13) and (1.14) are known as the Kohn-Sham equations. The quantities ψ_i and ε_i are quantities used in the calculation of the electron density and total energy and should not be confused with the wave function and energy of the real electrons.

An exact formulation for the exchange-correlation energy (Perdew & Schmidt, 2001), $E_{xc}|n(\mathbf{r})|$ is given by

$$E_{xc}|n(\mathbf{r})| = \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' n(\mathbf{r}) \frac{n(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \quad (1.17)$$

If we introduce a coupling constant α which varies from 0 to 1, which relates to the real interacting systems and to Kohn-Sham non-interacting systems respectively, then

$$n_{xc}(\mathbf{r},\mathbf{r}') = \int_0^1 d\alpha n_{xc}^\alpha(\mathbf{r},\mathbf{r}') = n_x(\mathbf{r},\mathbf{r}') + n_e(\mathbf{r},\mathbf{r}') \quad (1.18)$$

is the average over the coupling constant α of the density at \mathbf{r}' of the exchange-correlation hole about an electron at \mathbf{r} :

$$n_{xc}^\alpha(\mathbf{r},\mathbf{r}') = \frac{\langle \psi_\alpha | \hat{n}(\mathbf{r}) \hat{n}(\mathbf{r}') | \psi_\alpha \rangle}{n(\mathbf{r})} - \delta(\mathbf{r} - \mathbf{r}') \quad (1.19)$$

and $n_{xc}^{\alpha=0}(\mathbf{r},\mathbf{r}') = n_x(\mathbf{r},\mathbf{r}')$ is the exchange hole. Here, ψ_α is the correlated ground state wave function for a system with equivalent spin densities as compared to the real system, yet with the electron-electron interaction reduced by a factor α .

Due to the Pauli exclusion principle and Coulomb repulsion, an exchange-correlation hole forms leading to a confirmation of the sum rule

$$\int dr' n_{xc}^{\alpha}(\mathbf{r}, \mathbf{r}') = -1 \quad (1.20)$$

Yet the main flaw in DFT is that even though the theory is exact in principle, the form of the correlation potential term is never known. Even for the simplest of systems in computational procedures approximations are made for the exchange-correlation potential. Due to this process, approximations may have to be made in order to arrive at some form of consistent solution.

In order to solve the Kohn-Sham equations one starts by assuming some charge density $n(\mathbf{r})$ and from this calculate $V_{xc}|n(\mathbf{r})|$ from which Equation (1.13) is solved for the wave functions of the system $\psi_i(\mathbf{r})$, from the new wave functions obtained a new charge density can be calculated

$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 \quad (1.21)$$

Thus this process continues until convergence in the charge density is achieved within some predefined region.

The self-interaction error in DFT

As in the Hartree-Fock theory, DFT contains the Hartree potential with the self-interaction error. In the formal DFT theory, an exact exchange-correlation functional potential cancels the self-interaction term, just as in the Hartree-Fock derivation. However, for all practical purposes in DFT one approximates the exchange and correlation terms with an approximate exchange-correlation functional. This approximate functional does not likely cancel the self-interaction error in the Hartree term, which introduces the self-interaction error.

1.4.1 The Hohenberg-Kohn theorems, a formal derivation

The Hohenberg and Kohn theorems were formulated as an approach to describe density functional theory as an exact theory of many-particle systems. It specifically applies to all systems of interacting particles which are placed in an externally applied potential, denoted as $V_{ext}(\mathbf{r})$.

This includes problems related to any system of interacting electrons and nuclei whereby the Hamiltonian of such a system can be written as follows:

$$\hat{H} = -\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 + \sum_i V_{ext}(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (1.22)$$

This section formulates the theorems and proofs first shown by Hohenberg and Kohn in (Hohenberg & Kohn, 1964a) (Hohenberg & Kohn, 1964b).

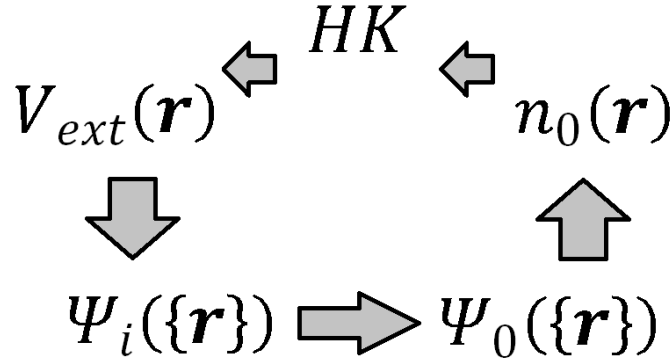


Figure 1.1: A representation of the Hohenberg-Kohn theorem.

The diagram above represents the use of the Hohenberg-Kohn theorem. It can be seen that if the potential $V_{ext}(\mathbf{r})$ is known then all states of the system, $\Psi_i(\{\mathbf{r}\})$ including the ground state, $\Psi_0(\{\mathbf{r}\})$ can be determined. From the ground state of the system the ground state energy density $n_0(\mathbf{r})$ can also be found.

The formal Hohenberg-Kohn theorems and corollaries state:

Theorem I: Consider some external potential $V_{ext}(\mathbf{r})$ which is acting on a system of interacting particles. That potential $V_{ext}(\mathbf{r})$ can be determined from the particle density $n_0(\mathbf{r})$ of the system in its ground state. Furthermore, this potential is unique, except for a constant.

Corollary I of Theorem I: From Theorem I, it follows that the Hamiltonian of the interacting system may be calculated from the particle density $n_0(\mathbf{r})$ alone, except for the value of the unknown constant shift to the total energy. From this fact, the many-body wave functions for both the ground state and all excited states of the system can be fully determined.

Thus in conclusion it is shown that all properties of an interacting system of particles can be uniquely determined as long as only the ground state particle density $n_0(\mathbf{r})$ is known.

Theorem II: It is possible to define a functional in terms of the particle density $n(\mathbf{r})$ of a system of interacting particles for the energy of that system given as $E[n]$. This functional will be applicable for all external potentials $V_{ext}(\mathbf{r})$. The exact ground state energy $n_0(\mathbf{r})$ is the global minimum value of this functional, for that particular $V_{ext}(\mathbf{r})$ over all particle densities $n(\mathbf{r})$. The ground state particle density $n_0(\mathbf{r})$ is the particle density which is found to minimize the functional.

Corollary II of theorem II: All that is required to determine the precise ground state and particle density is the functional $E[n]$. It is only valid for the ground state and all excited states of the particles in question must be determined by some other process.

The proofs of the two Hohenberg-Kohn theorems are now shown:

Proof of Theorem 1: density as a basic variable:

Suppose that there are two different external potentials $V_{ext}^{(1)}(\mathbf{r})$ and $V_{ext}^{(2)}(\mathbf{r})$ which are different by more than some constant and which will result in two different Hamiltonians, $\hat{H}^{(1)}$ and $\hat{H}^{(2)}$, which both have different ground state wave functions, $\psi^{(1)}$ and $\psi^{(2)}$, which we hypothesize to have an equivalent ground state density $n_0(\mathbf{r})$. Since $\psi^{(2)}$ is not the ground state of $\hat{H}^{(1)}$ it follows that:

$$E^{(1)} = \langle \psi^{(1)} | \hat{H}^{(1)} | \psi^{(1)} \rangle < \langle \psi^{(2)} | \hat{H}^{(1)} | \psi^{(2)} \rangle \quad (1.23)$$

The strict inequality follows if the ground state is non-degenerate.

Since:

$$\langle \psi^{(2)} | \hat{H}^{(1)} | \psi^{(2)} \rangle = \langle \psi^{(2)} | \hat{H}^{(2)} | \psi^{(2)} \rangle + \langle \psi^{(2)} | \hat{H}^{(1)} - \hat{H}^{(2)} | \psi^{(2)} \rangle \quad (1.24)$$

$$= E^{(2)} + \int d^3r [V_{ext}^{(1)}(\mathbf{r}) - V_{ext}^{(2)}(\mathbf{r})] n_0(\mathbf{r}) \quad (1.25)$$

$$E^{(1)} < E^{(2)} + \int d^3r [V_{ext}^{(1)}(\mathbf{r}) - V_{ext}^{(2)}(\mathbf{r})] n_0(\mathbf{r}) \quad (1.26)$$

And if we consider $E^{(2)}$ in exactly the same way, we find the same situation arises

$$\begin{aligned}
\langle \psi^{(1)} | \hat{H}^{(2)} | \psi^{(1)} \rangle &= \langle \psi^{(1)} | \hat{H}^{(1)} | \psi^{(1)} \rangle + \langle \psi^{(1)} | \hat{H}^{(2)} - \hat{H}^{(1)} | \psi^{(1)} \rangle \\
&= E^{(1)} + \int d^3 r [V_{ext}^{(2)}(\mathbf{r}) - V_{ext}^{(1)}(\mathbf{r})] n_0(\mathbf{r}) \\
E^{(2)} &< E^{(1)} + \int d^3 r [V_{ext}^{(2)}(\mathbf{r}) - V_{ext}^{(1)}(\mathbf{r})] n_0(\mathbf{r})
\end{aligned} \tag{1.27}$$

Now adding the two results together we arrive at the contradiction, that is:

$E^{(1)} + E^{(2)} < E^{(1)} + E^{(2)}$ and from this we arrive at the desired result: there cannot be two different external potentials, which are identical except for the value of a constant that will yield the same non-degenerate ground state charge density. The particle density therefore uniquely determines the external potential, with the exception of some constant value.

The corollary of theorem II follows from the above proof by implication that if the Hamiltonian is determined, with the exception of some constant, by the ground state particle density, then by solving the Schrödinger equation by making use of this Hamiltonian, one can determine the wave function of any required state. Within the set of all solutions which are applicable to the particle density used, the solution with the lowest energy is simply the unique ground state of the system.

Even though this result is indeed profound, it holds no indication as to how to determine the external potential as all that was shown was that $n_0(\mathbf{r})$ uniquely determines $V_{ext}(\mathbf{r})$.

Proof of Theorem 2: Existence of a universal functional for the energy $E[n]$ in terms of the density $n(\mathbf{r})$

Theorem II is specifically targeted at particle densities, $n(\mathbf{r})$, which are acted upon by an external potential, $V_{ext}(\mathbf{r})$ and forms part of the electron Hamiltonian defined set of ground state densities. This set of particle densities forms a basis within which the functionals of the particle density can be written. From theorem I we know that if the particle density $n(\mathbf{r})$ is known, all other properties of the system can be determined and thus the total energy functional can be described in terms of these properties.

$$\begin{aligned}
E_{HK}[n] &= T[n] + E_{int}[n] + \int d^3 r V_{ext}(\mathbf{r})n(\mathbf{r}) + E_{II} \\
&\equiv F_{HK}[n] + \int d^3 r V_{ext}(\mathbf{r})n(\mathbf{r}) + E_{II}
\end{aligned} \tag{1.28}$$

where E_{II} is the interaction energy between the ionized nuclei of the system. $F_{HK}[n]$ is the functional created to encompass all internal energies of the system, namely the kinetic and potential energies; this functional is by construction applicable in all respective cases:

$$F_{HK}[n] = T[n] + E_{int}[n]. \quad (1.29)$$

Assuming there exists some system with a ground state particle density of $n_0^{(1)}(\mathbf{r})$, which also has naturally an attributed externally acting potential $V_{ext}^{(1)}(\mathbf{r})$. The resulting expectation value of the Hamiltonian which describes this system, in its ground state with an associated wave function denoted by $\psi^{(1)}$, is equal to the Hohenberg-Kohn functional.

$$E^{(1)} = E_{HK}[n^{(1)}] = \langle \psi^{(1)} | \hat{H}^{(1)} | \psi^{(1)} \rangle \quad (1.30)$$

Now assume there is some second particle density, $n^{(2)}(\mathbf{r})$ which by construction has its own attributed wave function, $\psi^{(2)}$. The total energy of the second state must therefore be greater than that of the first state since:

$$E^{(1)} = \langle \psi^{(1)} | \hat{H}^{(1)} | \psi^{(1)} \rangle < \langle \psi^{(2)} | \hat{H}^{(1)} | \psi^{(2)} \rangle = E^{(2)} \quad (1.31)$$

The resulting energy found by the use of Equation (1.28) which was calculated for the ground state particle density $n_0(\mathbf{r})$ is found to be lower than that of the value of Equation (1.31) for any other particle density $n(\mathbf{r})$ as expected.

Thus the implication of this is that if the functional $F_{HK}[n]$ was known, the exact ground state particle density and energy can be found by minimizing the total energy of the system with respect to the density function $n(\mathbf{r})$.

1.4.2 The Kohn-Sham auxiliary system

DFT is the most widely used method today for electronic structure calculations due to the approach of Kohn and Sham in 1965. The idea is to replace the original many body problem by an auxiliary independent-particle problem. This principle in effect leads to exact calculations of properties of many body systems, using the independent particle methods. In practice it has made possible approximate formulations that have proved to be remarkably successful.

As a self-consistent method, the Kohn-Sham approach, involves independent particles but an interacting charge density. The idea that follows is the process and formulation of the Kohn-Sham (KS) approach and the exchange-correlation functional $E_{xc}[\mathbf{n}]$.

The approach of the Kohn Sham method is to assume that the ground state density of the original interacting system is equal to that of some chosen non-interacting system. This results in a set of independent particle equations for a non-interacting system and all of the many-body terms incorporated into an exchange-correlation functional of the density. By solving the equations one finds the ground state density and energy of the original interacting system with the accuracy limited only by the approximation in the exchange-correlation functional.

The K-S construction of the auxiliary system is firmly based upon 2 assumptions:

1. Assuming we have some secondary system of non-interacting particles in their own ground state, the exact ground state of our system in question can be calculated in terms of the secondary system.
2. The secondary non-interacting system has an attributed Hamiltonian which is chosen to have the standard kinetic energy term and a potential $V_{eff}^\sigma(\mathbf{r})$. This potential which acts upon the electron is dependent of the spin orientation σ of that electron and it is dependent on the distance r from the particle.

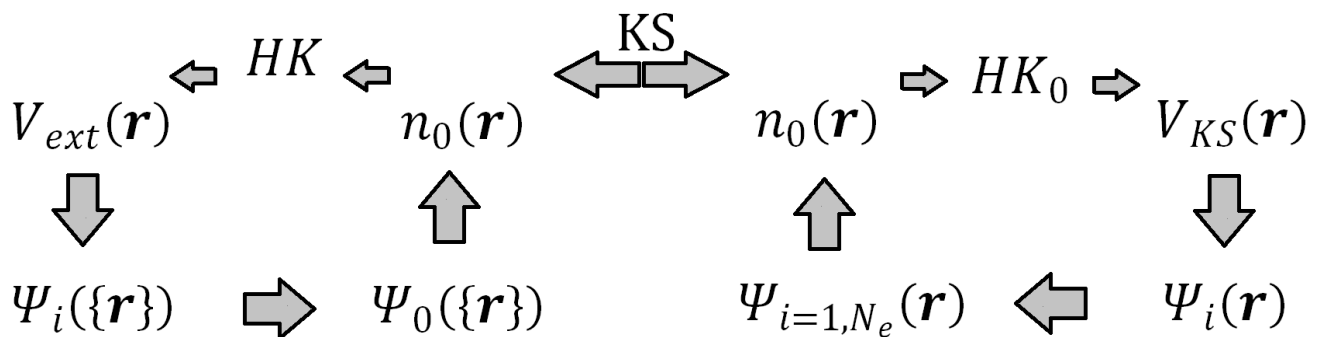


Figure 1.2: A representation of the Kohn-Sham theorem. Showing how a solution to the independent-particle Kohn-Sham system determines all the properties of the fully interacting system.

The calculations for the Kohn-Sham auxiliary system are applied to the independent non-interacting particle system defined by that systems own Hamiltonian:

$$\hat{H}_{aux}^{\sigma} = -\frac{1}{2}\nabla^2 + V^{\sigma}(\mathbf{r}) \quad (1.32)$$

The potential, $V^{\sigma}(\mathbf{r})$ so far is not specified and the expressions must therefore apply for all $V^{\sigma}(\mathbf{r})$ in some range, in order to define functionals for a range of densities.

For a system containing $N = N^{\uparrow} + N^{\downarrow}$ independent electrons, where N^{\uparrow} is the number of electrons in the spin up configuration and N^{\downarrow} the number of electrons in the spin down configuration, obeying this Hamiltonian, the ground state of this system has one electron in each of the N^{σ} orbitals $\psi_i^{\sigma}(\mathbf{r})$ with the lowest eigenvalues ϵ_i^{σ} of the Hamiltonian Equation (1.22). The electron density of the auxiliary system can now be represented as a probability density of the orbitals

$$n(\mathbf{r}) = \sum_{\sigma} n(\mathbf{r}, \sigma) = \sum_{\sigma} \sum_{i=1}^{N^{\sigma}} |\psi_i^{\sigma}(\mathbf{r})|^2 = 2 \sum_{i=1}^{N^{\sigma}} |\psi_i^{\sigma}(\mathbf{r})|^2 \quad (1.33)$$

The independent non-interacting particle kinetic energy T_s is given by:

$$T_s = -\frac{1}{2} \sum_{\sigma} \sum_{i=1}^{N^{\sigma}} \langle \psi_i^{\sigma} | \nabla^2 | \psi_i^{\sigma} \rangle = -\frac{1}{2} \sum_{\sigma} \sum_{i=1}^{N^{\sigma}} \int d^3r |\nabla \psi_i^{\sigma}(\mathbf{r})|^2. \quad (1.34)$$

The classical Coulomb interaction energy of the electron density, interacting with itself is defined as:

$$E_{Hartree}[n] = \frac{1}{2} \int d^3r d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \quad (1.35)$$

The Kohn-Sham approach to the full interacting many body problem is to re-write the Hohenberg-Kohn expression for the ground state energy functional Equation (1.28) in the form:

$$E_{KS} = T_s[n] + \int d\mathbf{r} V_{ext}(\mathbf{r})n(\mathbf{r}) + E_{Hartree}[n] + E_{II} + E_{xc}[n] \quad (1.36)$$

Here the external potential $V_{ext}(\mathbf{r})$ is due to the nuclei and all other external fields (assuming no dependence on spin) and E_{II} is the interaction between the nuclei. The sum of the terms involving, E_{II} , V_{ext} and $E_{Hartree}$ forms the neutral grouping that is well defined. The independent particle energy T_s is given explicitly as a function of the orbitals. Yet by application of the Hohenberg-Kohn theory, T_s for each spin σ must be a unique functional of the density $n(\mathbf{r}, \sigma)$.

All many-body effects of exchange and correlation are grouped into E_{xc} and now from Equation (1.36) the Kohn-Sham expressions for the total energy shows that E_{xc} can be written in terms of the Hohenberg-Kohn functional equation (1.29) as:

$$E_{xc}[\mathbf{n}] = F_{HK}[\mathbf{n}] - (T_s[\mathbf{n}] + E_{Hartree}[\mathbf{n}]) \quad [1.37]$$

Or alternatively:

$$E_{xc}[\mathbf{n}] = \langle \hat{T} \rangle - T_s[\mathbf{n}] + \langle \hat{V}_{int} \rangle - E_{Hartree}[\mathbf{n}] \quad (1.38)$$

Here $[\mathbf{n}]$ denotes a functional of the density $\mathbf{n}(\mathbf{r}, \sigma)$ which is dependent on both position and spin. From this we can see $E_{xc}[\mathbf{n}]$ must be a functional since the right-hand sides of the equations are functionals. The latter equation shows explicitly that the difference of the kinetic and the internal interaction energies of the true interacting many-body system, from those of the alternative independent-particle system with electron-electron interactions are replaced by the Hartree energy. If the functional $E_{xc}[\mathbf{n}]$ defined in Equation (1.38) were to be known then the exact ground state energy and density of the many-body electron problem could be found by solving the Kohn-Sham equations for independent particles.

Thus to the extent that as an approximate form for the energy $E_{xc}[\mathbf{n}]$ describes the true exchange-correlation energy, the Kohn-Sham method provides a fairly accurate approach to calculating the ground state properties of the many body system.

1.4.3 The Kohn-Sham variational equations

The solution to the Kohn-Sham auxiliary system in its ground state can be seen as the minimization of either the particle density $\mathbf{n}(\mathbf{r}, \sigma)$ of the system or the minimization of the systems' effective potential $V_{eff}^\sigma(\mathbf{r})$. Since the independent particle kinetic energy, T_s , is explicitly expressed as a functional of the orbitals and all other terms present are considered to be functionals of the density, the wave functions can be allowed to vary and by applying of the chain rule applied to equation (1.36), one can derive the variational equation:

$$\frac{\delta E_{KS}}{\delta \psi_i^{\sigma*}(\mathbf{r})} = \frac{\delta T_s}{\delta \psi_i^{\sigma*}(\mathbf{r})} + \left[\frac{\delta E_{ext}}{\delta n(\mathbf{r}, \sigma)} + \frac{\delta E_{Hartree}}{\delta n(\mathbf{r}, \sigma)} + \frac{\delta E_{xc}}{\delta n(\mathbf{r}, \sigma)} \right] \frac{\delta n(\mathbf{r}, \sigma)}{\delta \psi_i^{\sigma*}(\mathbf{r})} = 0 \quad (1.39)$$

Subject to the orthonormalization constraints:

$$\langle \delta\psi_i^\sigma | \delta\psi_j^{\sigma'} \rangle = \delta_{ij}\delta_{\sigma,\sigma'} \quad (1.40)$$

Now in order to proceed further, the $V_{KS}(\mathbf{r})$ potential must be obtained for a given density $n(\mathbf{r})$. To do this, the variation of the energy functional $E[n(\mathbf{r})]$ must vanish with respect to the variation of the one electron orbitals $\psi_i(\mathbf{r})$, which is constrained by the orthonormality constraint Equation (1.40). Thus let's consider the constrained functional,

$$E'_{KS} = E_{KS} - \sum_{ij} \lambda_{ij} \left(\int \psi_i^*(\mathbf{r})\psi_j(\mathbf{r})d^3\mathbf{r} - \delta_{ij} \right) \quad (1.41)$$

Where λ_{ij} are Lagrangian multipliers, its functional derivatives must vanish, thus:

$$\frac{\delta E'_{KS}}{\delta \psi_i^*(\mathbf{r})} = \frac{\delta E'_{KS}}{\delta \psi_i(\mathbf{r})} = 0 \quad (1.42)$$

Now using the definition of a functional derivative and expression (1.33) and (1.34) for $n(\mathbf{r}, \sigma)$ and T_s which gives:

$$\frac{\delta T_s}{\delta \psi_i^{\sigma*}(\mathbf{r})} = -\frac{1}{2}\nabla^2\psi_i^\sigma(\mathbf{r}); \quad \frac{\delta n^\sigma(\mathbf{r}, \sigma)}{\delta \psi_i^{\sigma*}(\mathbf{r})} = \psi_i^\sigma(\mathbf{r}) \quad (1.43)$$

And expression (1.35) with respect to $\psi_i^{\sigma*}(\mathbf{r})$ yields the result:

$$\frac{\delta E_{Hartree}[n]}{\delta \psi_i^{\sigma*}(\mathbf{r})} = \frac{1}{2} \int d^3\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \psi_i^\sigma(\mathbf{r}) \quad (1.44)$$

Then finally using (1.41-42) and (1.43-1.44) applied into equation (1.39) yield the equation:

$$\left(-\frac{1}{2}\nabla^2 + V_{Hartree}(\mathbf{r}, \sigma) + V_{xc}[n(\mathbf{r}, \sigma)] + V_{ext}(\mathbf{r}, \sigma) \right) \psi_i^\sigma(\mathbf{r}) = \sum_j \lambda_{ij} \psi_j^\sigma(\mathbf{r}) \quad (1.45)$$

Where the Hartree and exchange-correlation potentials are defined as:

$$V_{Hartree}(\mathbf{r}) = \int d^3\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} ; \quad V_{xc}[n(\mathbf{r}, \sigma)] = \frac{\delta E_{xc}}{\delta n(\mathbf{r}, \sigma)} \quad (1.46)$$

Now multiplying both sides of equation (1.45) by $\psi_k^\sigma(\mathbf{r})$ and integrating, one can obtain the Lagrangian multipliers. This leads to the Kohn-Sham equations:

$$(H_{KS}^\sigma - \varepsilon_i^\sigma) \psi_i^\sigma(\mathbf{r}) = 0 \quad (1.47)$$

Where the $\delta_{ij}\epsilon_i^\sigma = \lambda_{ij}$ are the eigenvalues and H_{KS}^σ is the effective Hamiltonian (in Hartree atomic units)

$$H_{KS}^\sigma = \left(-\frac{1}{2}\nabla^2 + \frac{\delta E_{Hartree}}{\delta n(\mathbf{r}, \sigma)} + \frac{\delta E_{xc}}{\delta n(\mathbf{r}, \sigma)} + V_{ext}(\mathbf{r}, \sigma) \right)$$

$$H_{KS}^\sigma = \left(-\frac{1}{2}\nabla^2 + V_{Hartree}(\mathbf{r}, \sigma) + V_{xc}[n(\mathbf{r}, \sigma)] + V_{ext}(\mathbf{r}, \sigma) \right) = -\frac{1}{2}\nabla^2 + V_{KS}^\sigma \quad (1.48)$$

The Kohn-Sham eigenvalues have no physical meaning except for the highest occupied eigenvalue which is the negative of the ionization energy. The Kohn-Sham Hamiltonian is related to the functional derivative of the energy by the expression:

$$\frac{\delta E_{KS}}{\delta \psi_i^\sigma(\mathbf{r})} = H_{KS}^\sigma \psi_i^\sigma(\mathbf{r}) \quad (1.49)$$

1.4.4 Solving the Kohn-Sham equation: The self-consistent Kohn-Sham equations

The Kohn-Sham equations can be seen as a set of independent-particle equations which must be solved subject to the condition that the effective potential $V_{eff}^\sigma(\mathbf{r})$ and the density $n(\mathbf{r}, \sigma)$ are consistent. From now on the explicit reference to spin shall be dropped and the notation V_{eff} and n will be assumed to designate both spatial and spin dependencies. The actual calculation procedure utilizes a numerical procedure which successively changes V_{eff} and n to approach the self-consistent solution.

Figure 1.3 outlines the steps to solve the KS equations. The computationally intensive process begins at “Solve KS equations” for a given potential V_{eff} . In this step for some V^{in} the KS equations are solved uniquely for exactly one n^{out} . Conversely, for a given form of the exchange functional, any density n determines a unique potential V_{eff} as can be seen in the second box. The problem lies however that, unless for the exact solution, the input and output densities will not agree, thus to arrive at the solution we define a new operational potential $n^{out} \rightarrow V^{new}$ and so the cycle begins again and repeats iteratively with each new choice of operational potential converging until self-consistency within the results is obtained.

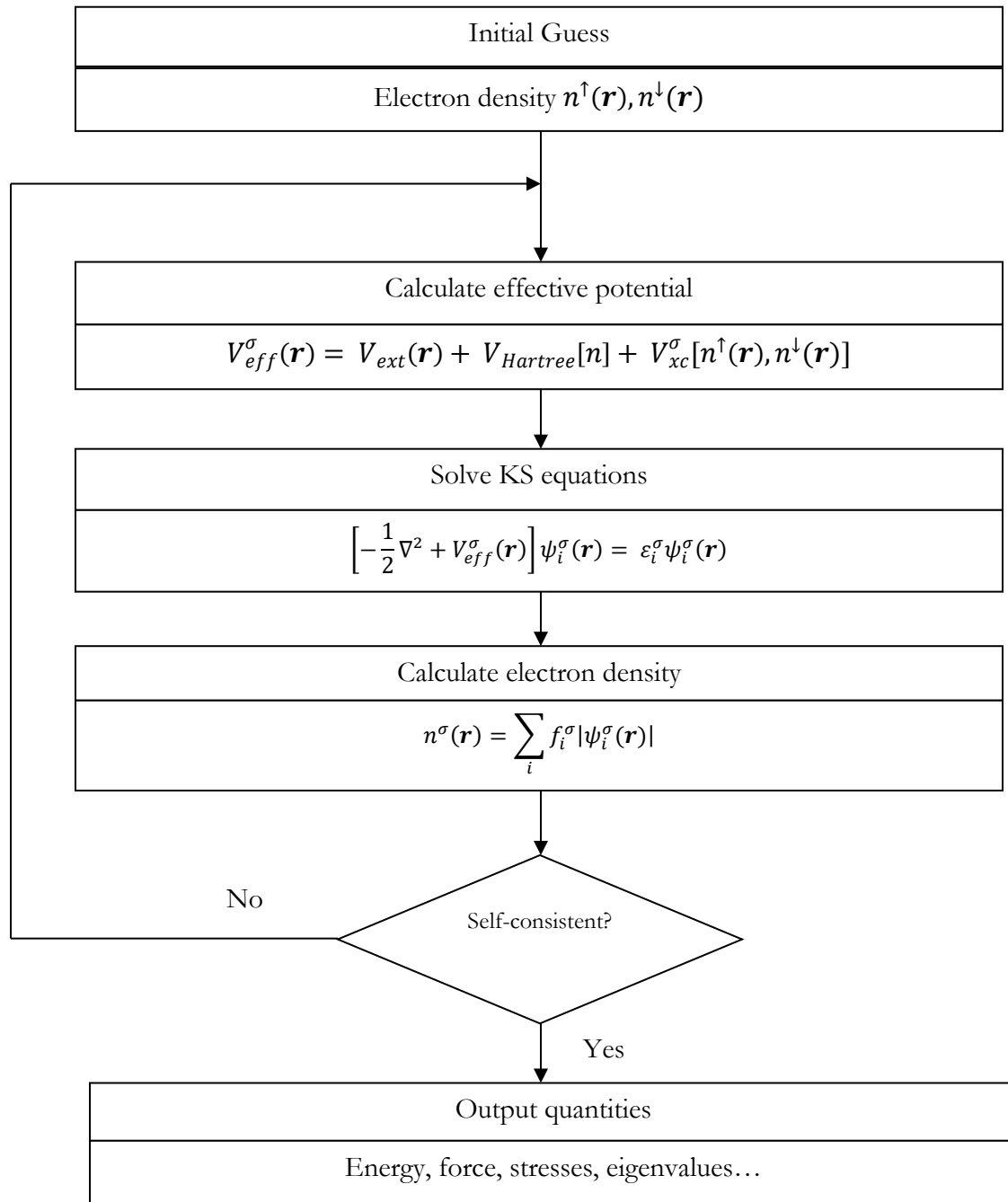


Figure 1.3: Schematic representation of the self-consistent loop for the solution of the Kohn-Sham equations.

In order to solve the KS equations we must define a basis set. The Kohn-Sham equations are expressed in terms of reciprocal lattice vectors \mathbf{G} and a Bloch vector \mathbf{k} in the Brillouin Zone. The basis set is defined as follows:

$$\langle \mathbf{r} | \mathbf{k} + \mathbf{G} \rangle = \frac{1}{\sqrt{V}} e^{i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r}}$$

$$\frac{i}{2} |\mathbf{k} + \mathbf{G}|^2 \leq E_{cut} \quad (1.50)$$

Where V is the total volume of the system and E_{cut} is the kinetic energy cut-off of the plane-waves. These plane-waves do not accurately produce localized functions, this is problematic when considering the core electrons in the calculations, as the high energies required for producing accurate results do not exist in the core electrons due to their lower kinetic energies and thus these higher energies that are needed are only found in the valence shell electrons. Yet since most applications of DFT focus on small energy variations due to changes in the distribution of the valence electrons, the core of each atom can be replaced by a pseudo-atom representation which closely mimics the tightly bound core electrons.

In order to solve the problem of the rapidly varying charge density about each core of the system in question, the pseudopotentials are used which also contain the core electrons. In which case the problem is re-expressed as a smooth function and an auxiliary localized function. This method keeps the full electron wave function but all integrals are solved as a combination of integrals of smooth functions which extend throughout space and localized contributions.

If the projection operators are evaluated in reciprocal space, the number of calculations performed increases with the size of the basis set of plane-waves. If however it is done in real space, the projection operators are confined to a sphere around each atom and the number of calculations performed does not increase with the size of the system. (Martin, 2004)

The Charge density $n(\mathbf{r}, \sigma)$ in a periodic system is found by a summation over an infinite number of these \mathbf{k} -points:

$$n(\mathbf{r}, \sigma) = \sum_{\mathbf{k}} \sum_i |\psi_{\mathbf{k},i}^{\sigma}(\mathbf{r}, \sigma)|^2 \quad (1.56)$$

where the i index is for the occupied bands. In a system where the number of unit cells tends to infinity, the discrete sum over \mathbf{k} becomes an integral. However this integral can be approximated

well by the choice of a discrete sum over a small number of \mathbf{k} -points for semiconductors and insulators

The most efficient way to determine the charge density is to determine the Bloch states on the \mathbf{k} -points in real space. From this the charge density can be solved from the square of the wave function. To solve this, a method of Fast Fourier transform (FFT) is used to transform from reciprocal space to real space.

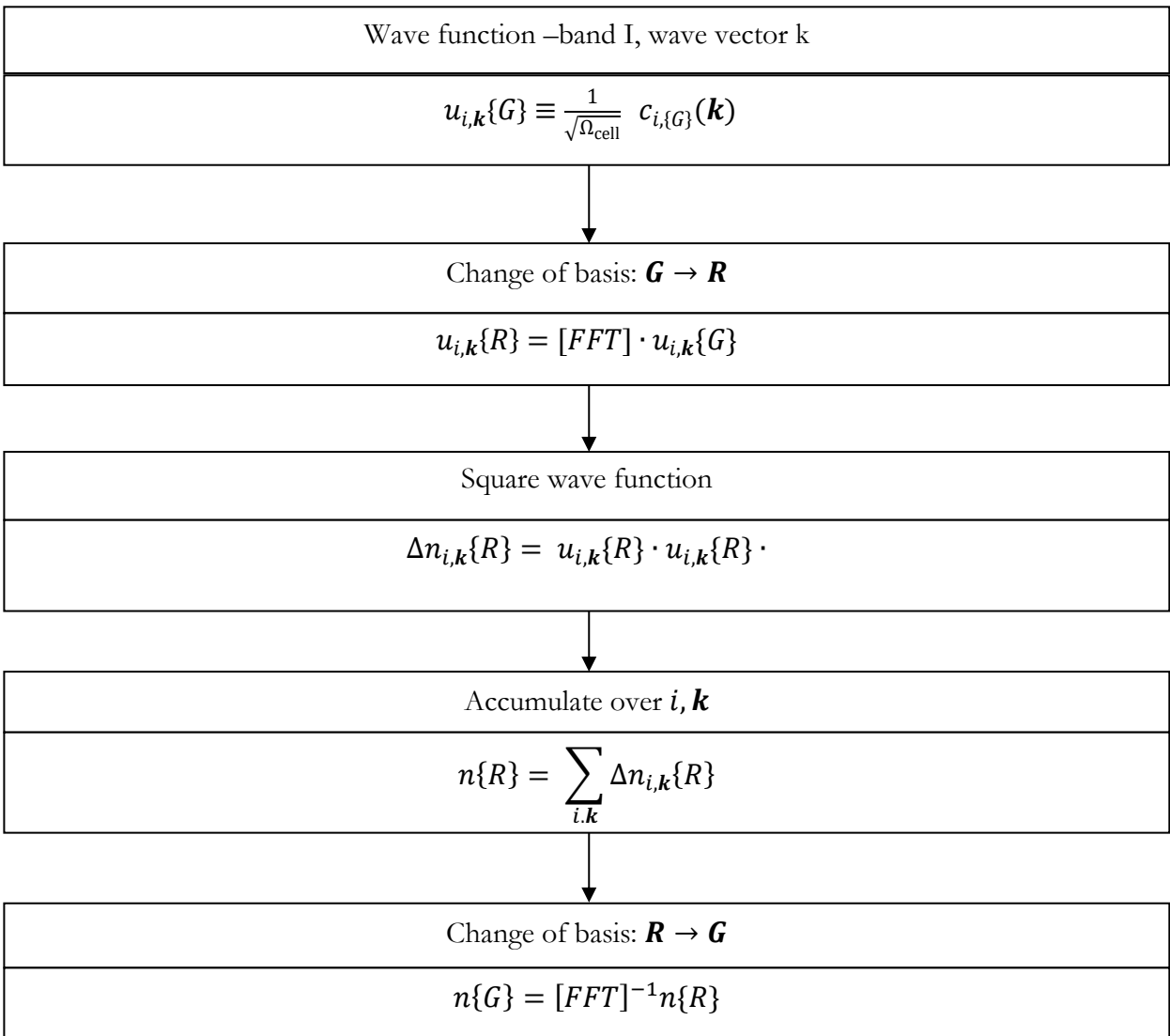


Figure 1.4: Calculation of the density using Fourier transforms and grids. The notation $\{\mathbf{G}\}$ and $\{\mathbf{R}\}$ denotes the sets of N \mathbf{G} vectors and N grid points \mathbf{R} (Martin, 2004).

1.5 DFT functionals

Density Functional theory (DFT) is currently the standard model for computing material properties and has been for many past studies in the field of solid state physics. As described in the previous section DFT is an exact theory, which states that ground-state properties of a material can be obtained from the functionals of the charged density alone. However in computational practice, exchange-correlation functionals describing the many-body electron interactions must be approximated. Common choices for these are the local density approximation (LDA) and the generalized gradient approximations (GGA) and most recently hybrid functionals. All of which cannot accurately describe material properties of interest. This is what leads to the shortcomings of DFT.

DFT functionals produce a variety of non-conforming or non-uniform results for reasons often left unclear or due to very subtle physics related factors. DFT is often thought to only have problems with highly correlated materials exhibiting complex electronic structures. However DFT has been largely infamous for its failure to compute band gaps accurately for any material. (Yakovin & Dowben, 2007)

The crucial quantity in the Kohn-Sham approach is the exchange-correlation energy which is expressed as a functional of the density $E_{xc}[n]$. Several choices for this exchange-correlation energy are used, ranging from energies based on rigorous physical principles, to deriving the energy from empirical data.

Density functional approximations for the exchange correlation energy, E_{xc} are described below for the exchange-correlation energy as a function of the electron density:

$$E_{xc}[n_{\uparrow}, n_{\downarrow}] = \int d\mathbf{r} \varepsilon_{xc}([n_{\uparrow}, n_{\downarrow}]; \mathbf{r}) \quad (1.51)$$

where the integrand $n\varepsilon_{xc}$ is an exchange-correlation energy density and ε_{xc} is the exchange correlation energy per electron.

1.5.1 The LSDA approximation

Kohn and Sham pointed out that solids can often be considered as close to the limit of the homogeneous electron gas. In that limit, it is known that the effects of exchange and correlation are local in character and from this they proposed making the Local Density Approximation (LDA), in which the exchange-correlation energy is simply an integral over all space with the exchange-correlation energy density at each point assumed to be the same as in a homogeneous electron gas with that density

$$\begin{aligned}
 E_{xc}^{LDA}[n^\uparrow, n^\downarrow] &= \int d^3r n(\mathbf{r}) \epsilon_{xc}^{hom}(n^\uparrow(\mathbf{r}), n^\downarrow(\mathbf{r})) \\
 &= \int d^3r n(\mathbf{r}) [\epsilon_x^{hom}(n^\uparrow(\mathbf{r}), n(\mathbf{r})^\downarrow) + \epsilon_c^{hom}(n^\uparrow(\mathbf{r}), n(\mathbf{r})^\downarrow)]
 \end{aligned}
 \tag{1.52}$$

The LSDA can be formulated in terms of either two spin densities $n^\uparrow(\mathbf{r})$ and $n^\downarrow(\mathbf{r})$, or the total density $n(\mathbf{r})$ and the fractional spin polarization $\zeta(\mathbf{r})$ defined as:

$$\zeta(\mathbf{r}) = \frac{n^\uparrow(\mathbf{r}) - n^\downarrow(\mathbf{r})}{n(\mathbf{r})}
 \tag{1.53}$$

For un-polarized systems the LDA is found simply by setting $n^\downarrow(\mathbf{r}) = n^\uparrow(\mathbf{r}) = n(\mathbf{r})/2$

Now application of the definition of the functional to Equation (1.51):

$$\frac{\delta E_{xc}}{\delta n(\mathbf{r})} \equiv \mu_{xc}(n^\uparrow(\mathbf{r}), n^\downarrow(\mathbf{r})) = \left(\epsilon_{xc}^{LDA}(n) - n \frac{d\epsilon_{xc}(n)}{dn} \right)_{n=n(\mathbf{r})}.
 \tag{1.54}$$

The $\epsilon_{xc}^{LDA}(n(\mathbf{r}))$ is chosen to be the exchange-correlation energy per electron of a homogeneous electron gas. The exchange part is given by Equation (1.52) where the density of the homogeneous electron gas is known analytically (Dirac, 1930),

$$\epsilon_x^{LDA}(n(\mathbf{r})) = -\left(\frac{3}{4}\right) \left(\frac{6}{\pi} n(\mathbf{r})\right)^{\frac{1}{3}}
 \tag{1.55}$$

The correlation part has various ways of being found, yet most commonly is a highly accurate QMC fit of the electron gas and is the preferred method as analytic expressions for the correlation energy are not known except in the high- and low- density limits (Parr & Yang, 1994).

The LDA approach has proven to be largely successful, however it overestimates bond strengths and cohesive energies, on the order of 20% and this has a follow on effect of underestimations of

the bond lengths as compared to experimental results. To correct this, other methods have been introduced such as the GGA group of functionals. In the GGA approach the exchange-correlation functional is written in terms of both, the local charge density and the local gradient of the charge density. This inclusion of the local gradient term greatly improved the accuracy of the results obtained.

1.5.2 Generalized gradient approximation (GGA)

The generalized gradient approximation (GGA):

$$E_{xc}^{GGA}[n_{\uparrow}, n_{\downarrow}] = \int d\mathbf{r} n(\mathbf{r}) \varepsilon_{xc}^{LDA}([n_{\uparrow}, n_{\downarrow}, \nabla n_{\uparrow}, \nabla n_{\downarrow}]; \mathbf{r}) \quad (1.56)$$

Initially was seen as a huge improvement over LDA for atomization energies and became a standard method in many fields. The leading gradient correction for exchange and correlation is second order, for slowly varying densities. Initially the Perdew-Wang exchange and correlation functional (PW91) (Perdew & Wang, 1986) used analytic expansions to second order yet it was found to violate exact properties of the exchange-correlation holes. This violation forced the development of a numerically parameterized GGA functional to satisfy exact hole constraints, called Perdew-Burke-Erzenhoff PBE (Perdew & Enzerhof, 1996). GGA is a semi-local functional of density since it requires the density in an infinitesimal neighborhood around \mathbf{r} .

In 2007 Perdew et al. (Perdew, 2008) released an expanded version of the GGA-PBE functional which was intended to rectify the overestimation issues in the PBE functional. Accurate results with respect to atomic exchange energies however required violating the gradient expansion for slowly varying densities. The violation is an important point to make for solids. Thus the PBEsol model was introduced which restored the gradient expansion for exchange but at the same time violated the gradient expansion for correlation, by fitting a parameter in the correlation functional for the jellium exchange-correlation (XC) surface energy. This second violation was corrected in later work by Zhao (Zhao & Truhlar, 2008) whereby a complete restoration of the gradient expansion for the exchange and correlation was derived.

1.5.3 Hybrid functionals

A Hybrid functional is a functional which explicitly and implicitly depends upon the electron density. A hybrid functional is a linear combination of an orbital dependent Hartree-Fock exchange

potential and any number of explicit density dependent exchange functionals. The first hybrid functional was developed by Becke (Becke, 1988), the “Becke, three-parameter, Lee-Yang-Parr” (B3LYP) (Becke, 1988) (Zhang & Yang, 2000) and was of the form:

$$E_{XC}^{Becke} = E_{XC}^{LDA} + \alpha_0(E_X^{exact} - E_X^{LDA}) + a_X \Delta_X^{B88} + a_C \Delta_C^{PW91} \quad (1.56)$$

where α_0 , a_X , a_C are semi-empirical coefficients that are determined by and approximated for the experimental data, E_X^{exact} is the exact exchange energy, Δ_X^{B88} is the Becke 1988 gradient correction to LDA for the exchange term (Perdew, et al., 1982) and Δ_C^{PW91} is the Perdew and Wang gradient correction for correlation (Stowasser & Hoffman, 1999). The main conditions in choosing how to implement hybrid functionals are the selection of coefficients. This is mainly because the coefficients that determine the mixing ratio cannot be obtained by ab-initio techniques.

Another hybrid functional currently in use is known as the screened hybrid of Heyd-Scuseria-Ernserhof (HSE) functional (Heyd, et al., 2003). This functional uses a screened, short-range HF exact exchange instead of the full exact exchange. The screened terms in this functional are found from the splitting of the Coulomb operator into short and long-range terms as shown below.

$$\frac{1}{r} = \frac{erfc(\omega r)}{r} + \frac{erf(\omega r)}{r} \quad (1.57)$$

where the first and second terms are the short and long range terms respectively. The complimentary error functions $erfc(\omega r) = 1 - erf(\omega r)$ and ω determines the range. The long-range term is zero and the short range term is equivalent to the full Coulomb operator when $\omega = 0$. For $\omega \rightarrow \infty$ the opposite is true. The HSE exchange-correlation energy is given by,

$$E_{XC}^{HSE} = aE_X^{HF,SR}(\omega) + (1 - a)E_X^{\omega PBE,SR}(\omega) + E_X^{\omega PBE,LR}(\omega) + E_C^{PBE} \quad (1.58)$$

where $E_X^{HF,SR}$ is the hybrid functional short range exchange term, $E_X^{\omega PBE,LR}$ and $E_X^{\omega PBE,SR}$ are the long range and short range components of the PBE exchange functional, ω is the splitting parameter and $a = 0.25$ is the mixing coefficient. The form of this functional can be seen as an adiabatic connection functional only for the short-range portion of exchange; the long-range exchange and correlation are treated at the PBE generalized-gradient approximation GGA level (Savin, et al., 1998)

Using sufficient terms and optimizing the fitting parameters, hybrid functionals are capable of obtaining accurate results for a large variety of systems. However due to the empirical nature and lack of physical insight, these functionals are not transferable and thus become application specific.

1.6 Basis sets

The wave function required to find the solution of the many-body Schrodinger equation has to be expanded in terms of a set of known basis functions.

Accordingly a single electron wave function can be written as,

$$\psi_i(\mathbf{r}) = \sum_{j=1}^{\infty} c_j \varphi_j(\mathbf{r}) \quad (1.59)$$

where φ_j belongs to a complete set of functions. In practice solving for an infinite number of basis functions is impractical thus only a finite number of functions are used. The basis functions have the same characteristics as the real wave function.

1.6.1 The plane-wave basis set

In any periodic system, the potential of that system follows the following relationship

$$V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}) \quad (1.60)$$

where \mathbf{R} is some arbitrary lattice vector. The ions in a perfect crystal are arranged in a regular periodic way at $0K$, from this the external potential felt by the electron will also be periodic, where this periodicity will have the same length as a unit cell, 1, thus the external potential experienced by an electron at \mathbf{r} can be expressed as

$$V(\mathbf{r}) = V(\mathbf{r} + 1) \quad (1.61)$$

From this we can make use of Bloch's theorem (Patterson & Bailey, 2010). Bloch's Theorem makes it possible to express the wave function of the crystal in terms of wave functions at reciprocal space vectors of a Bravais Lattice. This method reduces the infinite number of one-electron wave functions to be calculated to simply the number of electrons in the unit cell of the crystal lattice.

The wave function is written as the product of the periodic part and the wavelike part:

$$\psi_i(\mathbf{r}) = e^{ik \cdot \mathbf{r}} f_i(\mathbf{r}) \quad (1.62)$$

The first part of the equation relates to the wavelike part of the wave function; the second part relates to the periodic part. The periodic part however can be further expanded by re-writing it as a finite number of plane-waves whose wave vectors are reciprocal lattice vectors of the crystal.

$$f_i(\mathbf{r}) = \sum_{\mathbf{G}} c_{i,\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \quad (1.63)$$

Where now \mathbf{G} are the reciprocal lattice vectors. Thus the electronic wave function is written as a sum of plane-waves

$$\psi_i(\mathbf{r}) = \sum_{\mathbf{G}} c_{i,\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \quad (1.64)$$

Thus the problem of the infinite number of electrons has now been mapped onto the problem of expressing the wave function in terms of an infinite number of reciprocal space vectors within the first Brillouin zone of the periodic cell, \mathbf{k} . This problem is solved by a method known as Brillouin zone sampling which forms the basis of the next section.

The electronic wave functions at each point are now expressed in terms of a discrete plane-wave basis set. In principal this Fourier series is infinite, however introduction of a plane-wave energy cut-off reduces the basis set to a finite size. The coefficients for the plane-waves $c_{i,\mathbf{k}+\mathbf{G}}$ each have an associated kinetic energy $(\hbar^2/2m)|\mathbf{k} + \mathbf{G}|^2$. The plane-waves with a smaller kinetic energy typically have a more important role than those with a very high kinetic energy.

By implementing the plane-wave basis set on the Kohn-Sham equations gives

$$\left\{ \sum_{\mathbf{G}'} \frac{\hbar^2}{2m} |\mathbf{k} + \mathbf{G}|^2 \delta_{\mathbf{G}\mathbf{G}'} + V_{Hart}(\mathbf{G} - \mathbf{G}') + V_{ext}(\mathbf{G} - \mathbf{G}') + V_{XC}(\mathbf{G} - \mathbf{G}') \right\} \times c_{i,\mathbf{k}+\mathbf{G}'} = c_{i,\mathbf{k}+\mathbf{G}} \varepsilon_i \quad (1.65)$$

From this it can be seen that the reciprocal space representation of the kinetic energy is diagonal and the various potentials appear in terms of their corresponding Fourier components. From this the plane-wave expansion of the Kohn-Sham equations is solved by diagonalization of the Hamiltonian matrix whose elements are shown above in (1.65). The size of the Hamiltonian matrix is proportional to the energy cut-off.

1.7 Brillouin zone sampling

For periodic systems the calculations used to solve for the total energy are always performed on a single unit cell with periodic boundary conditions. From Bloch's theorem, the electronic structure calculation can be changed from one of trying to calculate an infinite number of states to one of calculating finite number of bands at an infinite number of \mathbf{k} points within the Brillouin zone (Paier, et al., 2006) (Heyd, et al., 2003). A \mathbf{k} -point is a reciprocal lattice vector which is used to label an eigenstate. From equation (1.62) $f(\mathbf{r})$ can be expressed as

$$f(\mathbf{r}) = \frac{\Omega_{cell}}{(2\pi)^3} \int_{BZ} \mathbf{F}(\mathbf{k}) d\mathbf{k} = \sum_j w_j \mathbf{F}(\mathbf{k}_j) \quad (1.66)$$

Where, Ω_{cell} is the cell volume, $\mathbf{F}(\mathbf{k})$ is the Fourier transforms of $f(\mathbf{r})$ and w_j are the weighting factors. The volume of the Brillouin zone Ω_{BZ} , is related to the volume of the supercell Ω_{cell} by

$$\Omega_{BZ} = \frac{(2\pi)^3}{\Omega_{cell}} \quad (1.67)$$

For larger supercells such as the cells used for semiconductor defect ab-initio calculations the Brillouin Zone becomes very small and so fewer \mathbf{k} -points can be used to sample the Brillouin Zone.

1.8 Pseudopotentials

Pseudopotentials replace the effects and interactions of the core electrons, (non-valence electrons), of an atom and its nucleus with an effective replacement potential. This replacement potential contains the modified effective potential terms, including the effect of the core electrons and smoothing the wave function, thus having less nodal points and leading to the requirement of fewer Fourier components. This effective potential replaces the Coulombic potential in the Schrodinger equation.

In order to construct the effective potential the region close to the atomic nucleus needs to be closely looked at since firstly due to the Pauli exclusion principal, the valence electron wave function must be orthogonal to the core electron wave functions. Thus the frequency of oscillations of the valence electron wave functions within the core region will be very high. Secondly the electron potential varies in inverse proportion to the distance from the center of the nucleus and thus it diverges as that radius tends toward zero.

In order to describe these wave functions, a large basis set of plane-waves is required, as well as a large number of plane-waves being required to describe the tightly bound core states.

This is where a pseudopotential approximation can be used effectively. The approximation uses the fact that most physical properties are determined by the valence structure of the atom; the strong Coulomb potential and the core electrons are replaced with an effective potential that is weaker, the pseudopotential. Within the core region the valence wave functions are replaced by smooth nodeless pseudo-wave functions that are identical to the real wave function outside the core region.

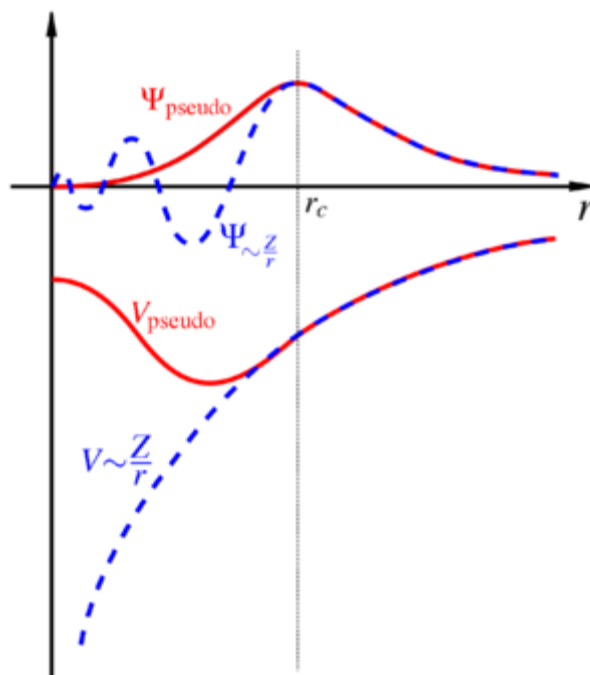


Figure 1.5: Comparison of a wave function in the Coulomb potential of the nucleus (dotted) to the pseudopotential (solid), the real and the pseudo wave function and potentials match above a certain cut-off radius. (Alpha, 2006)

This process lowers the complexity of the problem in three fundamental ways: firstly, the number of wave functions needed for the calculation is drastically reduced since the core electrons are not considered. Secondly, the potential no longer diverges upon approaching the core of the nucleus and finally the valence wave functions are smooth within the core region. The pseudopotential method however has one specific requirement; the pseudo-wave function must be identical to the all-electron wave function outside the core, to produce the same energy difference (Louie, 1982).

Chapter 2

The Science of semiconductors

2.1 Introduction

A semiconductor is a material which has the electrical properties between those of conductors and insulators. The most desirable properties of semiconductor materials is how the electrical conductivity can be changed by applying variations to the environmental conditions by either varying temperature, applied external fields or by the creation of defects within the semiconductor.

2.2 Doping of semiconductors

Intrinsic semiconductors are semiconductors that have no impurities within their atomic structure (or at least so few that they do not influence the free carrier density). The free carriers in this case are generated by the process of electron-hole pair generation. In intrinsic semiconductors, the number of holes will always be the same as the number of electrons.

Extrinsic semiconductors are semiconductors in which the carrier density is determined by impurity atoms in the host lattice. These impurity atoms can be introduced in various ways and are used to alter the properties of the semiconductor. The process of introduction of impurities onto a semiconductor is called “doping”. In general doping produces two groups of semiconductors. If the impurity has more valence electrons than the host atoms, this leads to an excess of electrons with positively charged impurity atoms (called donors) and the material is called n-type. Similarly, in p-type material, an impurity with less valence electrons will lead to positively charged holes forming in the valence band, as electrons are captured by the impurity atoms called acceptors, which become negatively charged.

A donor is classified as an impurity defect within a semiconductor which contributes (donates) free electrons to the semiconductor. In a semiconductor such as germanium (Ge) donors include group V elements (e.g. phosphorus (P), arsenide (As) and antimony (Sb)) acting as substitutional defects in

the lattice. The process by which donors donate electrons to the semiconductor is well understood. The group V elements have an extra valence electron when compared to the lattice atoms (the case of Ge comprise of Group IV atoms). This extra electron is not tightly bound to the donor and can easily be ionized thermally. Much in the same way as Group V atoms donate 1 electron per defect and Group VI impurities donate 2 electrons. These impurity atoms are known as double donors.

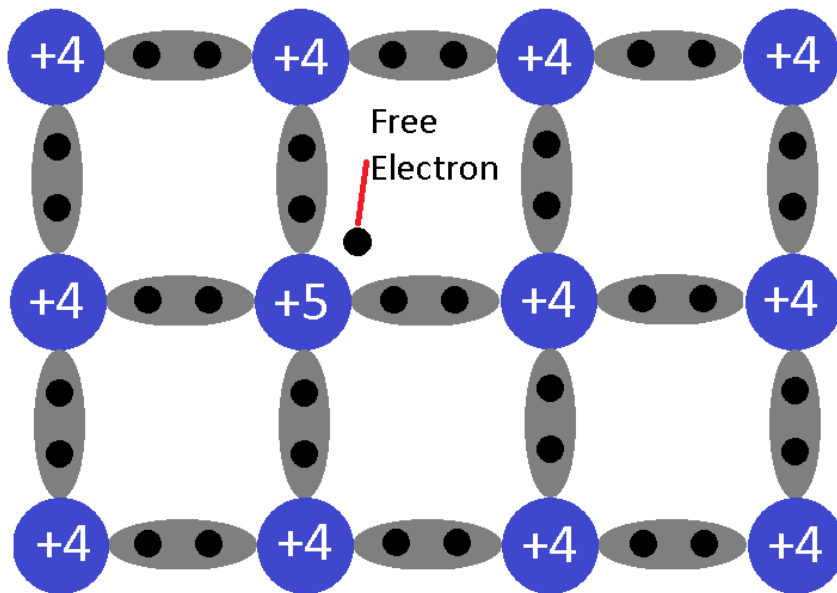


Figure 2. 1: Schematic representation of a semiconductor doped with a donor. A free electron is now present.

Acceptors are defect impurities which contribute to holes in the semiconductor. Acceptor doped semiconductors are referred to as p-type semiconductors. In germanium examples of acceptor impurities are the Group III elements (boron (B), aluminum (Al), gallium (Ga)). These elements have one less electron in the valence shell when compared to germanium. This causes a vacant state close to the valence band, which, by thermal excitation, can be filled by an electron from the valence band leaving a hole. The hole is not tightly bound to the impurity atom and thus can be ionized at temperatures above 0 K. Group II atoms add 2 holes in the system and hence are referred to as double acceptors.

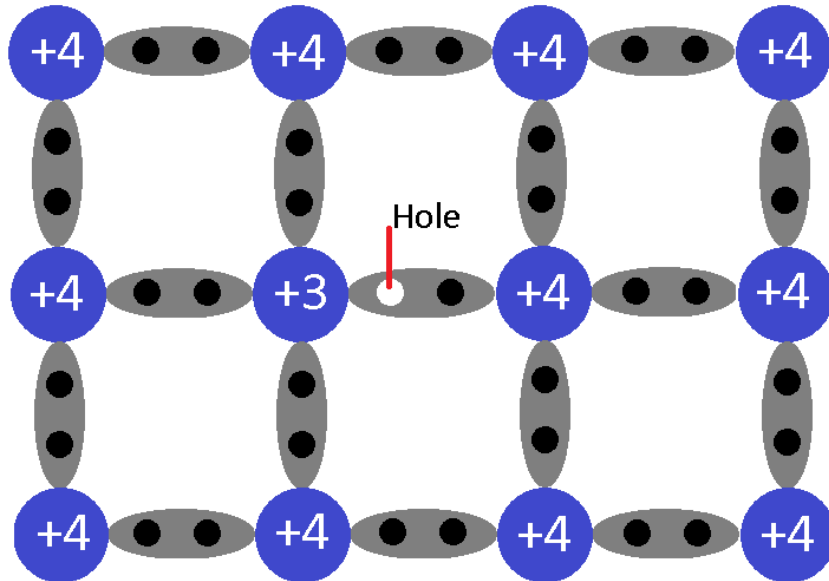


Figure 2. 2: Schematic representation of a semiconductor doped with an acceptor. A free hole is now present.

2.3 Structural properties of semiconductor defects

As already stated defects in semiconductors are structures which can have positive or detrimental effects on the properties of the semiconductor. These defects need to be classified in their structural properties to better understand the nature of the benefits or hindrances they offer. Defects can occur as either point defects or as part of larger defect complexes. Point defects are lattice imperfections caused by missing atoms (vacancies), an interstitial atom or impurity atom. While defect complexes are a combination of two or more point defects that occur (relatively) close to each other such as: interstitial-vacancy pairs, impurity-vacancy pairs, substitutional-vacancy pairs and divacancies, are to name but a few.

2.3.1 Primary and secondary defects

Primary defects are defects which are initially formed in semiconductors as a result of irradiation. These defects are mainly point defects. Secondary defects are mainly defect complexes which are formed when point defects combine into larger more complex structures, e.g. vacancies and

interstitials. These tend to be highly mobile in the semiconductor lattice and interact with other defects in the lattice.

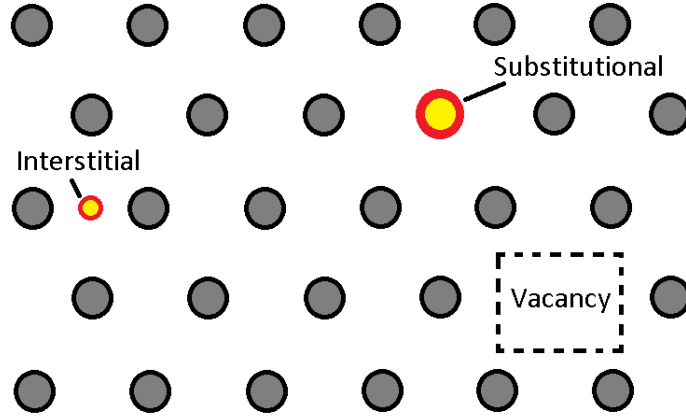


Figure 2. 3: Point defects in semiconductor lattice.

Secondary defects are defect complexes formed when point defects e.g. vacancies and interstitials, which are highly mobile within the lattice, interact with other defects in the lattice forming defect complexes

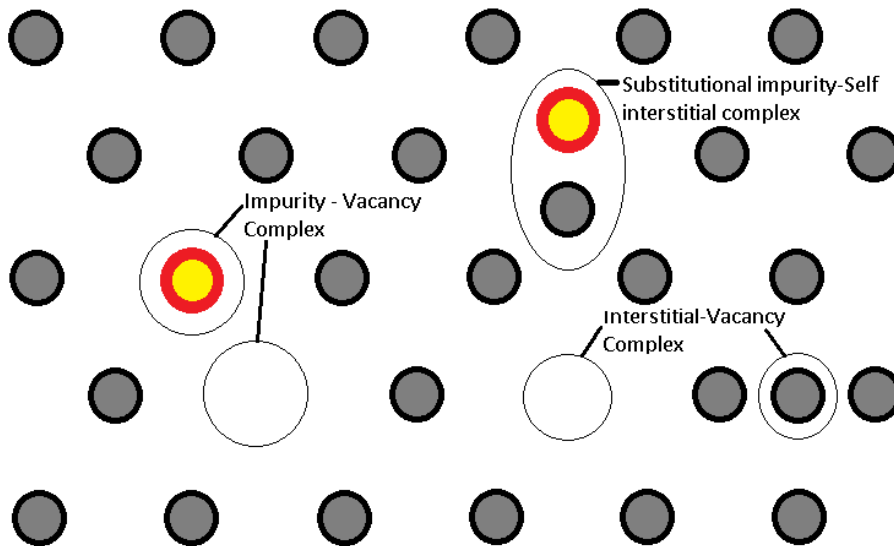


Figure 2. 4: Secondary defect complexes in the semiconductor lattice.

2.4 Electronic properties of semiconductor defects

Defects in semiconductors have been studied for many years and have been found to have notable effects on the electronic properties of the device with whom they are attributed. Two factors which determine the properties of a defect are namely the **local factors**: such as the chemical makeup and size of the defect. Defects lead to changes in the translational symmetry of the semiconductor. **Host factors**: these factors involve the properties of the host atomic makeup, whether the makeup be either n-type or p-type material.

In semiconductors there may exist the following two types of electronically active defects:

1. Shallow level defects
2. Deep level defects

This distinction is based on the position energy level of the defect in the band gap with respect to the conduction or valence band. Deep level defects are classified to have highly localized wave functions whereas shallow level defects wave functions are considered as the far-reaching Coulomb potential. Antimony (Sb) is classified as a shallow level donor impurity in germanium.

Shallow level defects

Doping can affect the state of the intrinsic point defect in crystals. The dislocation behavior and associated impurities can eventually result in new possibilities of controlling the properties of a semiconductor material and of the structures built on it. Thus doping with shallow-level impurities controls the conduction type and the charge carrier concentration in a semiconductor.

A shallow-level defect has features similar to that of the hydrogen atom where we have a positive nucleus binding to a valence electron. There however are two differences to the hydrogen model which come about

1. Renormalizing of the mass of the electron due to its interactions with the lattice, thus we get a smaller effective mass than that of the free electron.
2. The crystal reduces the binding potential due to its associated dielectric constant.

These effects reduce the known ionization energy (ground state binding energy) of the donor atom/s so that they require little energy for ionization, energy that is available even at room

temperature. The system still displays other hydrogenic atom qualities however, such as, a series of bound excited states and the ionization into the energy continuum, i.e. the conduction band in the case of the donor.

Deep level defects

The second types of defects in semiconductors are those centers whose electronic structures are affected significantly by a short-range (central cell) potential. They generate the bound state with well-localized wave functions. These defects are called deep or localized defects. They can act as recombination-centers or deep level-traps or localized levels. They limit the lifetime of the carriers and may compensate for the shallow donors or acceptors or they may induce an effect known as Fermi level pinning (the center of the band gap becomes pinned to the Fermi level). Their energy levels lie “deep” within the forbidden energy band and are much more difficult to understand than the shallow-level counterparts (Pantelides, 1992). Impurity atoms which are improperly placed in the lattice distort the lattice extensively such that the binding energy greatly exceeds that of the shallow hydrogenic defects. Many atoms around the distorting defect must be considered when a first-principles calculation is attempted; the defect wave function extends over many lattice constants. The carriers interact strongly with the lattice. The effects of deep level defects are to drastically reduce minority carrier lifetimes and to act as traps for charge carriers.

2.5 Charge distribution in semiconductors

When studying a semiconductor’s electrical properties or any defect’s electrical properties it is important to know the Fermi level E_F position with respect to either that of the valence E_V or conduction E_C band. The reasons for this are many but depending of the semiconductor type, either n- or p-type and the applied bias, differing outcomes will present themselves, as the Fermi level will shift either towards the valence band or the conduction band. Application of a bias across a Schottky barrier changes the difference between the position of the Fermi level of the semiconductor and the metal in the depletion region.

2.5.1 Schottky barriers

The concept of a rectifying Schottky barrier is based on which the barrier being large enough that there is a depletion region or a insulating layer with the conductive semiconductor material where the mobile charge carriers have diffused away or been forcibly moved by application of an external field, near the interface layer of the semiconductor. This barrier is a layer of high resistance when a small voltage is applied but under larger applied voltage, the electric current is governed by the laws of thermionic emission. Depending on the direction of the applied bias as well as the type or rectifying junction it is applied to, the electric current may be allowed to pass the barrier and not be allowed to flow in reverse. One must note that there is a point at which the barrier can be overcome and this process results at high enough applied bias's that can cause breakdown of the depletion region and loss of rectification properties.

Fabrication of a Schottky barrier is a science all to its own, but occurs when a metal or in some cases dielectric contact is evaporated by processes such as E-beam deposition, resistive evaporation or metallic sputtering. A potential barrier is formed at the metal-semiconductor interface (Streetman & Banjeree, 2000). Since the metal and semiconductor have different work functions, charge transfer will continue until the Fermi level of the metal E_{FM} aligns with that of the semiconductor E_{FS} . This alignment is referred too simply as the equilibrium. In n-type semiconductors, the work function of the metal is higher than that of the semiconductor and thus E_{FS} is shown to be higher than that of E_{FM} before the contact is made.

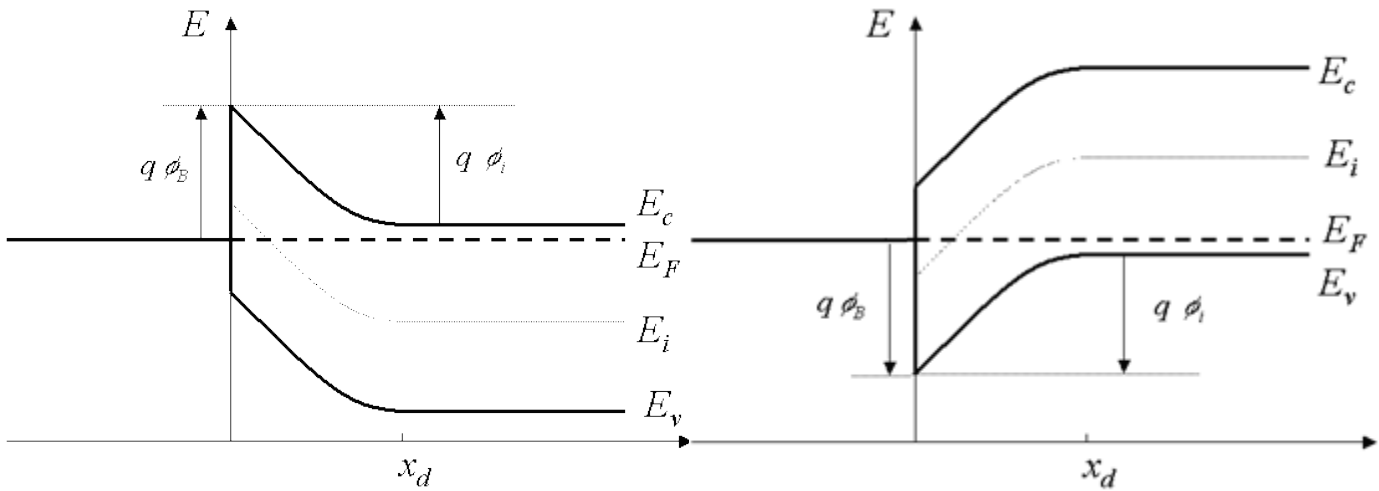


Figure 2. 5: Schematic representation of an n-type (left) and p-type (right) Schottky barrier at equilibrium $E_{FS} = E_{FM} = E_F$.

n-Type semiconductors

Under zero applied bias conditions, the electrons from both the metal and semiconductor junction require the same energy to move over the barrier height $q\phi_B$ relative to the Fermi energy, for no net flow of electrons in either direction. Application of a bias however results in variations to the width of the depletion region and also the amount of band bending experienced. The Fermi level therefore can be moved within the band gap by variations to the applied conditions.

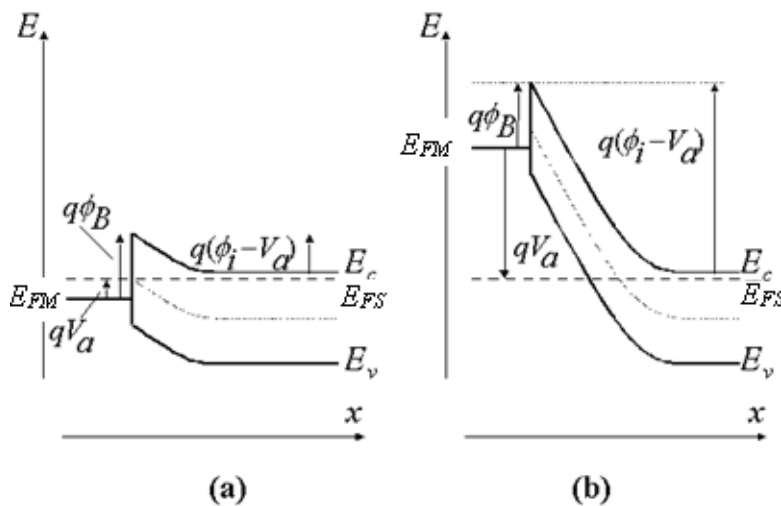


Figure 2. 6: Representation of a Schottky barrier under forward (a) and reverse (b) bias for an n-type semiconductor.

Under reverse bias conditions E_{FS} is lowered relative to E_{FM} . This results in band bending and an increase in the width of the depletion region.

p-Type semiconductors

For p-type semiconductors, there exist differences from their n-type counterparts. The Fermi level of the p-type semiconductor is close to the valence band edge and as a result, occupied defect states in the band gap emit electrons since holes are the majority charge carriers. By fabrication of a Schottky diode and formation of a depletion region, bands close to the interface bend upward since an electric field exists within the depletion region. Band bending is increased by application of a reverse bias, shifting the defect level to above that of E_{FM} . Defects within the band gap capture electrons and emit “holes” and thus for a large part of the depletion region, the semiconductor band gap lies above the E_{FM} and below the Fermi level of the metal.

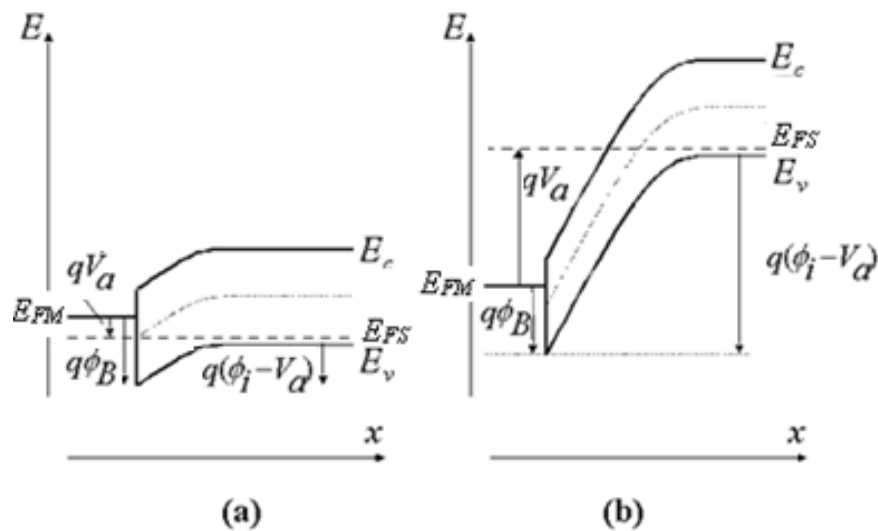


Figure 2. 7: Representation of a Schottky barrier under forward (a) and reverse (b) bias for a p-type semiconductor. Adapted from (Van Zeghbroeck, 2011).

2.6 Electrical characterization of semiconductor defects

Experimental techniques have evolved and adapted over time to measure and quantify the properties of semiconductor defects. Tools such as Capacitance Voltage (C-V) and Current Voltage (I-V) setups are common place in today's semiconductor laboratories worldwide. More complex techniques such as Deep Level Transient Spectroscopy (DLTS) (1974) as well as Laplace DLTS (1994) were developed years after the semiconductor became of interest to the scientific world.

The details for the following methods can be further explored with reference to Schroder et al. (Schroder, 2006) yet will not be explored in great detail as they form part of the experimental reference to this work only.

2.6.1 Capacitance–voltage technique (C-V)

This technique makes use of a p-n junction or Schottky barrier diode to create a depletion region, a region devoid of conducting holes and electrons. The depletion region may however contain ionized donors and electrically active defects or traps. This depletion region thus acts like a capacitor and for this reason we can measure its electrical properties by applying a varying voltage across it and the capacitance can be measured as a function of voltage. By varying the applied voltage we vary the width of the depletion region, this provides information on the electrical properties, such as the doping profile and any electrically active defect densities. (Diebold, 2001)

The capacitance per unit area \mathbf{A} of a Schottky barrier diode to an n-type substrate is given as

$$\frac{C}{A} = \left[\frac{qN_D\epsilon_s}{2\left(V_{bi}-V-\frac{kT}{q}\right)} \right]^{\frac{1}{2}} \quad [2.1]$$

This can be re-arranged to

$$\frac{1}{C^2} = \frac{2\left(V_{bi}-V-\frac{kT}{q}\right)}{A^2qN_D\epsilon_s} \quad [2.2]$$

Where V_{bi} , V , k , T , q , N_D , N_A and ϵ_s are the built-in voltage, applied voltage, Boltzmann constant, temperature in K, electron charge, majority carrier concentration, minority carrier concentration and the static dielectric constant of a semiconductor respectively. This assumes $N_D > N_A$. (Coelho, 2014)

2.6.2 Current–voltage technique

The simplest form of the current voltage relation is that of a resistor which according to Ohms law results in a linear relation between the applied voltage and the electric current measured. However for electronic devices that makes use of a p-n junction this relation is a bit more complex. A number of mechanisms exist within a semiconductor to capture slow-moving electrons of low energy and low voltage. Thus under larger applied currents these smaller effects are saturated and not observable, yet when one applies a lower current these capture methods end up representing an increasing percentage of the total throughput. Thus a method was developed to study semiconductors near to this insulation effect and to characterize the electrical properties depending on the type of p-n junction in question.

2.6.3 Deep level transient spectroscopy

Deep level transient spectroscopy (DLTS) is one of the most versatile tools for studying electrically active semiconductor defects. Based on the work of David Lang in 1974 (Lang, 1974). Only the basic principles of DLTS will be discussed here, a review of more advanced techniques can be found in the work of L. Dobaczewski (Dobaczewski, et al., 1994).

The basic DLTS method requires the construction of either by either direct fabrication or formation of a Schottky diode or a p-n junction. This method investigated defects in a depletion region of a semiconductor junction. In the measuring process the steady state diode's reverse polarization voltage is disturbed by a voltage pulse. This pulse causes a reduction in the electric field in the space charge region (depletion region) and this allows free charge carriers from the semiconductor bulk to leech into this area and recharge, causing them to take on a non-equilibrium charge state.

After the pulse, the defects emit the trapped charge carriers due to thermal emission. This is where the method of DLTS observes the device's space charge region capacitance where the defect charge state recovery causes the capacitance transient. The voltage pulse followed by the defect charge emission is cycled, allowing different signal processing methods to probe the recharging process.

2.7 Thermodynamic properties of defects

2.7.1 Formation energy

The formation energy of a defect is defined as the difference between the total energy of the supercell with the defect and the supercell without the defect, i.e. the perfect crystal. It reflects the cost in energy due to the lattice stress caused by the presence of the defect in the semiconductor. Introduction of defects results in an increase of entropy as the system now can have multiple microstates unlike the pristine system. The concentration of defects c_i must adhere to the following relation:

$$c_i \propto \exp\left[-\frac{\Delta G_f^i}{k_B T}\right] \quad (2.3)$$

Where i is the defect in question and ΔG_f^i is the Gibb's free energy of formation of the defect Also referred to as the defect formation energy. The expression, $k_B T$ has the usual meaning. (Van de Walle, 2004)

2.7.2 Activation energy of defect migration

Defect migration deals with the concept of the motion of point defects and defect complexes within semiconductors. The mobility of such defects depends on many aspects of the defect in question, namely; the size of the defect; the makeup of the complex and the interatomic spacing within the lattice among other aspects. By considering the point defect, the simplest being that of a vacancy within the lattice; the ability of the vacancy to move from site A to site B within the lattice requires energy. This energy can come from latent thermal energy or applied energy. The process however will start at the minimized energy for the defect in Site A and it must progress through the saddle point energy at its maximum, and to the corresponding Site B and returns to a minimum. The activation energy for the defect migration is equal to the difference between the maximum of the saddle point and the minimum of the equilibrium state. (Matsushita, et al., 2007)

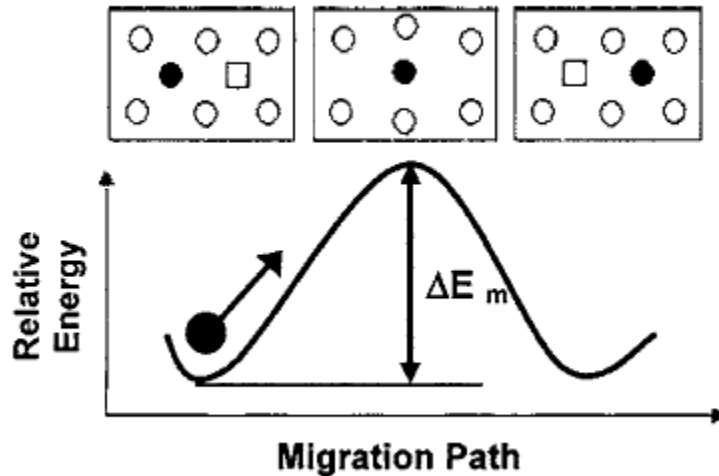


Figure 2. 8 Visual representation of the migration activation energy of a point defect in a lattice. Adapted from (Matsushita, et al., 2007).

2.7.3 Binding energy of defect complexes

Defects in semiconductors do not only exist as point defects. Electronically active defects often exist as defect complexes which are made up of more than one defect. The binding energy of a defect complex is the energy required to dissociate the defect into its constituent parts.

$$E_B = E^f(A) + E^f(B) - E^f(AB) \quad (2.4)$$

The binding energy of defect AB is thus equal to the formation energy of the point defect A and the point defect B minus the formation energy of the defect complex AB.

2.8 Annealing properties

Annealing or heat treatment of defects is a process by which defects can be removed from a semiconductor. To be effective the defect concentration is required to be high compared to that of the thermodynamic equilibrium concentration. Annealing is always characterized by activation energies. This process occurs due to various factors; these factors include mobile defect migration and recombination with their counterparts as well as the dissociation of defects into their constituent parts and finally, defects diffusing deeper into the bulk or towards the surface. For any given temperature T_a , defect annealing corresponds to a decrease in defect concentration as a function of time.

2.9 The Poole-Frenkel effect

When an external electric or magnetic field is applied to an electron trapped in a potential well, the electron is subjected to the sum of properties of both fields. This process causes the shape of the potential well to be distorted by increasing the height of the barrier on the one side of the defect and lowering it on the other. This method known as the Poole-Frenkel effect is the simplest way to enhance the emission of an electron from a potential well.

Frenkel developed the original version of the theory in 1938 (Frenkel, 1938) which only dealt with the 1 dimensional case.

According to that model, the ionization energy of a Coulombic well subjected to an externally applied electric field F , is lowered by:

$$\Delta E_T = \frac{1}{kT} \sqrt{\frac{qF}{\pi\epsilon}}. \quad (2.5)$$

The emission rate of the defect can then be expressed as:

$$e(F) = e(0) \exp\left(\frac{1}{kT} \sqrt{\frac{qF}{\pi\epsilon}}\right) \quad (2.6)$$

Where $e(0)$ is the emission rate at a zero electric field strength, k the Boltzmann constant and T is the absolute temperature.

The emission enhancement due to the Poole-Frenkel effect is frequently used by experimentalists to estimate the range of the defect potential. A potential with a longer range would show a far stronger Poole-Frenkel enhancement than a shorter one.

The dependence of the emission rate e on the electric field strength F has been used as experimental evidence to differentiate between donor and acceptor defects. The linearity of the log of the emission rate on the square root of the field is characteristic of a charge leaving a center of opposite sign. In n-type semiconductors this would imply a donor type defect and in p-type semiconductors it would imply an acceptor type defect (Meyer, 2007).

Chapter 3

Defect metastability

3.1 Introduction

Defects can either be classified as being stable or metastable. Stable defects exist where the defect can be characterized with one minimum-energy atomic configuration at a particular charge state, but whose characteristics are still dependent on extrinsic or applied properties (i.e. applied fields, bias's and/or annealing properties).

Defects have more than one minimum energy atomic configuration depending on the experimental conditions applied to them. The property of metastability in defects is often not noticed or goes unreported. Metastable defects have levels which are easily removed or introduced by applications in experimental conditions and thus can be manipulated extensively by applied fields and biases as well as pre and post fabrication annealing.

Applications of metastability have recently come to the forefront of semiconductor research with the effect claiming larger amounts of responsibility in breakthroughs in the field of higher conductivity in existing semi conductors such as gallium doped ZnO (Zakutayev, 2013).

Ab-initio defect studies have in the past mainly focused on well understood and simple defect complexes and stable defects. Yet as the science progresses further, more work is now being conducted in the field of modeling possible metastability of known defect structures, in the hope of mapping the results once thought to have been understood, to provide both the experimental and theoretical results.

3.2 Bistability and Multistable defects

The definition of metastability states that the defects will have more than one stable configuration state in at least one of its charged states. Figure 3.1 left (a) and (b) shows a configuration co-ordinate diagram of stable defects whereby the position of the energy minima does not depend on the charge state of the defect. Stable defects have a unique stable atomic configuration, with respective lattice relaxations for some charge state. Large lattice relaxations can lead to significant changes in configuration with charge state variations, yet there is still only one minimum configuration energy in the observed charge state.

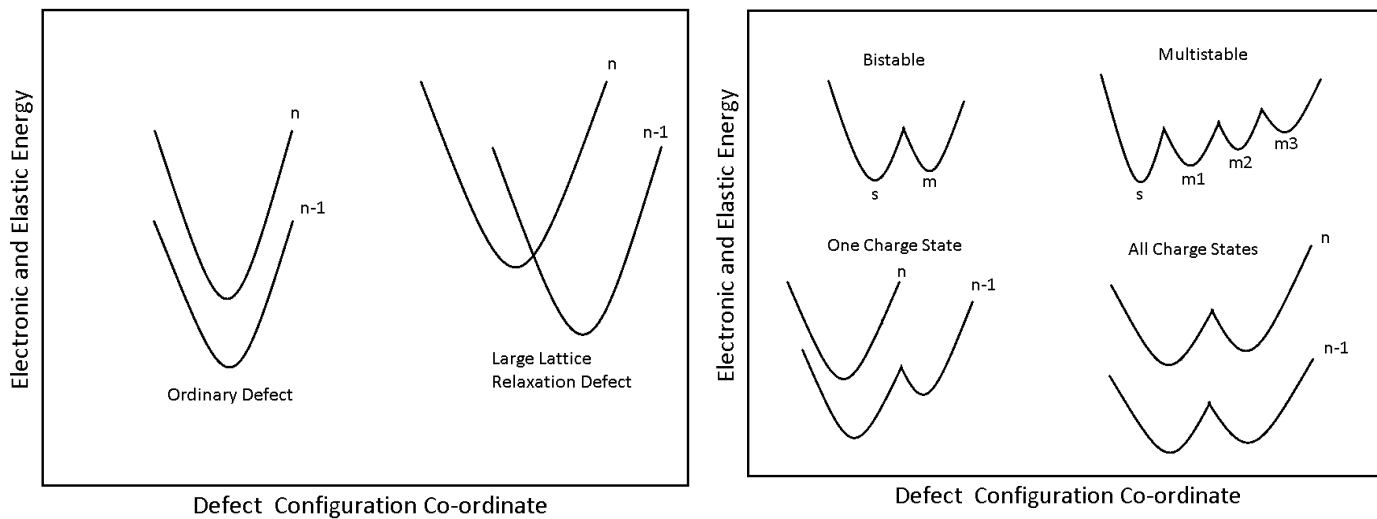


Figure 3. 1: Left: configuration diagrams from usual semiconductor defects. Right: potential configuration diagrams for charge state dependent metastability.

Defects can be classified as configurationally bistable if, there exists two possible fixed atomic configurations for one particular charge state. The lowest energy configuration is referred to as the stable configuration whilst the higher configuration is referred to as the metastable configuration. The two configurations must be separated by a finite barrier such that only through a thermally activated process, can the metastable configuration decay into the stable state over time.

3.3 Charge-state-controlled metastability

A defect which has two (or more) different minimum energy atomic configurations for the neutral and two charged states is referred to as a charge state controlled metastable defect. In order for charge-state controlled metastability to occur, the defect will have two or more stable atomic configurations for the neutral state with each of the two defect charge states considered.

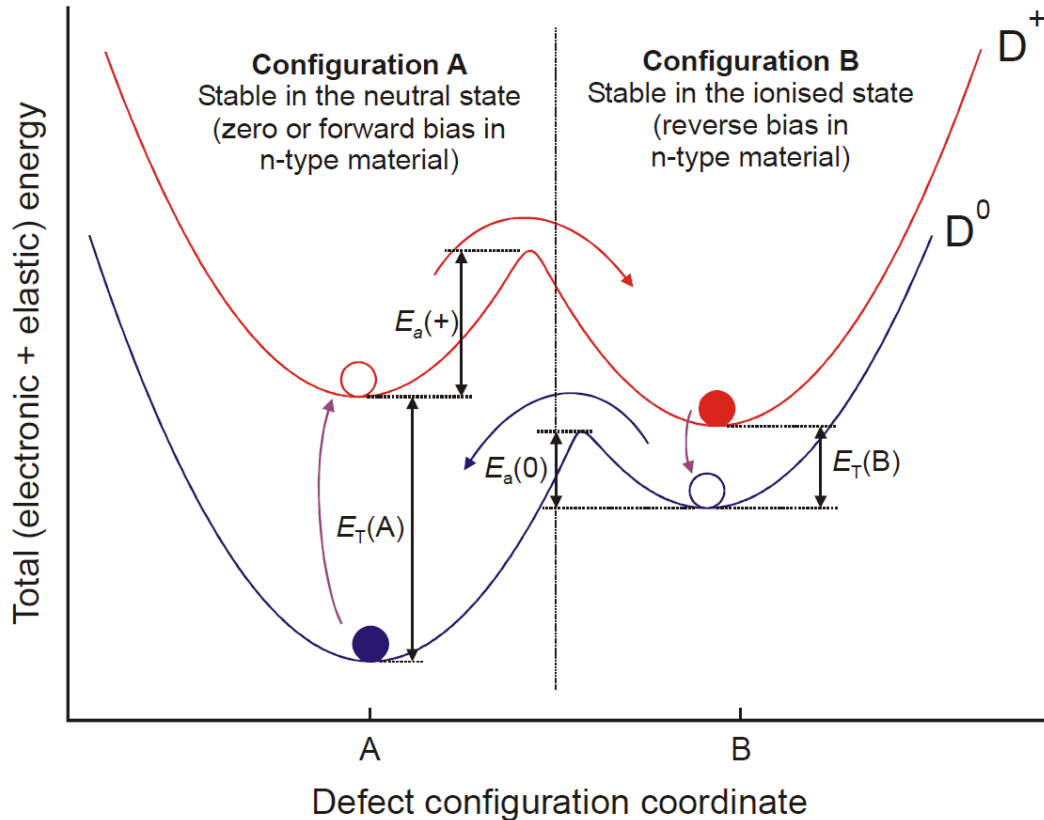


Figure 3. 2: A configuration co-ordinate diagram showing charge state controlled metastability in two configurations namely A and B. Blue and red curves are for the neutral and ionized states respectively adapted from (Meyer, 2007).

The mechanism of charge state controlled metastability can be explained with the aid of figure 3.2. The neutral state denoted as configuration A is the stable minimum energy while in the ionized state denoted as configuration B is the minimum energy configuration.

To illustrate further, assuming the barriers of $E_a(0)$ and $E_a(+)$ are easily overcome at 300 K and the defect is frozen in its current state at 100 K, the defect can change between the two shown configuration states, namely configuration A and B by changing the charge state of the defect from one of its charged states to a neutral state or vice versa. This process of changing the charge state of a defect can be done with the application of a suitable forward or reverse bias to the Schottky contact.

Depending on whether the semiconductor is n- or p-type, we can influence its charge state with the application of a forward or reverse bias. For the n-type semiconductor, application of a forward bias leads to the filling of a defect level with an electron leaving the defect in its neutral state and consequently configuration A is more energetically favorable in its neutral state than an unionized state. Most defects will therefore favor configuration A. Similarly application of a reverse bias allows the defect to emit an electron, thus leaving the defect in its ionized state. In this state configuration B is more energetically favorable, hence these defects will be found more often in this configuration.

By cooling the semiconductor down to 100 K while keeping the applied bias conditions constant, it will allow for the stable configuration to be frozen in. If, at this low temperature, the carrier emission and capture rate of the defect is much higher than the transformation rate, the charge state can be changed without affecting the defects' configuration by changing the applied bias.

3.4 Negative-U defects

A negative-U defect is capable of trapping two electrons (or holes), with the second trapped charge carrier bound more strongly than the first. Negative-U ordering of defect levels leads to interesting metastability phenomena.

By making use of a hypothetical argument consider a defect which can exist in three charge states namely the +1, 0, -1 state. If the defect is fully ionized, i.e. in the +1 state, it can capture two electrons, the first of which corresponds to a donor level, which is positive when above the Fermi level and neutral when below and the second level corresponds to an acceptor level. The energy difference between the conduction band minimum and the donor level is the binding energy of the first electron and the energy difference between the conduction band edge and the acceptor level is the binding energy of the second electron. U is defined as the difference between the binding

energies (binding energy of first electron minus binding energy of the second electron) and is also called the interaction-energy between the electrons. Coulomb interaction between the two captured electrons would naturally result in the second electron being more loosely bound than the first and thus the acceptor level would lie above the donor level in the band gap. Since the Coulomb interaction is repulsive the expectation is that U should always be positive. However experimental observations have seen these defect levels are inverted, thus a negative- U system exists, which implies that the second electron is more tightly bound than the first, as if an attractive potential existed between the electrons. (Meyer, 2007).

3.4.1 Mechanisms leading to negative- U behavior

Negative- U properties of chalcogenide glasses was first explained by (Anderson, 1975) with a model showing that in addition to the restoring force that acts between atoms, the level of a state depends linearly on the displacement x between the two atoms. This model was expanded upon by (Street & Mott, 1975) and applied to point defects in amorphous glassy semiconductors.

The Anderson model described the potential energy of an atom in the following form:

$$V = -\lambda x(n_{\uparrow} + n_{\downarrow}) + \frac{1}{2}cx^2 \quad (3.1)$$

where n_{\uparrow} and n_{\downarrow} are the occupancies for spin-up or spin-down bond orbitals and x is representative of position but can be generalized to a configuration co-ordinate. By setting the derivative of the potential energy $\delta V/\delta x = 0$ and substituting this back into Equation (3.1) it was found that the potential energy is lowered by $-\lambda^2/2c$ for single occupancy and $-2\lambda^2/c$ for double occupancy. The net effective correlation energy, which is defined as the difference between two singly occupied bonds and the disproportionated state (one empty and one doubly occupied) becomes:

$$U_{eff} = U - \frac{\lambda^2}{c} \quad (3.2)$$

where U is the normal Hubbard correlation energy that would be used if no lattice relaxation were present. It therefore follows that if the lowering of the single occupancy potential energy is large enough, the correlation energy would become negative and the disproportionate state will become energetically more favorable than the two singly occupied states (Meyer, 2007).

3.4.2 Properties of negative-U defects

The neutral state of a negative-U defect is no longer thermodynamically stable. If two isolated neutral defects diffuse into contact with each other (via the conduction band), they can lower their respective energies by ionizing and releasing the energy U



Thus, in the ground state, a negative-U defect will always exist in a particular charge state, this implies that the ground state is either the negatively ionized or the positively ionized charged state and the neutral state then becomes defined as an excited state of the defect (Meyer, 2007).

Chapter 4

Modern techniques applied to computational defect studies

4.1 Ab-initio techniques applied to bulk properties

Defects within semiconductors can be found at any position within the bulk and even at the surface layers. The properties of semiconductors relating to both physical and electrical properties can be predicted extremely well using DFT yet there is strong evidence that these predictions are not always precisely accurate and this inaccuracy has been linked to approximations made in the exchange-correlation energy function as it relies on the unknown form of the functional required to complete the calculation. Properties which are predicted with minor inaccuracy are related to lattice constants, bulk moduli and internal stress and strain of the structure in question, so this must be kept in mind when reviewing results. The exchange-correlation functional choice will determine whether a property is over or underestimated. Functionals such as the LDA functional underestimates lattice constants (Van de Walle & Ceder, 1999) which in turn overestimates the bulk modulus. While functional classes such as the GGA collection of functionals, generally overestimate lattice constants and underestimate the bulk moduli (Haas, et al., 2009). Even with these inaccuracies in place DFT has proven to be one of the most accurate methods for theoretically determining properties of semiconductors in recent history.

4.2 Ab-initio techniques applied to electronic properties

All modern standard approximations to the exchange-correlation functional within DFT are known to underestimate the band gap of semiconductors by about 40% (Perdew & Levy, 1983). In order to correct for this error a class of functionals known as the hybrid functionals was developed to more accurately predict the properties of semiconductors and a more accurate calculation of the band gap was a result of this development.

4.2.1 Errors related to prediction of the band gap and correction methods

The band gap is defined as the difference in energy between the bottom of the conduction band and the top of the valence band of the semiconductor. There are two ways in which we can calculate the band gap by using ab-initio techniques. Firstly by using the Kohn-Sham energy levels E_{KS} and secondly by the quasi-particle method E_{QP} . By using the Kohn-Sham levels the band gap can be obtained from: $E_{KS} = E_C - E_V$ where E_V and E_C are the values for the highest occupied Kohn-Sham level (HOMO) and the lowest unoccupied Kohn-Sham level (LUMO) (Lany & Zunger, 2008). The band gap energy can also be defined as the difference between the first ionization energy and the electron affinity. For large supercells, comprising structures of 32 – 256 atoms and larger, the ionization potential (IP) and electron affinity (EA) are defined as follows:

$$IP = E_H(-) - E_H(0) \quad (4.1)$$

$$EA = E_H(+) - E_H(0) \quad (4.2)$$

Where $E_H(0)$ is the energy of the homogeneous host with no defect present, with reference for the charge state of the host. This method is referred to as the “so-called” quasi-particle (QP) band gap energy (Lany & Zunger, 2008) and is given by:

$$E_{QP} = EA - IP \quad (4.3)$$

The exchange-correlation (XC) functional

Inaccuracies associated to LDA and GGA are due to incorrect assumptions and the self-interaction errors discussed in chapter 1.5 (Perdew, et al., 1982) (Perdew & Levy, 1983) (Sham & Schluter, 1983). One of the well-known problems with both LDA and GGA XC functionals is the inability to

correctly predict the band gap in semiconductor materials. An example of this is the case where germanium is predicted to have no band gap at all and is modelled to be metallic rather than a semiconductor. This error however is not attributed to one XC functional and is a general problem that all XC functionals share (Batista, et al., 2006).

Hybrid functional methods

The Hybrid functional is a mixture of the non-local Hartree-Fock exchange potential and the semi-local or local DFT exchange-correlation functional and the weighting of the mixture as a percentage leads to better correlation to experimental results. It has been reported that the use of hybrid functionals when applied to semiconductors as well as insulators results in a much improved accuracy in the prediction of the band gap (Heyd, et al., 2003) (Heyd, et al., 2005). From section 1.5.3 equation 1.56 the choice of the mixing ratio α is the key to the accuracy of the results that the hybrid functional technique will return.

A more direct reference to the applicability of the success of the HSE06 functional to predict the band gap of semiconductors is the ability to correctly predict the band gap of germanium (Heyd, et al., 2005). Hybrid functionals have become increasingly more popular as their accuracy is on par with the like of quantum Monte Carlo (QMC) techniques (Batista, et al., 2006) but not as computationally expensive thus calculations can be completed in far shorter periods.

Application of hybrid functionals to defect studies is of great importance in correctly characterizing the defect structures and optimization of the defect complexes. Correctness of these defect calculations depends on whether the original formation energy was computed using LDA or GGA methods. It has already been mentioned that LDA functionals as well as GGA functionals underestimate the band gap of semiconductors (Xiao, et al., 2011) which cause large uncertainties in the calculation of defect levels in the band gap. Application of appropriate corrections outlined by Komsa (Komsa, et al., 2010), (Komsa & Pasquarello, 2011) and Deák (Deák, et al., 2010) results in a far more accurate description of the defect properties when theoretical results are compared to experimental ones. These methods, however, show there is still a shift which exists between the calculated levels and accepted levels as the gap approaches the experimental value and requires that

the defect level to be aligned for the best accuracy. The average shifts of the defect levels are usually at most 0.2 eV irrespective of material.

4.3 Defect properties

From this point the conceptual ideas of how the theory applies to achieving accurate results are put behind us and more emphasis is placed on the topics which are more meaningful to the overall accuracy of the calculations that were chosen for this investigation.

4.3.1 Formation energy

The formation energy of a defect within otherwise homogeneous semiconductor bulk is defined as the difference in the total bulk energy before and after the formation of the defect. It represents the energy required to break existing atomic bonds and distort lattice stresses imposed as a result of the presence of the defect. The energy required to add or remove atoms or electrons from the system also is a contributing factor to the total formation energy. Opposing this energy penalty is the increase in entropy, because a crystal containing a defect has more possible microstates than the pristine state. The formation of a defect in some charge state q can be obtained by ab-initio techniques and can be represented as follows according to Zhang and Northrup (Zhang & Northrup, 1991):

$$\Omega_{D,q} = E_{D,q} - E_H - \sum_i n_i \mu_i + q(E_{VBM} + \Delta E_F + \Delta V_{pa}) + E_{MP} \quad (4.4)$$

Where $E_H, E_{D,q}$ are the total energies of the pristine host state and the host state containing the defect respectively, n_i is the number of atoms of type i that have been added to ($n_i > 0$) or removed from ($n_i < 0$) the pristine host state to create the defect. μ_i are the corresponding chemical potentials for the respective atom species. ΔE_F is the position of the Fermi level energy of the semiconductor in question calculated as shown in section 4.2.1. and will vary from a value of 0 to the conduction band minimum of the semiconductor.

E_{VBM} denotes the energy of the valence band maximum of the supercell, calculated as follows:

$$E_{VBM} = \lim_{N \rightarrow \infty} [E_H(N) - E_H(N - 1)] \quad (4.5)$$

This is the difference in total energy between a pristine supercell with N valence electrons and a $+1$ positively charge identical supercell and determines a reference energy of the electron reservoir of the system (Lany & Zunger, 2009).

In order to obtain an accurate formation energy, one needs to account for the relative position of the average potential in the defect calculations as well as that of the pristine host. Thus the ΔV_{pa} term is required to correct for this change in electrostatic potential due to the addition or removal of electrons from the system. For practical calculations a reference potential is chosen at an atomic site, which is far enough removed from the defect, in order to keep potential energy variations due to the defect presence at a minimum, which aims to serve as a “potential marker”. The potential alignment term is found as follows:

$$\Delta V_{pa} = V_D - V_H \quad (4.6)$$

Which is the average difference of the reference potential in the defect supercell V_D and the pristine supercell V_H . As ΔV_{pa} is a function of electron interaction with the bulk, naturally it scales with the magnitude of added or removed electrons (Lany & Zunger, 2009).

E_{MP} (eq. 4.4) is the Makov-Payne correction term, it is defined as a 3rd order correction term which accounts for the interaction of the delocalized part of the defect-induced charge state. However not much consensus has been reached in the academic community regarding the applicability of this 3rd order term until recently, when supercells of 1728 atoms were used by (Lany & Zunger, 2009) whereby a very accurate convergence could only be found by inclusion of the Makov-Payne correction term.

The defect formation energy is one of the fundamental properties of a defect in a semiconductor that may be determined by ab-initio techniques. It is used to determine configurational defect stability of the defect and is used to obtain ionization levels to characterize the electrical nature of the defect.

4.3.1.1 Setting of boundary conditions

In order to study defects by making use of DFT a selection of boundary conditions must be made in order to compromise between accuracy of the result and computational time. These conditions

attempt to model the defect in an environment as close to as possible of the infinite host lattice. There exist here two main techniques namely the cluster approach and the supercell approach.

Cluster approach

A cluster is defined as a finite section of a semiconductor lattice whereby the surface is passivated to eliminate the possibility of bonding with foreign atomic species. Due to the proximity to the surface there exist strong defect-surface interactions. These interactions can be minimized using large clusters but this leads to far longer computational times and limitations for defect configuration studies (Jones & Mitchell, 1992).

Supercell approach

A supercell attempts to recreate the semiconductor crystal lattice in theory. It consists of a predefined volume or unit cell which is then repeated in 3D-space built from these unit cells. In the natural world a defect is then introduced into the lattice by some external means, at some part of the lattice while the remaining bulk is left in the pristine condition. But in the supercell technique, the defect is placed in the unit cell and then repeated in all periodic images of the unit cell. Emphasis is placed on ensuring that no interaction between defects in neighboring supercells takes place. The major benefit of this technique is that direct comparison between the state containing the defect and the pristine state can be made with ease (Van de Walle, 2004). However to get the best representation of the natural world, a supercell large enough needs to be chosen to correctly mirror the defect densities found in modern semiconductors. Even though 1 defect per 64 atom supercells are often used for this technique, this is far short from reality where defect concentrations are often measured in parts per million (ppm). However surprisingly accurate results can still be achieved. It is important to ensure the cell is large enough to minimize inter-defect interaction between the supercells periodic nature (Van de Walle, 2004).

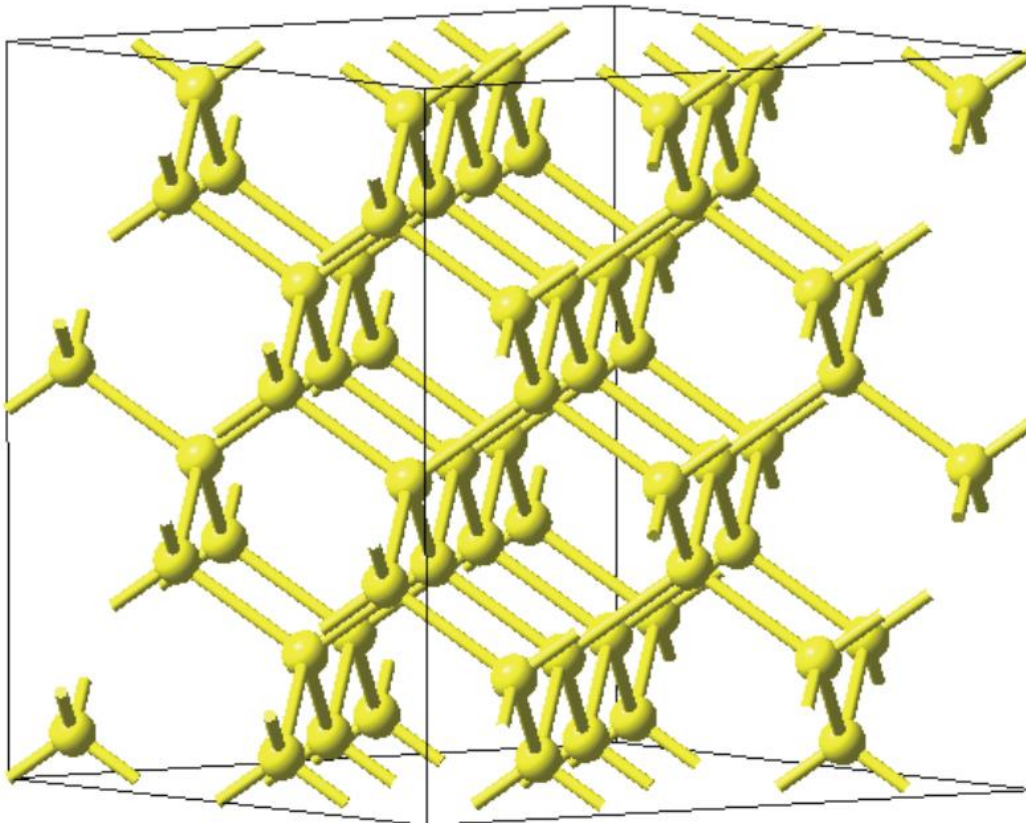


Figure 4. 1: A representation of the pristine germanium 64-atom unit cell. From this initial structure changes will be made and defects added to create the secondary defected structures used for comparative analysis.

4.3.1.2 Brillouin zone sampling

Two main methods exist for increasing the accuracy of calculated results by ab-initio techniques. These are a) increase the number of atoms of the unit cell, or b) increase the number of k-points used to sample the Brillouin zone (Puska, et al., 1998). Computationally it is more efficient to achieve greater accuracy by increasing the number of k-points rather than the number of atoms of the unit cell, yet the study of defects often still requires large cell sizes to ensure that there is no interaction between defects with neighboring cells. Thus a trade off is often used whereby k-point sampling and cell size are chosen to yield the best accuracy in an acceptable calculation time.

Γ -point sampling

Special k-point selection schemes are used to determine how the Brillouin zone is sampled. This technique allows both for faster more efficient calculations as well as more accurate ones. The k-space of the lattice needs to be represented by the correct choice of k-points used to sample the Brillouin zone and the simplest way of doing this is to center around the Γ -point. Γ -point sampling saves computational time since the wave functions are calculated in real rather than reciprocal space. However since one (1) k-point is used it often fails to take factors into account such as charge densities and delocalized states. Thus it becomes ineffective for more complicated defect analysis and often it is necessary for more k-points to be used for more complex studies (Makov & Payne, 1995).

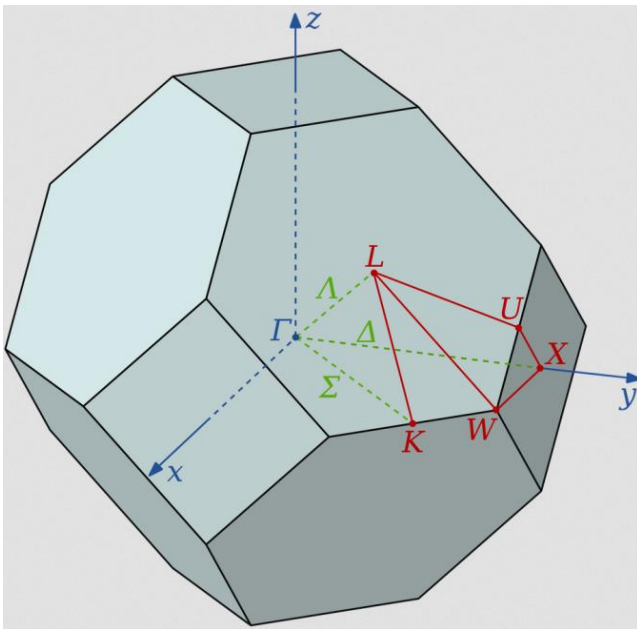


Figure 4. 2: Schematic representation of the Wigner Sietz Brillouin zone, the location of the Γ -point and the various other points. Depending on the structure under investigation more or less points should be used to sample the BZ (Freed & Moore, 2013).

4.3.1.3 Charged defect states

Defect formation energy is charge state dependent as shown in equation (4.4). For some chosen Fermi level, the stable charge state will be the one with the lowest formation energy. The Fermi level E_F for equation (4.4) will span the entire band gap calculated for the semiconductor such that $E_V < E_F < E_C$ where E_V (energy of the valence band) and E_C (energy of the conduction band) have

their usual meanings and E_F (energy of the Fermi level) refers to the top of the valence band; i.e.: $E_V = 0$. Since defects are known to be electrically active, they often affect the band structure of the semiconductor. The valence band maximum energy found from the supercell with the defect present, must relate in such a way that the effect that the defect has on the pure host band structure is taken into account. The effect the defect has can be corrected for by calculation of the valence band energy of the pure semiconductor and aligning the electrostatic potentials of the supercell with the defect and that of the pure host supercell. This effect arises due to the nature of the long-range Coulomb potential and the periodic boundary conditions imposed with the supercell approach (Van de Walle, 2004). This alignment factor and correction is represented by ΔV_{pa} in equation (4.4).

4.3.1.4 Spin orbit coupling - spin polarization

In this method the angular momentum of the spin component of an electron is combined with the angular momentum of its orbit. An electron's spin interactions with its neighboring electrons can cause shifts in the electrons atomic energy levels due to electromagnetic interaction. This interaction with nearest neighbors will affect the total entropy of a system and will result in the accuracy of the total energy resulting from our DFT model being affected. Inclusion of spin orbit coupling will allow for a better defined band structure, when modelling semiconductor's using DFT and this becomes more prevalent in situations whereby semiconductors such as germanium are known to be difficult to model due to a very small band gap (Heyd, et al., 2005).

4.3.2 Thermodynamic transition levels

A thermodynamic transition level is the position of the Fermi level where the formation energies of the same defect in differing charge states are equal.

The energy of the transition level may be calculated as:

$$\varepsilon(q/q') = \frac{\Omega_{D,q} - \Omega_{D,q'}}{q - q'} \quad (4.7)$$

The thermodynamic charge transition level is related to the activation enthalpy as determined from the experimental process of DLTS.

4.3.3 Optical transition levels

An optical transition level can be defined by calculating the energy of the final state q' based on the atomic configuration of the initial state q . The final state is not relaxed to equilibrium since the Frank-Condon principal is taken into account which states that electronic charge states can switch at far faster rates than ions can be relaxed.

4.3.4 Defect diffusion and propagation

Defect diffusion and propagation within any bulk material is largely a research field all to its own. But defect diffusion can be modeled and predicted using ab-initio techniques such as exposing a potential energy surface (PES) to a transition state search (TSS). The PES method can give information on all the stable and metastable defect configurations and the location of the saddle points, migration barriers migration path or minimum energy paths can be determined. TSS on the other hand looks more specifically at a predetermined path that can be studied in detail by first creating a set of extrapolated images between an initial and final configuration of a defect. These images are then subjected to the standard minimization techniques to relax the structure at each stage of the transition. From this an energy profile and transition map can be built to analyze the path of the diffusing. From this, local energy minima can be located as well as the exact diffusion path for interstitial or substitutive defects.

Chapter 5

Computational conditions for this study

5.1 Summary of calculations and boundary conditions.

The criteria in this chapter were used for the DFT calculation of the germanium antimony-vacancy (Sb-V). Similar conditions were used in the calculation for all other defects considered in this study.

The Vienna Ab-initio Simulation Package (VASP) has been used to study the defect complexes of interest. VASP has been used extensively around the world in the field of ab-initio DFT calculations (Kresse & Furthmuller, 1996a) (Kresse & Furthmuller, 1996b) (Kresse & Joubert, 1999). VASP is a package designed for performing ab-initio electronic structure calculations by making use of Vanderbilt pseudopotentials or the projector augmented wave method and a plane-wave basis set. The basic process makes use of density functional theory (DFT) calculating an approximate solution to the many-body Schrödinger equation using the solution of the Kohn-Sham equations yet post DFT corrections such as the latest hybrid functionals, mixing DFT and Hartree-Fock exchange (Paier, et al., 2006). Many body perturbation theory and dynamical electronic correlations are also implemented in VASP.

VASP was pioneered by Mike Payne of MIT pre-1989 and since been under continuous development by a team under the leadership of Georg Kresse at the University of Vienna with the cooperation of Jürgen Furth Müller to be notably mentioned.

5.2 Convergence criteria

The size of the supercell which is chosen as well as that of the k-point sampling scheme influences the results of the energy calculations when making use of ab-initio techniques. The effect of these parameters on the total energy results requires investigation to select the optimum choice of parameters. A compromise often needs to be made between the accuracy of the computation and the time which it will take to complete the calculation.

5.2.1 Unit cell

The unit cell is optimized from the point of view that the k-point spacing, cut-off energy and lattice constants are chosen to yield a chosen accuracy of convergence for the energy of the system. For this study the electronic energy convergence criteria was chosen to be 10^{-6} eV for the HSE06 calculation as well as for the LDA/GGA in the standard DFT mode of calculation. This criterion is much less than the typical energy difference involved with deep-level defects in semiconductors, which are in the order of 10^{-3} eV. Similar convergence criteria were used previously (Ouma & Meyer, currently under review). Structure optimization was kept at a base standard of 0.02 eV/Å for all calculations.

5.2.2 Supercell size

All supercells of 64 atoms used in this study were built from the optimized unit cell. Before the size of supercell can be determined in any calculation, consideration must be given to prevent close defect-defect interactions. Usually one standard unit cell of unaffected bulk its used in-between the defect and the edge of the supercell.

The appropriate supercell size can be determined, from a scalable set of calculations which compare the minimized formation energy of the system and the size of the supercell. Since defect calculations are computationally expensive the choice of supercell should yield acceptable results without compromising on the accuracy of the result.

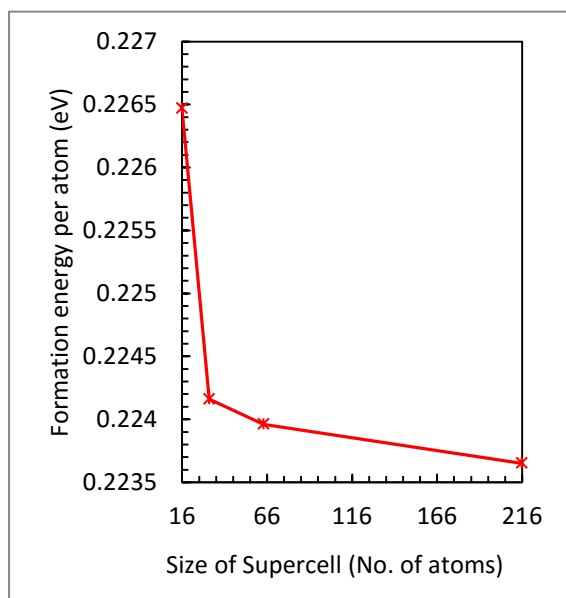


Figure 5. 1: Diagram of formation energy per atom as the size of the homogeneous germanium supercell increases. From a choice of 32 atoms onwards the formation energy difference between structures becomes small relative to the energy differences considered in the calculation.

5.2.3 Plane-wave cut-off energy

The energy cut-off parameter applies to the completeness of the plane-wave basis set. At each k -point only the plane-waves which satisfy the equation are included.

$$\frac{1}{2} |\mathbf{G} + \mathbf{k}|^2 < E_{cutoff} \quad (5.1)$$

However using a higher cut-off energy will always yield more accurate results. The length of applied computational time will exponentially increase. In the case of this parameter a trade off in calculation time to acceptable accuracy must be made.

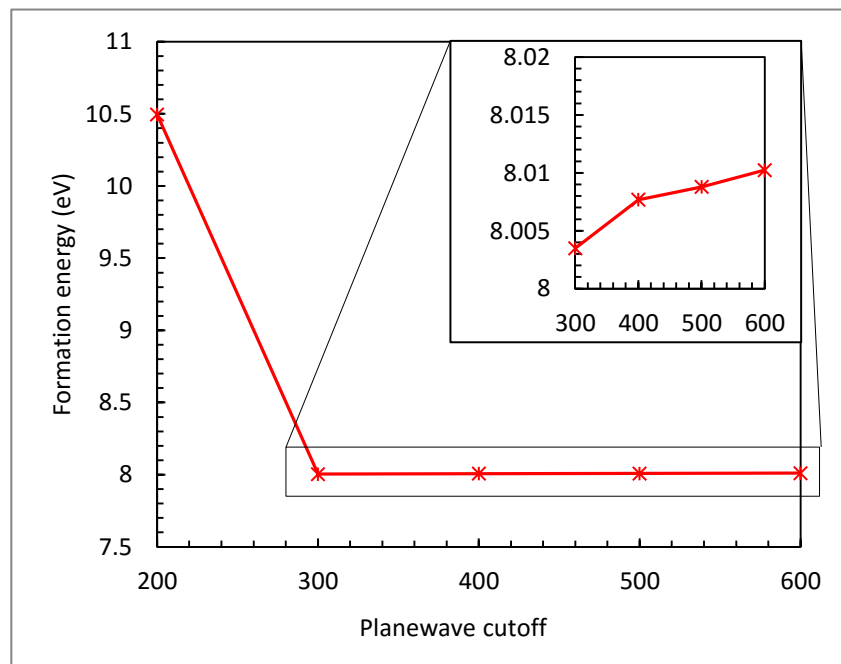


Figure 5. 2: The formation energy of an 8 atom homogeneous germanium cell as a function of the plane-wave cut-off energy. The difference in formation energy was deemed negligible after 300 eV and a plane-wave cut-off of 400 eV was chosen for all further calculations.

5.2.4 Brillouin zone sampling

The sampling of the Brillouin zone, also known as k-point sampling, influences defect formation energy when calculated by means of ab-initio techniques. From the results below, making use of a 2x2x2 Monkhorst-Pack k-point mesh is shown to be just as effective at achieving acceptable results, as using a 3x3x3 or 4x4x4 mesh with a significant reduction in expended computational time. Figure 5.3 shows that the difference in formation energy per atom for the k-point mesh of 2x2x2 or larger becomes less than 1 meV this was chosen to be acceptable as the trade off in computational time deteriorates very rapidly beyond this range.

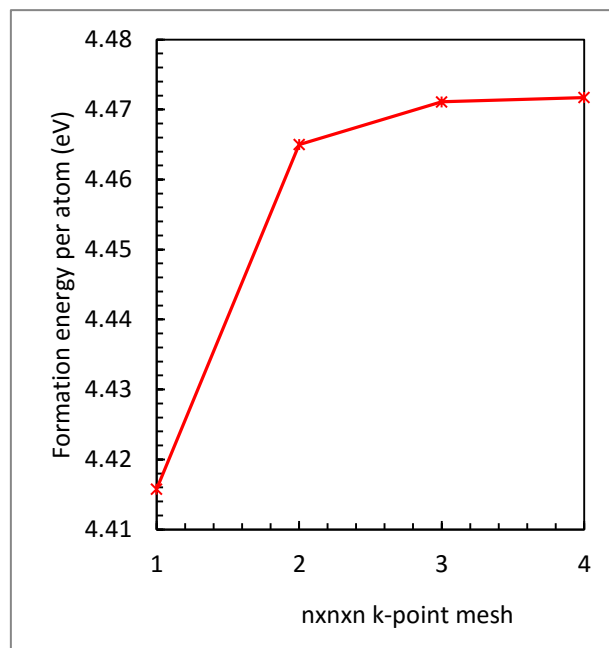


Figure 5. 3: The energy convergence of a 64 atom homogeneous germanium supercell with cut-off energy of 400 eV with regards to the k-point mesh chosen

5.3 Using DFT to predict metastability

The process on how defect metastability can be investigated using ab-initio techniques needs to be outlined. Noteworthy at this point is that defects are classified as metastable if and only if the defect has more than one minimum atomic energy configuration depending on experimentally imposed conditions. Such conditions for example would be an applied directional bias, or an applied external electric field.

5.3.1 The minimum energy configuration

Since metastable defects are stable in more than one configuration, all such configurations need to be identified. Finding the minimum energy of formation is a general problem in DFT simulations and with multiple minima, as in the situation of a metastable defect, this problem becomes more difficult as the ability to discern global minima from local minima is not a trivial task. The configurations of a defect complex can be obtained by experimentally proposed configurations (Markevich, et al., 2004), (Fage-Pedersen & Nylandsted Larsen, 2000) or from ab-initio atomic configurations, whereby a configuration is theoretically assumed, by virtue of local minima in formation energy calculations (Jones & Briddon, 1998), (Latham, et al., 1999).

This study takes on the exploratory role to investigate the possible presence of metastability of the Sb-V complex in germanium. If metastability is found, various results would have to be uncovered for a definitive conclusion as will now be explained in the remainder of this chapter. For charge state controlled metastability, the minimum energy configuration would have to depend on the charge state of the defect complex. This can be determined by calculating the formation energy for the defect configuration in each charge state and comparing the results.

Table 5.1: An example of HSE06 calculated formations energies (eV) of two different defect states. The minimum energy is shown in bold in each case.

Configuration	-1	0	1
S1	5.1	4.9	4.85
S2	5.7	5.1	4.5

From this example assume we have two configurations of a defect which are given as configuration S1 and S2, assume that S1 and S2 are some hypothetical orientation of the defect for now. The S1 configuration is the minimum energy configuration for both the -1 and neutral charge states. Yet the S2 configuration becomes the minimum energy configuration for the +1 charge state. From this result there is theoretical evidence that charge state controlled metastability exists. This result can be confirmed experimentally by examination of applied field dependence of the defect emission rate (Nielsen, 2005), as well as by observing the defect behavior by DLTS, after applied bias annealing cycles.

5.3.2 Formation energy difference and occupation

For the occupation of two defect configurations of the same defect in the same charge state to differ significantly, the difference in the respective formation energies should be approximately $k_B T$. With a difference of approximately $3k_B T$ or larger, the ratio of occupation would be 1:10 or greater as long as no degeneracy is present. This implies that the configuration with the lowest energy will be occupied and be more prevalent over the other and thus only the minimum energy configuration is likely to be experimentally observable. If the formation energy of the defects differs by much less than $3k_B T$ then both configurations will be observable simultaneously.

5.3.3 Application of electrical bias and relation to charge state controlled metastability

As mentioned in Section 3.4 the position of the Fermi level is a critical factor to note in describing the properties of both stable and metastable defects. The application of an external bias may change the Fermi level position in the region below a Schottky diode. The behavior of the Fermi level under applied bias is dependent on whether the semiconductor is p- or n-type.

In the case of the Sb-V complex which usually implies an n-type semiconductor, the Fermi level E_F in the bulk lies close to the conduction band edge E_C thus $E_F \approx E_C$ this phenomenon causes any occupied defect states in the band gap to fill with electrons (i.e. to emit holes). When a Schottky diode is fabricated a depletion region is formed. The bands close to the interface bend upwards due to the electric field in the depletion region. Application of a forward bias to the Schottky diode decreases band bending, shifting the conduction band to below the bulk Fermi level thus we have $E_C \approx E_V$ this results in defects in the depletion region with levels within the band gap; emitting electrons.

In a theoretical framework, to understand the process of a changing Fermi level on the charge state of a defect, the formation energies at each charge state are calculated for varying values of the Fermi level. Thus we examine the region of $E_V < E_F < E_C$. For example:

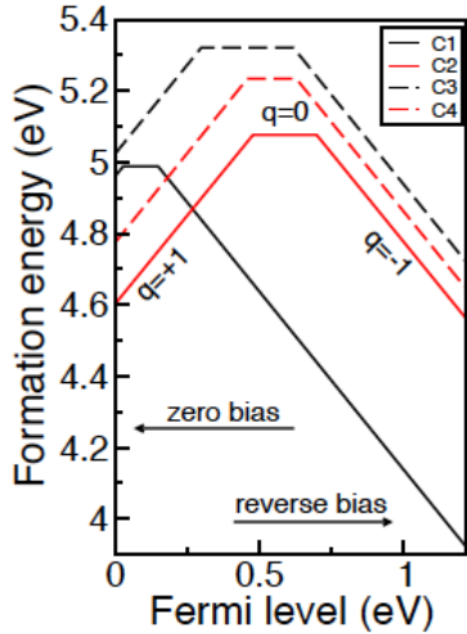


Figure 5. 4: Formation energy as a function of Fermi level for varying charge states of the boron-vacancy complex in p-type silicon. (Ouma & Meyer, currently under review).

C1, C2, C3 and C4 are the configuration designators used by (Ouma & Meyer, currently under review) to denote the configuration of the silicon-boron vacancy cluster.

As $E_F \rightarrow E_V$ the C2 configuration in the $q = +1$ charge state is the minimum energy configuration. As $E_F \rightarrow E_C$ the C1 configuration in the $q = -1$ charge state is the minimum energy configuration.

If the energy difference criteria from 5.3.2 are met whereby the respective difference in formation energy of the two defects is in the region of $k_B T$ or less and ratio of occupation is approximately 1:1 then both C1 and C2 configurations will be observed experimentally. The minimum energy configuration will be dominant if and only if the $3k_B T$ condition is met. Since the formation energy of C3 and C4 is high compared to the energies of C1 and C2 these will not be observed experimentally.

Chapter 6

The antimony-vacancy (Sb-V) complex, the *E*-center defect in germanium

6.1 The experimental results of Fage-Pederson and Larsen

Fage-Pederson and Larsen (Fage-Pedersen & Nylandsted Larsen, 2000) initially studied the irradiation-induced defects in n-type germanium in the year 2000 and several majority-carrier traps and one minority-carrier trap was characterized using DLTS. The antimony-vacancy complex (Sb-V) defect was found to anneal in a way which was different to that observed in silicon as the annealing profile in germanium was distinctly different under reverse bias, when compared to the same conditions applied to the silicon sample. Differences in the carrier concentration between the silicon and the germanium sample were measured and found to vary in the region of an order of magnitude.

6.1.1 Experimental details and results of the paper

Two types of material were used in the study and named as *Sb1* and *Sb2* with impurity concentration of 3.5×10^{14} and 1.4×10^{15} cm⁻³ respectively. Schottky diodes were fabricated onto the substrates by e-beam or resistive evaporation and defects were characterized by DLTS. An overview of the results by Larsen is shown in Table 6.2. The main result of interest in the work will be that of the E-center ($E_{0,37}$).

Table 6.1: The properties of relevant electron and hole traps observed by Larsen in the irradiation study.

Label	E_{na} (eV)	σ_{na} (cm ²)	Annealing (°C)	Identification	Occurrence
$E_{0.37}$	0.37	1.1×10^{-14}	↓150	<i>E</i> -center	<i>Sb1, Sb2</i>
$E_{0.23}$	0.23	2.0×10^{-15}	↑RT, ↓110	Sb and <i>I</i> related	<i>Sb1, Sb2</i>
$E_{0.19}$	0.19	1.5×10^{-14}	↑RT, ↓RT	Sb and <i>I</i> related	<i>Sb1, Sb2</i>
$E_{0.13}$	0.13	3.2×10^{-15}	↑RT, ↓RT	Sb and <i>I</i> related	<i>Sb1, Sb2</i>
$E_{0.21}$	0.21	7.1×10^{-14}	↑90, ↓180	Sb related?	<i>Sb1, Sb2</i>

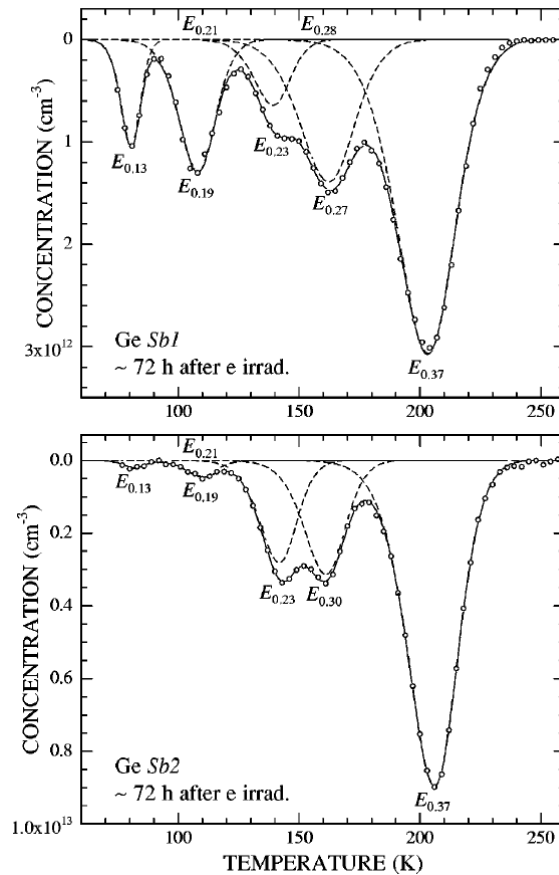


Figure 6. 1: DLTS spectra recorded on Sb1 and Sb2 three days after 4×10^{13} cm⁻² electron irradiation. (Fage-Pedersen & Nylandsted Larsen, 2000).

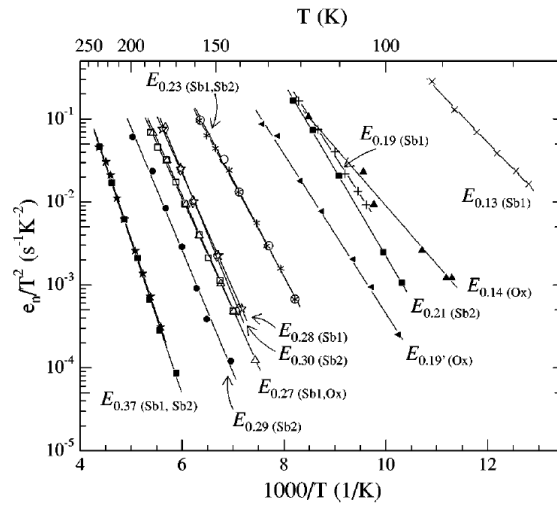


Figure 6. 2: DLTS signatures of all observed electron traps in Sb1 and Sb2 (Fage-Pedersen & Nylandsted Larsen, 2000).

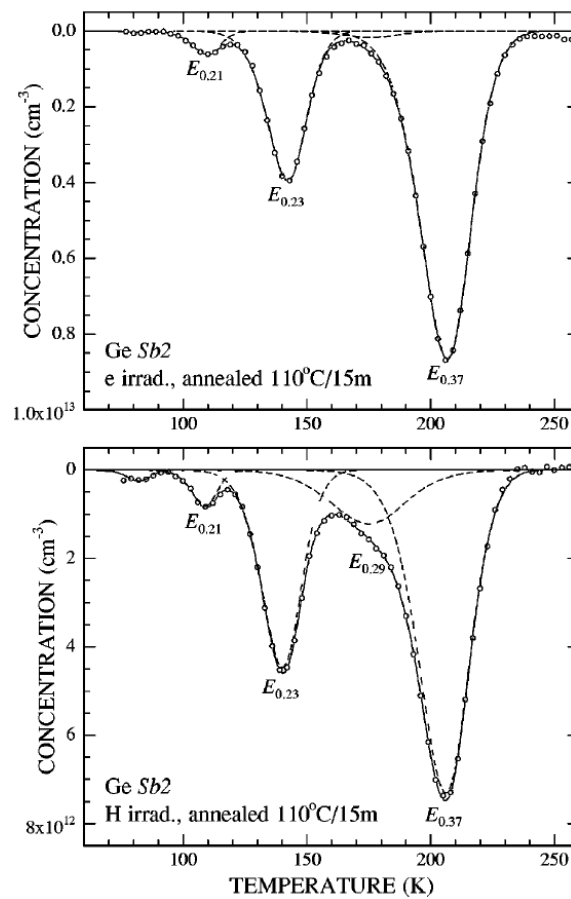


Figure 6. 3: DLTS spectra from Sb2 as a comparison of electron and proton irradiation and annealed at 110°C for 15 minutes (Fage-Pedersen & Nylandsted Larsen, 2000). Defect $E_{0.29}$ is visible only after annealing.

Figure 6.1 and 6.2 are included here in this work as an indication of the properties required to better understand the E -center complex. Defects had been introduced into the samples by irradiation of the diodes by either 2-MeV electrons or 2-MeV protons. *Sb1* was only irradiated with electrons, *Sb2* was irradiated with electrons and followed by a measurement process and then a secondary irradiation with protons. The beam intensity was $\approx 100 \text{ nAcm}^{-2}$ for the electrons and $\approx 0.5 \text{ nAcm}^{-2}$ for the protons. The interesting characteristics are that the concentration of the E -center defect is stable and does not vary greatly with regards to irradiation exposure time. However a lowering of concentration is observed after proton irradiation in *Sb2*.

6.1.2 The $E_{0.37}$ defect

The E -center was present in both materials *Sb1* and *Sb2*. The E -center was found to be the only observed defect in the sample, directly after electron irradiation. All other defects observed are secondary and the respective concentration growth of these other defects is proportional to time. Upon annealing at RT, a fraction of the $E_{0.37}$ defects disappear, yet annealing of the defect is still observable in the region of 150 °C. It was stated by Larsen that thermally activated dissociation or diffusion would not proceed over such a wide temperature span.

Larsen et al. explained in his work that the E -center peak is not a combination of contributions from other defects. There exists some mobile defect that consumes E -centers and it must be released at RT from some unstable source that is created during irradiation. The Ge self-interstitial becomes active at a much lower temperature and was ruled out as a candidate. Larsen speculates that the most likely cause is that interstitial groups are created during irradiation and a transient release of self-interstitials takes place at RT. It is seen that the E -center has a very small, temperature dependent capture cross section and this implies that the E -center has a repulsive potential. Thus $E_{0.37}$ is assumed to be a double acceptor level ($-/\equiv$) of the Sb E -center.

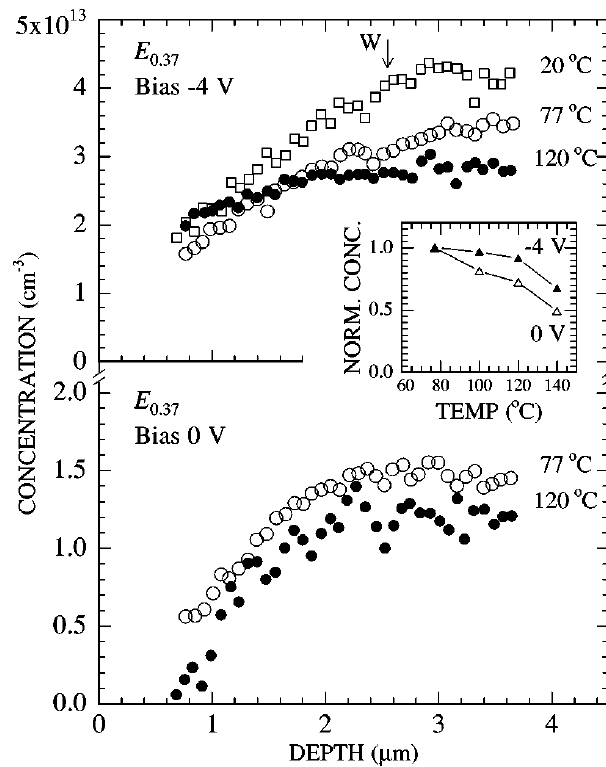


Figure 6. 4: E center depth profiles in two diodes on Sb2, which were annealed with either a 4V reverse bias or with no bias. The depletion layer width W at 4 V is shown. Diodes were irradiated with different doses. Profiles were measured 15 days post irradiation (\square), three months post 77 °C/5min annealing (\circ) and after 120 °C/10min anneal (\bullet). Results were obtained at 205 K with a DLTS double-pulse technique with a constant reverse bias of 20V with a pulse difference $\Delta V = 1.00$ V. Inserted are the annealing curves with (\blacktriangle) and without (\triangle) bias, as observed by DLTS using -4 V to 0 V applied reverse bias.

Table 6.2: Ge was doped with either the indicated elements and irradiated with the indicated proton or electron source. Each entry contains the name, apparent enthalpy (eV), apparent capture cross section (cm²) at T = ∞ and annealing behaviour (°C).

	Larsen <i>et al.</i> <i>Ref J:</i>	Bourgion and co-workers <i>Ref A, Ref G, Ref I</i>	Fukuoka and co-workers <i>Ref B, Ref C, Ref F</i>	Marie <i>Ref E</i>	Nagesh <i>Ref D</i>	Zistl <i>Ref H</i>
	Sb, O H, e	Sb, group V e	Sb, O e	Sb Pb, Ne	Sb, O, P η, γ	Sb e
<i>E</i> -center	E _{0.37} 0.37 1.1×10 ⁻¹⁴ ↓150	E ₂ 0.53 4×10 ⁻¹¹ ↓150	E(0.40) 0.4 ↓97(hs)	ET5 0.46-0.47 0.4-1.1×10 ⁻¹² ↓150	E ₄ 0.35 ↓125	ET5 0.34-0.39 ↓150
Sb and / related	E _{0.23} 0.23 2.0×10 ⁻¹⁵ ↑RT, ↓110	E ₁ 0.23 1×10 ⁻¹³ ↓110	E(0.23) 0.23 ↑70, ↓110	ET3 0.29 2.2×10 ⁻¹⁴ ↓90	E ₂ 0.17 ↑RT, ↓100	ET3 0.22-0.23 ↑RT, ↓110
Sb related?	E _{0.21} 0.21 7.1×10 ⁻¹⁴ ↑90, ↓180			ET2 0.27-0.28 2.0-8.2×10 ⁻¹² ↑80, ↓160	E ₆ (?) 0.15 ↑80, ↓170	

Ref A: (Bourgoin, et al., 1980)

Ref B: (Fukuoka & Saito, 1982)

Ref C: (Fukuoka, et al., 1983)

Ref D: (Nagesh & Farmer, 1988)

Ref E: (Marie, et al., 1993)

Ref F: (Fukuoka & Saito, 1981)

Ref G: (Mooney, et al., 1983)

Ref H: (Zistl, 1997)

Ref I: (Poulin & Bourgion, 1982)

Ref J: (Fage-Pedersen & Nylandsted Larsen, 2000)

Possible evidence of this idea is seen in annealing experiments. From the E -center profiles in Figure 6.4 the application of a 4 V reverse bias in the inserted graph was shown to impede annealing of the E -center in Ge, as the concentration of the defect remained noticeably higher than the scenario when there was no applied bias. This impediment was also observed in the E -center annealing of silicon.

It was found from the slope of the profile in Figure 6.4 that a number of vacancies from the irradiation-generated Frenkel pairs had diffused to the surface before they were trapped by Sb atoms. Comparison of the application of the reverse bias shows the reduction of the role of the surface as a sink for E -center annealing. At an annealing temperature of 120 °C the Fermi level is 0.08 eV below the E -center.

Two possible scenarios were proposed by the authors to explain the observed depth profiles of the E -center compatible with the idea that the E -center is a double acceptor ($-/=$).

Firstly it is proposed that the E -center anneals by diffusion. Migration of E -centers toward the surface will at some point, result in the defect being driven back into the bulk as a result of strengthening electric fields closer to the surface. Thus only very close to the surface is the E -center concentration expected to decrease.

Secondly they proposed that the E -center anneals by dissociation and the vacancy defect is then lost to the surface. Dissociation rates should therefore be low when a reverse bias is applied, meaning that the ($-$) state would be more stable than ($=$). This stabilization should be stronger than the opposite tendency to separate the positive Sb^+ component from the negative vacancy component. By extension the mobility of the vacancy could be reduced within the depletion region.

Experimentally, charge states of defects can be indicated by analysis of the DLTS capture cross section, the true capture cross section as well as the capture barrier height and dependence of the emission rate on the electric field.

Field effect emission measurements can also give details on the charge state of defects. The Poole-Frenkel effect is applicable in this circumstance.

6.1.3 Summary of findings of Larsen

Irradiation-induced impurity point defect complexes were investigated in antimony doped n-type germanium crystals with characterization of the majority-carrier traps by means of deep-level transient spectroscopy (DLTS). An annealing study into the E -center ($E_{0.37}$) defect had found that the way in which this defect annealed was fundamentally different from that in silicon since it shows unexpected behavior when under reverse bias. It was speculated in conclusion that $E_{0.37}$ is the double-acceptor level of the E -center.

Due to the unusual annealing properties and the metastable properties of the structurally similar B-V center in Si (Fage-Pedersen & Nylandsted Larsen, 2000), it was decided to study the E -center in Ge theoretically.

6.2 The modelling results

Modelling of the E -center defect was done by using the methods outlined in Chapter 5. Firstly the formation energies of the vacancy defect (V_{Ge}) were calculated in the charge states -2 to $+2$. Secondly, the formation energies of the antimony substitutional defect (Sb_{Ge}) which is known to be a shallow level donor in germanium were calculated. When these two defects interact they form the Sb-V complex. The results for the Sb-V complex defect are presented at the end of this chapter.

In all cases the atomic positions were relaxed but the volume of the supercell was not allowed to vary from that of the pristine cell to best model the conditions of a cell present in a bulk lattice. The reason for this choice is because in a bulk lattice the supercell cannot change its volume much, since it is embedded in a rigid host lattice. Thus, the cell's shape and volume were kept fixed. Finally, the defect formation energy as a function of the Fermi level, at various charge states for each defect, is shown at the end of each sub-section.

6.2.1 The V_{Ge} defect

Here the formation energies of the germanium vacancy defect are presented as obtained for each charge state, by making use of the HSE06 hybrid functional and are compared to that of known published results.

Table 6.3: Calculated standard energies of formation (eV) with the Fermi level at the VBM of the germanium vacancy defect in each of the explored charge states obtained by use of the HSE06 functional.

	Charge states				
	+2	+1	0	-1	-2
This work	3.16	3.02	3.03	3.16	3.62
Normalized relative to lowest energy state	0.14	0	0.01	0.14	0.6
(Spiewak, et al., 2011)	4.03	3.38	2.87	3.34	3.98

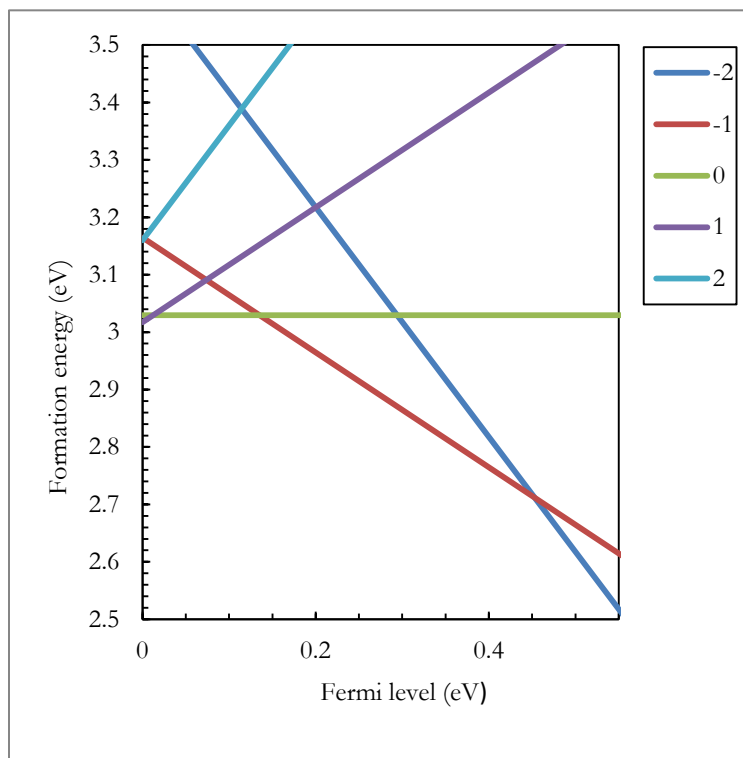


Figure 6. 5: Formation energy as a function of the Fermi level (eV) for various charge states for the vacancy defect in bulk germanium.

Table 6.4: Thermodynamic charge state transition levels (eV) relative to the valence band maximum (VBM) using HSE06 functionals for the vacancy defect in germanium.

Transition level	-2/-1	-1/0	0/+1	+1/+2
Present work	0.45	0.14	0.01	-0.14 (not in bandgap)
(Weber, et al., 2013)	0.38	0.16	0.15	0.14
Experimental (Mesli, et al., 2008)	0.44 From CBM	Only one level measured	Only one level measured	Only one level measured

The result obtained for the germanium vacancy defect is discussed in section 6.3.1.

6.2.2 The Sb-substitutional defect

Here the formation energies of the antimony substitutional defect in germanium are presented for each charge state by making use of the HSE06 hybrid functional.

Table 6.5: Calculated formation energies (eV) of the SbGe defect in germanium relative to the pristine bulk system of germanium in each of the explored charge states obtained by use of the HSE06 functional.

Formation energy	Charge state				
Charge state	2	1	0	-1	-2
Normalized (relative to lowest energy state)	0	0.21	1.47	2.83	3.97

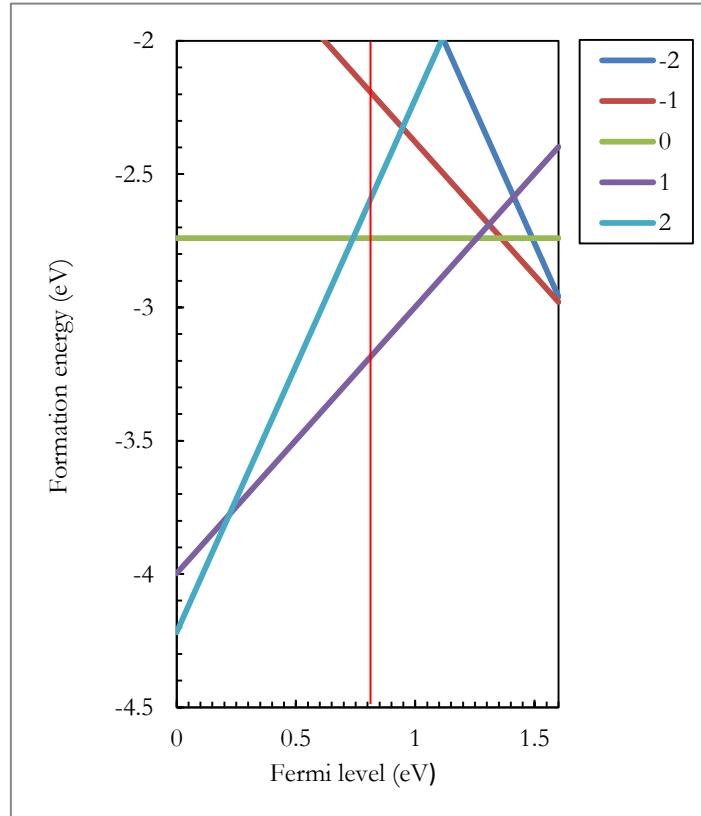


Figure 6. 6: Standard energy of formation as a function of the Fermi Level (eV) for the Sb substitutional defect in various charge states (The red line indicating the level of the CBM). Most transition levels lie with the conduction band and not within the band gap.

Table 6.6: Thermodynamic charge state transition levels (eV) relative to the valence band maximum (VBM) using HSE06 functionals for the antimony substitutional defect in germanium.

Transition level	-2/-1	-1/0	0/+1	+1/+2
Present work	1.62 (outside bandgap)	1.36 (outside bandgap)	1.26 (outside bandgap)	0.22

The result obtained for the antimony substitutional defect is discussed in section 6.3.2.

6.2.3 The Sb-V complex

In order to model the Sb-V complex, a number of configurations were investigated to better understand how the two constituent defects may interact. These configurations will be named C1, C2 and C3, whereby the constituent defects are at 1st, 2nd and 3rd nearest neighbour positions.

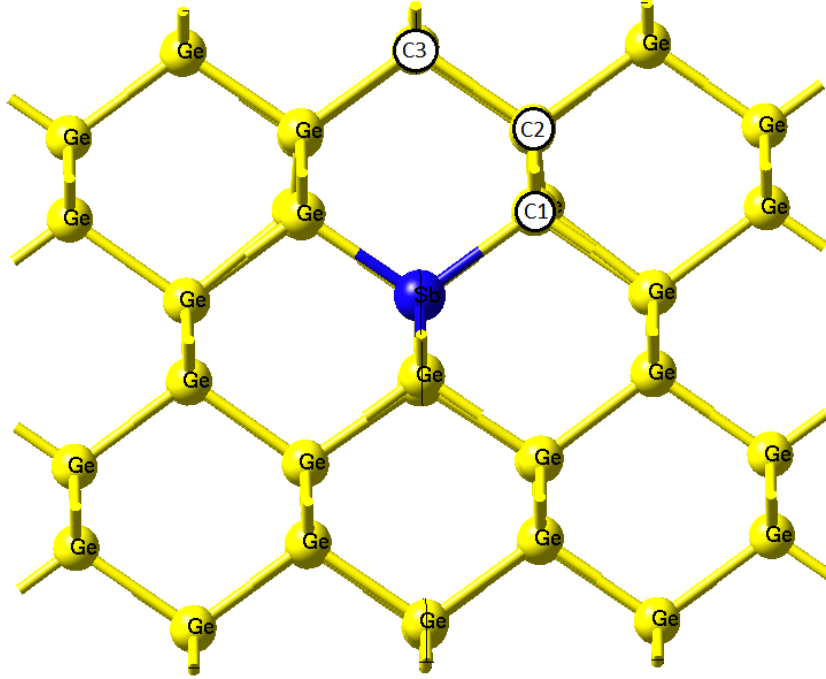


Figure 6. 7: Figure showing the defect configurations investigated for the V-Sb defect in germanium

The formation energies for the C1 through C3 configurations (defects) were calculated as outlined in sections 4.3.1 whereby the defect complex in a particular charge state q was calculated by application to the following formula:

$$\Omega_{D,q} = E_{D,q} - E_H - \sum_i n_i \mu_i + q(E_{VBM} + \Delta E_F + \Delta V_{pa}) + E_{MP} \quad (6.1)$$

All terms have the respective meanings as outlined in section 4.3.1 with E_{MP} being the Makov-Payne correction, which takes into account the electrostatic interactions resulting from the periodicity of the defect lattice.

Table 6.7: Calculated formation energies (eV) of the Sb-V complex in germanium in each of the explored charge states obtained by use of the HSE06 functional, with respect to the configuration of the defect. The energy state have been normalized relative to the lowest energy state for each charge state.

Formation energy		Charge states				
Configuration	-2	-1	0	1	2	
C1	0	0	0	0	0	
C2	0.58	0.51	0.48	0.61	0.53	
C3	0.71	0.65	0.62	0.74	0.65	

The binding energy of a defect complex is defined as the energy needed to dissociate the defect complex into its constituent parts. The binding energy of the Sb-V complex was found using the following formulation:

$$E_B = \Omega_{Sb} + \Omega_V - \Omega_{SbV} \quad (6.2)$$

Where Ω_{Sb} , Ω_V and Ω_{SbV} are the respective formation energies of the antimony substitutional defect, the vacancy defect and the Sb-V complex in germanium.

Table 6.8: The binding energies as calculated by making use of eq (6.2) for the Sb-V complex in germanium.

Configuration	Binding energy (eV)
C1	1.5
C2	1.02
C3	0.88

As seen in Table 6.8 the binding energies of the defect complex in all configurations and charge states are positive, indicating that the defect complex forms from its constituent point defects. From this we can determine that the complex forms spontaneously and is a stable bound defect complex. Also, the binding energy increases as the vacancy moves closer to the antimony.

Table 6.9: Thermodynamic charge state transition levels (eV) relative to the valence band maximum (VBM) using HSE06 functionals for the Sb-V complex.

Configuration	-2/-1	-1/0	0/+1	1/2
C1	0.52	0.4*	0.44*	0.02
C2	0.59	0.45	0.31	0.1
C3	0.58	0.43	0.31	0.1

*Negative-U: -1/+2 $U = -0.04$ eV

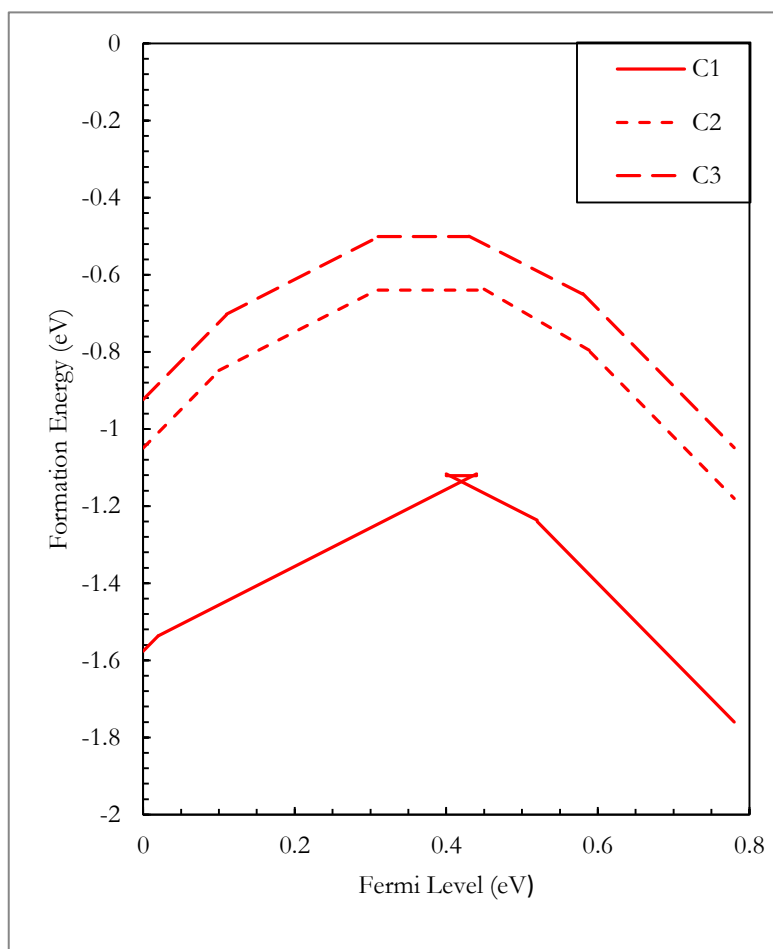


Figure 6. 8: Formation energy of the Sb-V complex as a function of the Fermi Level (eV) at various charge states for the three defined configurations.

6.3 Discussion of the Sb-V complex in germanium

6.3.1 The vacancy defect

The structure of the vacancy defect can be viewed in a different way to the traditional understanding of an empty void in the lattice. It can be also seen as a collection of strongly interacting dangling bonds which give rise to a single symmetric deep state and several degenerate states in the band gap (Weber, et al., 2013). The occupancy of these states will affect the formation energy of the vacancy and thus the interaction forces can be influenced due to the charge state of the defect.

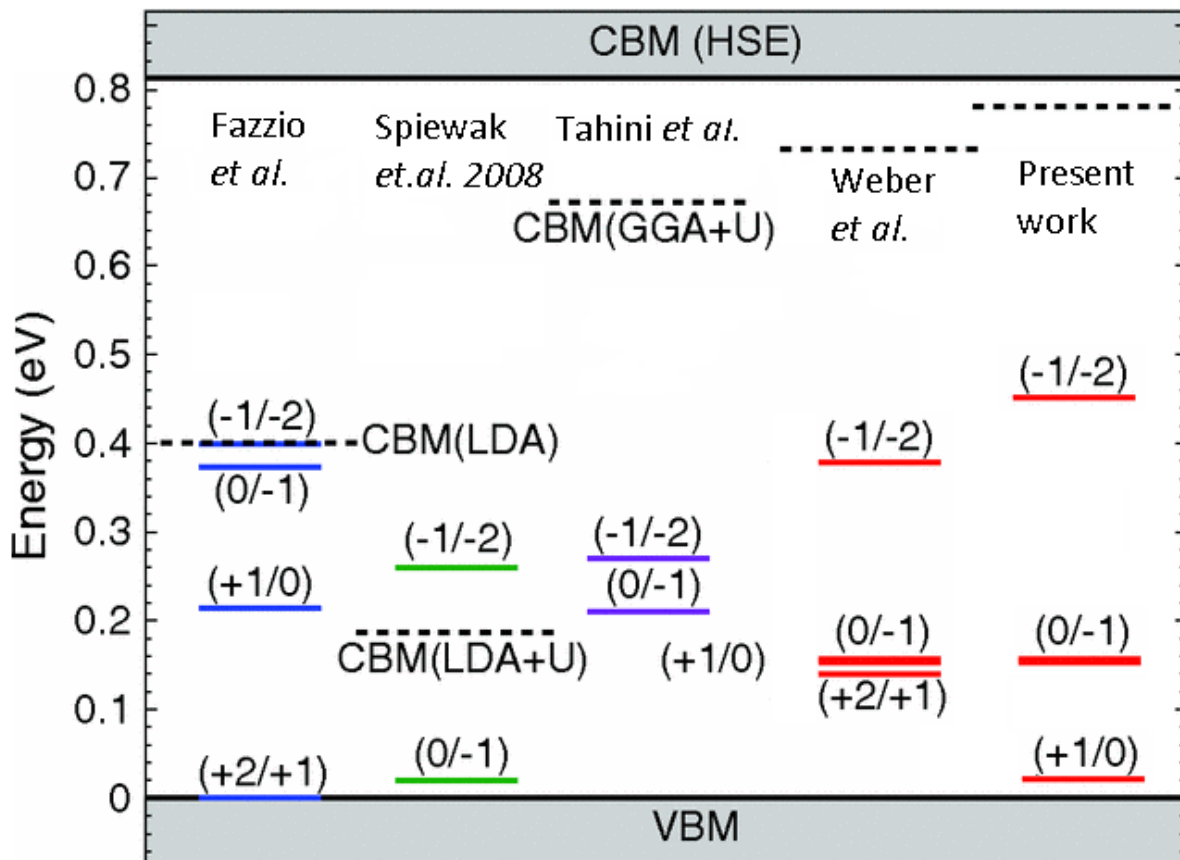


Figure 6.9: Comparison of the defect level found in different studies of the germanium vacancy. Dashed lines represent the level of the CBM in the respective calculations

Whereas Fazzio et al. (2000), used LDA with modifications to the germanium pseudopotential in order to correct the band gap calculated, this work used hybrid functionals with spin-polarization.

Spiewak et al. (2008) used the LDA+U approach to correct the band gap yet even with corrections the band gap remained underestimated. Similar positions are found for the (0/-1) and the (-1/-2) transitions in the Spiewak and Weber works and this is attributed to the interorbital repulsion and by making use of spin-polarized calculations.

Tahini et al. (2011) made use of a GGA+U functional that almost perfectly modelled the bandgap of germanium, yet their calculation lacked spin-polarization and failed to locate the transition of the (+2/+1) and the (+1/0) states in the band gap. Weber et al. (2013) makes several points as to the validity of the choices of the magnitude U, yet no real consensus is reached in his work on this point. It appears vital that more than merely accurately predicting the value of the bandgap is required to correctly determine the transition levels.

The most direct comparison from these results can be made with the work of Weber et al. (2013) since both made use of the HSE06 hybrid pseudopotential as well as spin-polarization. The correspondence between the results is, however, not as good as might be expected.

Weber et al. (2013) made use of a 216 atom supercell, whereas this study used only a 64 atom supercell. An issue could arise in the non-uniformity of defect to defect interactions whereby the 216 atom cell is 1x2x2 rectangular cell and the defect to defect interaction in the vertical plane can be different to the interaction in the horizontal plane as the model is being calculated, which can lead to some unintended consequences. On the other hand, the defect is not symmetrical either, so the asymmetry of the supercell might not be that much of a problem. The consequences of this have yet to be fully determined as the symmetry of the unit cell was not considered in the preliminary convergence studies as a possible limiting factor.

A second difference is that Weber et al. (2013) made use of Γ -point sampling rather than a k-point mesh. Γ -point sampling has been found to be insufficient to represent the whole Brillouin zone and the total energy of formation is significantly different from that calculated using different meshes (Shim, et al., 2005). However the method offers additional savings on computational resources as the wave functions used are purely real. The method is still used in many calculations based on the thinking that systematic errors in formation energies will tend to cancel out if the supercell sizes are

the same. But this is not always the case, as the errors do not always cancel. An example of such a case is when Γ -point sampling used to calculate the silicon vacancy and interstitial defects when taking the energy difference between the pristine bulk lattice and the supercell containing the defect (Puska, et al., 1998).

Having said this, the transition for the (0/-1) state is identical to that of Weber et al. and the only considerable differences is that our method results in a much lower position of the (+1/0) transition and the falling away of the (+2/+1) state. In comparison, all referenced works aside from Fazzio et al. have failed to determine this transition to lie within the band gap as well. It is determined by Weber et al. that the behaviour of the (+2/+1) and the (+1/0) transition levels has more to do with the repulsive and attractive forces created by the dangling bonds within the vacancy and thus are seen to be charge state transitions.

The formation energies of the vacancy defect were calculated and compared to later work of Spiewak et al. shown in Table 6.3 using hybrid functionals. The results obtained here agree with the results shown by Spiewak et al. Noteworthy however, Spiewak et al. did not include spin-orbit coupling and used a reduced plane-wave cut-off of 260 eV as apposed to the 400 eV used in this work. The size of the supercell and the k-point mesh used in both studies was the same. The lack of agreement in the results can be seen more clearly in the charge states of the defect and this is an indication that the spin orbit coupling does indeed play a significant role in the final result in each case.

6.3.2 The substitutional antimony defect

The results presented here are now based on the technique established by the comparison of the results of the vacancy defect to that of other published work. To our knowledge, there exists no LDA or GGA functional research applied to the Sb_{Ge} defect and even less in the way of hybrid functional studies into the matter. From this point on it is important to note that spin-orbit coupling was no longer included in the results as it became too computationally expensive for reasonable computational times.

The normalized formation energies of the antimony substitutional defect are summarised in Table 6.5. The results presented in Table 6.6 indicate that all but one transition level, namely the (+2/+1) lie outside of the band gap within the conduction band. The results do, however, correctly predict

that the +1 charge state should be the most stable state when the Fermi level is in the upper half of the band gap. The main effect influencing the results is that the supercell used is too small to accurately model the defect.

Antimony is a shallow level donor, which is typically modelled not by DFT but rather using the hydrogenic model, where the electron mass and vacuum permittivity in the Bohr model is replaced by the effective mass of the charges in the semiconductor and the semiconductor permittivity respectively. This leads to a large Bohr radius of several hundred angstrom and shallow energy levels. The large Bohr radius implies that the electron wave function is spread out over many unit cells and that modelling such a system in a small supercell will lead to large errors due to interaction between the donor and the periodic images.

It is therefore expected that the formation energy of the charge states is significantly over-estimated and that, as the supercell size increases, the energy due to the +1 charge state should drop.

The absence of electron spin could also have lead to additional errors. The results obtained were difficult to conclude as little comparison exists. Apart from our initial computational conditions it was seen that spin orbit coupling is required to ensure better accuracy. A clarification into the value of the chemical potential of antimony under these conditions is also required for accurate calculation of the formation energy term of substituted atomic species. This section specifically is now a topic of further work.

6.3.3 The Sb-V complex

The formation energies of the Sb-V complex are shown in Table 6.7 for the neutral state and four charge states. The binding energies were calculated according to Equation (6.2) and are shown in Table 6.8 and finally, the charge state transition levels are shown for each configuration of the defect in Table 6.9.

The binding energy of the defect complex is the energy required to dissociate the defect complex into its constituent point defect. The binding energy of the antimony-vacancy complex was found to be positive for all configurations. This result shows that the complex has formed its constituent parts and is defined as a stable bound defect which implies that once formed, it requires externally applied energy to dissociate it. The formation energies and charge state transition levels were found to be dependent on the position of the vacancy relative to the antimony substitutional defect, thus

configurationally dependent. The C1 configuration was found to be the lowest energy configuration in all charge states, followed by the C2 configuration and finally the C3 configuration as the highest energy configuration. The evidence from the formation energies shows that configuration C1 will form spontaneously from configuration C2 and C3. It can be seen from this result that charge state controlled metastability is not evident from these results obtained, as the evidence of apparent charge state controlled metastability depicted and explained in section 5.3.1 and 5.3.3 is not shown here.

In experimental terms, for the occupation of two distinct defect configurations in the same charge state to differ significantly, there is required to be a difference in the formation energies of both configurations of approximately kT . When this difference in formation energy is in the order of $3kT$ (≈ 0.06 eV at 215 K) or larger the occupation ratio between the two configurations would be in the region of 1:10 or larger. Thus implying one configuration dominates.

By making use of Figure 6.8 we can observe the energy of the defect complex in various configurations as a function of the Fermi level. The minimum energy configuration when the defect is under zero bias ($E_F = E_V$) is C1 and this remains to be the minimum energy configuration throughout the band gap. At no point does configuration C2 or C3 become the minimum energy configuration as we move through the band gap. As explained in section 3.4.1, the defect in configuration C1 shows some negative-U behaviour for the $+/0$ and $0/-$ transition levels. This effect is small, however still detectable which has been in the past linked to metastable behaviour in defects. The total value of the negative-U effect was found to be -0.04 eV.

Under zero applied bias conditions, the difference in formation energy between configurations C1 and C2 is 0.53 eV, which is far greater than $3kT$ at 215 K. This implies that the majority of defects will be in configuration C1. This is in agreement with Larsen et al. (Fage-Pedersen & Nylandsted Larsen, 2000).

The same scenario is observed under reverse bias conditions whereby the difference in configuration C1 and C2 at the conduction band is 0.58 eV which is again significantly larger than $3kT$. This implies an even larger concentration of defects in the C1 configuration. The transitions would now appear to be deeper in the band gap (closer to the VBM) than that of the previous applied conditions. The $(+2/+1)$ transition is now lying 0.02 eV from the VBM when under reverse applied

bias and thus is not likely to be observable by DLTS. Yet the three defects will still be detectable with $E_{0.37}$ having the largest concentration which is in agreement with Larsen.

As mentioned earlier, the $E_{0.37}$ had a DLTS signature of $E = 0.37$ eV and based on its small capture cross section the level was assumed to correspond to the -1/-2 transition. However the capture cross sections observed by others, vary widely, so this assignment may be incorrect. Further measurements by Auret et al. (to be published) seem to confirm the small capture cross section and also indicate the presence of a capture barrier.

It was found by Markevich et al. (2004) in an experimental DLTS study into the Sb-V complex, that the complex has 4 charge states, namely; 1) the doubly negative 2) the single negative 3) neutral and 4) positive state, with three corresponding energy levels in the gap. Thus three transitions would be found within the band gap. In comparison 4 transitions were found by use of DFT but the fourth transition, the $E(0/+1)$ level, lies just within the band gap. This can be attributed to a number of considerations already discussed as to the deviations from known experimental theory. The $E(=/-)$ was confirmed to be an electron trap at 0.37 eV and the $E(-/0)$ and $E(0/+)$ levels hole traps with energies 0.307 eV and 0.30 eV by Markevich. For the electron trap the position of the double acceptor was found to be temperature dependent but at 0 K and was calculated to be 0.386 eV from the conduction band minimum. The position predicted for the double acceptor level in this study was found to be 0.26 eV from the conduction band minimum. Under equilibrium conditions half occupancy of the double acceptor level occurred in the range of $E_c - (0.19 - 0.21)$ eV according to Markevich, whilst it was determined by this study that this level theoretically would be placed in the range of $E_c - (0.26 - 0.38)$ eV.

Chapter 7

Conclusions

We have investigated the properties of the Sb-V complex in germanium, from the formation energies of the constituent parts to their binding energies in various configurations, using the Heyd-Scuseria-Ernzerhof (HSE06) hybrid functional within DFT and compared the results where possible to both theoretically and experimentally published work.

With respect to the results for vacancy in germanium defect, it was shown that even though differences were obtained to the work of Weber et al. (Weber, et al., 2013) these differences were accounted for in the choices of initial computational conditions. The results obtained did agree well with the work of Weber et al. in the $(-2/-1)$ and the $(-1/0)$ charge states. The $(0/+1)$ state was found in this study to be positioned much deeper in the band gap than that found by Weber et al. and the $(+1/+2)$ transition was found to lie outside the band gap, inside the valence band by this study. The formation energies of the vacancy defect were found to match very well with that of more recent work by Spiewak et al. (Spiewak, et al., 2011) whereby hybrid functionals were used to determine the formation energy of the vacancy in germanium.

With respect to the antimony substitutional defect germanium defect, the nature of the defect being a shallow level donor implies that DFT is going to struggle to yield accurate results unless very large supercells are used. The +1 charge state was predicted to be the most stable configuration when the Fermi level lies closer to the CBM.

With respect to the antimony-vacancy complex (E-center) defect in germanium, the possible existence of metastability was explored by making use of the highly accurate hybrid functional technique within the VASP code. Experimental results hinted at possible metastability, yet no theoretical grounds for the metastability was found. These findings were consistent with the current experimental model of the Sb-V complex in germanium whereby no metastability has been observed experimentally. Yet consistencies were observed with the experimental results of Larsen and the

concentrations of the Sb- related defects in germanium detected by DLTS processes confirm the validity of our results.

However, the negative- U behavior of the C1 configuration suggests that further theoretical and experimental work should be done.

Suggested future work to expand upon this study:

1. Further investigation into the negative- U behavior of the C1 configuration.
2. Extending the modelling of germanium to include the interactions of its ten 3d orbital valence electrons, from the standard 4 valence electrons in the 4s and 4p orbitals shells, coupled with spin orbit coupling applied to the Sb-V complex will further increase the accuracy of the DFT process. Yet this increase from 4 valence electrons to 14 valence electrons has in trials done so far increased the computational time from several days to several months per calculation and to date had yet to yield reportable results.
3. Performing transition state searches (TSS) on the C1 to C2 and C2 to C3 configurations to accurately determine the dissociation and migration energies of the defect.

The results for all further purposes eliminate the E -center as a source of clearly observable charge state controlled metastability, yet the existence of the negative- U behavior implies further work is required. It allows us to also consider turning the search towards interstitials and interstitial clusters, which will form part of future a future study..

Works Cited

Alpha, W., 2006. *Wikipedia*. [Online]

Available at:

https://en.wikipedia.org/wiki/Pseudopotential#/media/File:Sketch_Pseudopotentials.png

[Accessed 2015].

Anderson, P. W., 1975. Model for the electronic structure of amorphous semiconductors. *Phys. Rev. Lett.*, Volume 34, p. 953.

Auret, F. D., Coelho, S. M. & Janse van Rensburg, P. J., 2008. Electrical characterization of defects introduced during metallization processes in n-type germanium. *Mat. Sci. in Semiconductor Processing*, 11(5-6), p. 348–353.

Batista, E. et al., 2006. Comparison of screened hybrid density functional theory to diffusion Monte Carlo in calculations of total energies of silicon phases and defects. *Phys. Rev. B.*, 74(12), p. 121102(R).

Becke, A. D., 1988. Density-Functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A.*, 38(6), p. 3098.

Bobrow, L., 1996. *Fundamentals of Electrical Engineering*. 2nd ed. New York: Oxford UP.

Born, M. & Oppenheimer, J. R., 1927. On the Quantum theory of Molecules. *Ann. Physik*, Volume 84, p. 457.

Bourgoin, J. C., Mooney, P. M. & Poulin, F., 1980. *Defect and radiation effects in semiconductors*. 1st ed. London: Institute of Physics.

Ceperley, D. M. & Alder, B. J., 1980. Ground State of the Electron gas by a Stochastic Model. *Phys. Rev. Lett.*, Volume 45, p. 9982.

Chantre, A., 1985. Configurationally bistable C center in quenched Si:B Possibility of a boron-vacancy pair. *Phys. Rev. B.*, 32(460-3).

Chantre, A., 1989. Introduction to defect bistability. *Applied Physics A*, 48(1), pp. 3-9.

Coelho, S. M., 2014. Electrical characterization of process induced defects in germanium. *University of Pretoria, PHD*.

Coelho, S. M., Auret, F. D. & Janse van Rensburg, P. J., 2013. Electrical characterization of defects introduced in n-Ge during electron beam deposition or exposure. *J. Appl. Phys.*, 114(17), p. 173708.

Dal-Pino, A., 1993. Ab-initio investigation of carbon-related defects in Silicon. *Phys. Rev. B.*, 47(19), p. 12554.

Deák, P. et al., 2010. Accurate defect levels obtained from the HSE06 range-separated hybrid functional. *Phys. Rev. B.*, 81(15), p. 153203.

Diebold, A. C., 2001. *Handbook of Silicon Semiconductor Metrology*. 1st ed. s.l.:CRC Press.

Dirac, P., 1930. *Note on Exchange Phenomena in the Thomas Atom*. s.l., s.n.

Dobaczewski, L., Kaczor, P., Hawkins, I. D. & Peaker, A. R., 1994. Laplace transform deep-level transient spectroscopic studies of defects in semiconductors. *J. Appl. Phys.*, Volume 76, pp. 194-198.

Fage-Pedersen, J. & Nylandsted Larsen, A., 2000. Irradiation-induced defects in Ge studied by transient spectroscopies. *Phys. Rev. B*, 62(15), p. 10116.

Fakuoka, N. & Saito, H., 1982. Defect States in n-Type Germanium Irradiated with 1.5 MeV Electrons. *Jpn. J. Appl. Phys.*, 21(6), p. 930.

Fazio, A., Janotti, A., de Silva, A. J. & Mota, R., 2000. Microscopic picture of the single vacancy in germanium. *Phys. Rev. B*, Volume 61, p. R2401.

Freed, D. S. & Moore, G. W., 2013. Twisted Equivariant Matter. *Annales Henri Poincare*, 14(8), pp. 1927-2023.

Frenkel, J., 1938. On pre-breakdown phenomena in insulators and electronic semi-conductors. *Phys. Rev.*, 54(8), pp. 647-648.

Fukuoka, N. & Saito, H., 1981. Radiation Defects in n-Type Germanium Studied by Deep Level Transient Spectroscopy. *Jpn. J. Appl. Phys.*, 20(7), p. L519.

Fukuoka, N., Saito, H. & Kambe, K., 1983. *Jpn. J. Appl. Phys.*, Volume 22, p. L353.

- Griffiths, D. J., 2005. *Introduction to Quantum Mechanics*. 2nd ed. s.l.:Pearson Education.
- Haas, P., Tran, F. & Blaha, P., 2009. Calculation of the lattice constant of solids with semilocal functionals. *Phys. Rev. B*, 79(8), p. 085104.
- Heyd, J., Peralta, J. E., Scuseria, G. E. & Martin, R. L., 2005. Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. *J. Phys. Chem.*, Volume 123, p. 174101.
- Heyd, J., Scuseria, G. E. & Ernzerhof, M., 2003. Hybrid Functionals based on screened Coulomb potential. *J. Chem. Phys.*, 118(18), p. 8207.
- Hichliffe, A., 2000. *Modelling Molecular Systems*. 2nd ed. West Sussex: John Wiley & Sons Ltd.
- Hohenberg, P. & Kohn, W., 1964a. Inhomogeneous electron gas. *Phys. Rev. B.*, 136(3B), p. 864.
- Hohenberg, P. & Kohn, W., 1964b. Self consistent equations including exchange and correlation effects. *Phys Rev*, 140(4A), p. A1133.
- Jones, R. & Briddon, P. R., 1998. The Ab-initio Cluster Method and the Dynamics of Defects in Semiconductors. In: *Identification of Defects in Semiconductors*. Boston: Academic Press.
- Jones, R. & Mitchell, P. C., 1992. Ab-initio Cluster Calculations of Defects in Solids. *Appl. Phys. Eng. Sci.*, 341(1661), p. 351–60.
- Komsa, H.-P., Broqvist, P. & Pasquarello, A., 2010. Alignment of defect levels and band edges through hybrid functionals: Effect of screening in the exchange term. *Phys. Rev. B*, 81(20), p. 205118.
- Komsa, H.-P. & Pasquarello, A., 2011. Assessing the accuracy of hybrid functionals in the determination of defect levels: Application to the As antisite in GaAs. *Phys. Rev. B*, 84(7), p. 075207.
- Kresse, G. & Furthmuller, J., 1996a. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Material. Sci.*, 6(1), pp. 15-50.
- Kresse, G. & Furthmuller, J., 1996b. Efficient iterative schemes for ab-initio total-energy calculations using a plane-wave basis set. *Physical Review B*, 54(16), p. 1169.

- Kresse, G. & Joubert, D., 1999. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3), p. 1758.
- Lang, D. V., 1974. Deep-level transient spectroscopy: A new method to characterize traps in semiconductors. *J. Appl. Phys.*, Volume 45, p. 3023.
- Lany, S. & Zunger, A., 2008. Assessment of correction methods for the band-gap problem and for finite-size effects in supercell defect calculations: Case studies for ZnO and GaAs. *Phys. Rev. B*, 78(23), p. 235104.
- Lany, S. & Zunger, A., 2009. Accurate prediction of defect properties in density functional calculations. *Modelling Simul. Mat. Sci. Eng.*, Volume 17, p. 084002.
- Latham, C. D. et al., 1999. Density-functional calculations of carbon diffusion in GaAs. *Phys. Rev. B*, 60(22), p. 15117–15122.
- Londos, C. A., 1992. Investigation of a New Metastable Defect in Boron-Doped Cz-Si. *Phys. Status Solidi*, 133(429).
- Louie, S. G., 1982. Nonlinear ionic pseudopotentials in spin-density-functional calculations. *Phys. Rev. B*, 26(4), p. 1738–1742.
- Makov, G. & Payne, M., 1995. Periodic boundary conditions in ab-initio calculations. *Phys. Rev. B*, 51(7), p. 4014.
- Marie, P., Levalois, M. & Bogdanski, P., 1993. Deep level transient spectroscopy of high-energy heavy ion irradiation-induced defects in n-type germanium. *J. Appl. Phys.*, Volume 74, p. 868.
- Markevich, V. P. et al., 2004. Electronic properties of antimony-vacancy complex in Ge crystals. *J. Appl. Phys.*, Volume 95, p. 4078.
- Martin, R. M., 2004. *Electronic Structure*. s.l.:Cambridge University Press.
- Matsushita, M., Sato, K., Yoshiie, T. & Xu, Q., 2007. Validity of Activation Energy for Vacancy Migration Obtained by Integrating Force–Distance Curve. *Materials Transactions*, 48(9), pp. 2362–2364.

- Mesli, A. et al., 2008. Low-temperature irradiation defects in germanium: In situ analysis. *Phys Rev B*, 78(16), p. 165202.
- Meyer, W. E., 2007. *Digital DLTS studies on radiation induced defects in Si, GaAs and GaN*. s.l.:University of Pretoria.
- Mikkelson, J. L., 1986. *Oxygen, Carbon, Hydrogen and Nitrogen in Crystalline Silicon*. Pittsburgh, MRS Symposia Proceedings No. 59.
- Mooney, P. M., Poulin, F. & Bourgion, J. C., 1983. Annealing of electron-induced defects in n-type germanium. *Phys. Rev. B*, 28(6), p. 3372.
- Nagesh, V. & Farmer, J. W., 1988. Study Of Irradiation-induced Defects In Germanium. *J. Appl. Phys.*, Volume 63, p. 1549.
- Nielsen, H. K., 2005. *Capacitance Transient Measurements on Point Defects in Silicon and Silicon Carbide*. Stockholm: Doctoral Thesis.
- Nyamhere, C., Das, A. G. & Auret, F. D., 2011. Deep level transient spectroscopy (DLTS) study of defects introduced in antimony doped Ge by 2 MeV proton irradiation. *Physica B-Condensed Matter*, 406(15-16), pp. 3056-3059.
- Ouma, C. N. & Meyer, W. E., currently under review. Metastability of the boron-vacancy (B-V) complex in silicon: A hybrid functional study. *Physica Status Solidi (b)*.
- Ouma, C. N. & Meyer, W. E., currently under review. Metastable properties of Eu²⁺ and Eu³⁺ defect complexes in GaN: An ab-initio study. *Materials Science in Semiconductor Processing journal*.
- Paier, J. et al., 2006. Screened hybrid density functionals applied to solids. *J.Chem. Phys.*, 124(15), p. 154709.
- Pantelides, S. T., 1992. *Deep Centers in Semiconductors: A State of the art approach*. 2 ed. Switzerland: Lausanne.
- Parr, R. G. & Yang, W., 1994. *Density Functional theory of atoms and molecules*. s.l.:Oxford University Press.

- Patterson, J. & Bailey, B., 2010. *Solid State Physics: Introduction to the Theory*. 2nd ed. New York: Springer.
- Perdew, J. P., 2008. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.*, 100(13), p. 136406.
- Perdew, J. P. & Enzerhof, M., 1996. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.*, Volume 105, p. 9982.
- Perdew, J. P. & Levy, M., 1983. Physical content of the exact Kohn-Sham orbital energies: Band Gaps and derivative Discontinuities. *Phys. Rev. Lett.*, 51(20), p. 1884.
- Perdew, J. P., Levy, M. & Balduz, J. L., 1982. Density Functional Theory for fractional particle number: Derivative Discontinuities of the energy. *Phys. Rev. Lett.*, 49(20), pp. 1691-1694.
- Perdew, J. P. & Schmidt, K., 2001. *Jacobs ladder of density functional approximations for the exchange-correlation energy*. Antwerp (Belgium), AIP Conf. Proc. 577, p. 1.
- Perdew, J. P. & Wang, Y., 1986. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B*, 33(12), p. 8800.
- Petersen, M. C. & Nylandsted Larsen, A., 2010. Divacancy defects in germanium studied using deep-level transient spectroscopy. *Phys. Rev. B*, 82(7), p. 075203.
- Poulin, F. & Bourgion, J. C., 1982. Characteristics of the electron traps produced by electron irradiation in n-type germanium. *Phys. Rev. B*, 26(12), p. 6788.
- Puska, M. J., Persola, M., Niemainen, R. M. & Po, S., 1998. Convergence of supercell calculations for point defects in semiconductors : Vacancy in silicon. *Phys. Rev. B*, 58(3), p. 1318.
- Savin, A., Umrigar, C. J. & Gonza, X., 1998. Relationship of Kohn-Sham eigenvalues to excitation energies. *Chem. Phys. Lett.*, 288(2-4), pp. 391-395.
- Schroder, D. K., 2006. *Semiconductor Material and Device Characterization*. 3rd ed. s.l.:Wiley.
- Sham, L. & Schluter, M., 1983. Density functional theory of the Energy Gap. *Phys. Rev. Lett.*, 51(20), p. 1888.

- Shim, J. et al., 2005. Density-functional calculations of defect formation energies using the supercell method: Brillouin-zone sampling. *Phys. Rev. B*, 71(24), pp. 245204/1-7.
- Song, L. W., Zhan, X. D., Benson, B. W. & Watkins, G. D., 1988. A Molecular rebonding bistable defect in Silicon: the Interstitial Carbon-Substitutional Carbon Pair. *Phys. Rev. Lett.*, Volume 60, p. 460.
- Spiewak, P., Vanhellefont, J. & Kurzydowski, K. J., 2011. Improved calculation of vacancy properties in Ge using the Hyed-Scuseria-Ernzerhof range-separated hybrid functional. *J. Appl. Phys.*, Volume 110, p. 063534.
- Spiewak, P. et al., 2008. First Principal calculations of the fomration energy and deep levels associated with the neutral and charged vacancy in germanium. *J. Appl. Phys.*, Volume 103, p. 086103.
- Sprenger, M., van Kemp, R., Sieverts, E. & Ammerlaan, C., 1987. Electronic and atomic structure of the boron-vacancy complex in silicon. *Phys. Rev. B.*, 35(1582-92).
- Springborg, M., 2000. *Methods of Electronic Structure Calculations: From Molecules to Solids*. 1 ed. New York: Wiley.
- Stowasser, R. & Hoffman, R., 1999. What do the Kohn-Sham Orbitals and Eigenvalues Mean?. *J. Am. Chem. Soc.*, 121(14), pp. 3414-3420.
- Streetman, B. G. & Banjeree, S., 2000. *Solid state electronic devices vol 4*. New Jersey: Prentice Hall.
- Street, R. A. & Mott, N. F., 1975. States in the gap in glassy semiconductors. *Phys. Rev. Lett.*, 35(19), p. 1293.
- Tahini, H. et al., 2011. Diffusion of E centers in germanium predicted using GGA+ U approach. *Appl. Phys. Lett.*, Volume 99, p. 162103.
- Van de Walle, C. G., 2004. First-principles calculations for defects and impurities:Applications to III-nitrides. *J. Appl. Phys.*, Volume 95, p. 3851.
- Van de Walle, G. & Ceder, A., 1999. Correcting overbinding in local-density-approximation calculations. *Phys. Rev. B.*, 59(23), p. 14992.

- Van Zeghbroeck, B., 2011. Principles of semiconductor Devices. In: s.l.:s.n.
- Weber, J. R., Janotti, A. & Van de Walle, C. G., 2013. Dangling Bonds and vacancies in germanium. *Phys. Rev. B*, 87(3), p. 035203.
- Xiao, H., Tahir-Kheli, J. & Goddard, W. A., 2011. Accurate Band Gaps for Semiconductors from Density Functional Theory. *Journ. Phys. Chem. Lett.*, 2(3), pp. 212-217.
- Yakovin, I. N. & Dowben, P. A., 2007. The problem of the band gap in LDA Calculations. *Surface review and Letters*, 14(3), pp. 481-487.
- Zakutayev, A., 2013. Metastable defects are the origin of high conductivity in gallium doped zinc oxide. *Appl. Phys. Lett.*, Volume 103, p. 232106.
- Zangenberg, N. R. & Nylandsted Larsen, A., 2005. On-line DLTS investigations of vacancy related defects in low-temperature electron irradiated, boron-doped Si. *Appl. Phys. A*, 80(5), pp. 1081-1086.
- Zhang, S. & Northrup, J., 1991. Chemical potential dependence of defect formation energies in GaAs: Application to Ga self-diffusion. *Phys. Rev. Lett.*, 67(17), p. 2339.
- Zhang, Y. & Yang, W., 2000. Perspective on Density functional theory for fractional particle number: derivative discontinuities of the energy. *Theor. Chem. Accounts Theory, Comput. Model. (Theortica Chim Acta)*, 103(3-4), pp. 346-348.
- Zhao, Y. & Truhlar, D. G., 2008. Construction of a generalized gradient approximation by restoring the density-gradient expansion and enforcing a tight Lieb-Oxford Bound. *American. Ins. Phys.*, Volume 128, p. 184109.
- Zistl, C., 1997. *Pb.D. thesis*. Berlin: Hahn Meitner Institut.