Defects and Their Properties in InP and InGaAs

An *ab initio* study

Christian Dam Vedel

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Engineering
College of Science and Engineering
University of Glasgow

June 2024

Abstract

Defects in III-V optoelectronic semiconductor devices are one the main bottlenecks for further device improvements and can in extreme cases even cause complete loss of function of the device. Some defects have very low formation energies, which makes avoiding them during device fabrication particularly hard and sometimes impossible.

One such case of defects are polytypic defects in indium phosphide, such as rotational twin planes, which besides their detrimental effects, also have unique properties that can possibly be utilised beneficially. One such possibility is crystal-phase quantum wells, which is a quantum well created by the crystal structure of a single material. Through *ab initio* studies in the state-of-the-art simulation software suite QuantumATK, it is in this thesis found that such crystal-phase quantum wells must have a minimal width of 10 nm. Another property which can potentially be utilised, is the dielectric constant anisotropy, which is discovered to be induced in Zincblende InP by the breaking of the A-direction symmetry by polytypic defects. One semiconductor property usually affected by defects, which is especially important in electronic devices like transistors, but which is also important in optoelectronic devices, is the current transport in the semiconductor. In the case of polytypic defects in InP, it is in this thesis discovered that for current transport along the [1 1 1]/[0 0 0 1]-direction, it is more beneficial to have more defects, rather than fewer of them, if they cannot be completely avoided, due to the Wurtzite structure having better transport properties. For transport in the defect plane, it is found that the defects increase the current conducting capabilities of InP, with a well-defined optimal fraction of polytypic defects.

A defect-related property which is especially important in optical devices such as PIN photodiodes, due to its detrimental effect on the photocurrent, is trap-assisted recombination. This recombination is studied in 23 different experimentally feasible point defects in $In_{0.53}Ga_{0.47}As$ in this thesis, and it is found that only 7 of them can host a notable carrier recombination. Out of those 7 defects, the most detrimental defects are found to be the antisites, and in particular the double antisites. This discovery is unfortunate for semiconductor devices, as the defect levels of the antisites also aligns well with the experimentally measured levels, suggesting that they are regularly occurring in device fabrication. In order to do this study, two different atomistic *ab initio* methods for simulating random alloys are also evaluated and compared, and it is found that the Special Quasi-random Structure (SQS) method performs better than the Virtual Crystal Approximation (VCA) for these types of studies, based on their implementation in QuantumATK U-2022.12 by Synopsys.

Contents

List of Tables

List of Figures

Acknowledgements

This PhD has been a roller-coaster of a journey for me, with many exciting ups and many depressing downs, including Covid. I am eternally thankful to the many people who have made the tough times more bearable and the good times more enjoyable during this journey. This obviously includes my academic supervisor Vihar P. Georgiev, and my industrial supervisor Søren Smidstrup, both of whom without this PhD would have never been possible, they have both been pillars of support and guidance for me throughout my PhD. I am also especially thankful to Tue Gunst, who has been like a third supervisor for me at Synopsys, without him, this thesis would have been insurmountable to finish. My thanks extend to all the people of Synopsys QuantumATK Denmark, who made me feel welcome and included during my 2 years working among them. I would also like to extend my thanks to the people of the Device Modelling group of University of Glasgow, who helped me through my year at Scotland during Covid. In that regard I am also extremely grateful to my social circle in Scotland, and in particular to my extended household during Covid, for keeping me company and away from depression in an otherwise lonely and hard time. Finally I am also grateful to my family and friends in Denmark, for their moral support and presence during this academic journey.

Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Printed name: Christian Dam Vedel

Chapter 1

Introduction

1.1 The semiconductor industry

The semiconductor industry has experienced an immense growth throughout the last couple of decades, with global sales growing from the 100 billion dollar range in the early 2000s, to being in the upper 500 billion dollar range in 2022 [1]. This growth has been spurred by a surge in demand for semiconductor chips across different market segments. With the rise of "big-data", 5G internet and "smart-everything", the number of chips used in common consumer products have spiked, as products which have previously not utilised chips have started incorporating them for various benefits [2]. Simultaneously, the automotive industry has sharply increased the number of chips used in vehicles, to between 1500 and 3000 per vehicle, for systems such as proximity sensors and autonomous driving systems [3]. The semiconductor industry has been struggling to meet these increased demands, causing a global semiconductor shortage crisis which have persisted since its instigation by the beginning of the COVID-19 pandemic. The pandemic caused factories all over the world to shut down or reduce their operation, creating a backlog of semiconductor chip requests, which were only worsened by the increased consumer demand in the same time period. Following the pandemic, the geopolitical situation between America and China and the Russia-Ukranian war, served to continuously hinder the semiconductor industry supply chains, thereby continuing the global semiconductor shortage crisis [4–7].

In order for the semiconductor industry to catch up with the backlog of chip requests, and to alleviate the global semiconductor shortage, new foundries are being constructed and currently existing ones are being expanded, in order to increase the global production capacity. Yet, this is only part of the solution to the global shortage problem, as the semiconductor industry also struggles with low yields with their current cutting-edge fabrication nodes. The semiconductor industry owes this issue to their own success, as the lower yield is a direct consequence of the industry's rapid improvement of the computational power of microchips. The semiconductor industry has, for the past few decades, doubled the transistor density on a chip every 18 months, as can be seen in the data and projections from Intel in Figure 1.1 [8]. This doubling of transistor density has resulted in much faster computer chips and has thereby fuelled the industry's immense growth. This trend of doubling the transistor density

has become known as "Moore's Law" [9], and has worked more as a guiding target for the industry, rather than the future prediction that Moore initially coined it as.

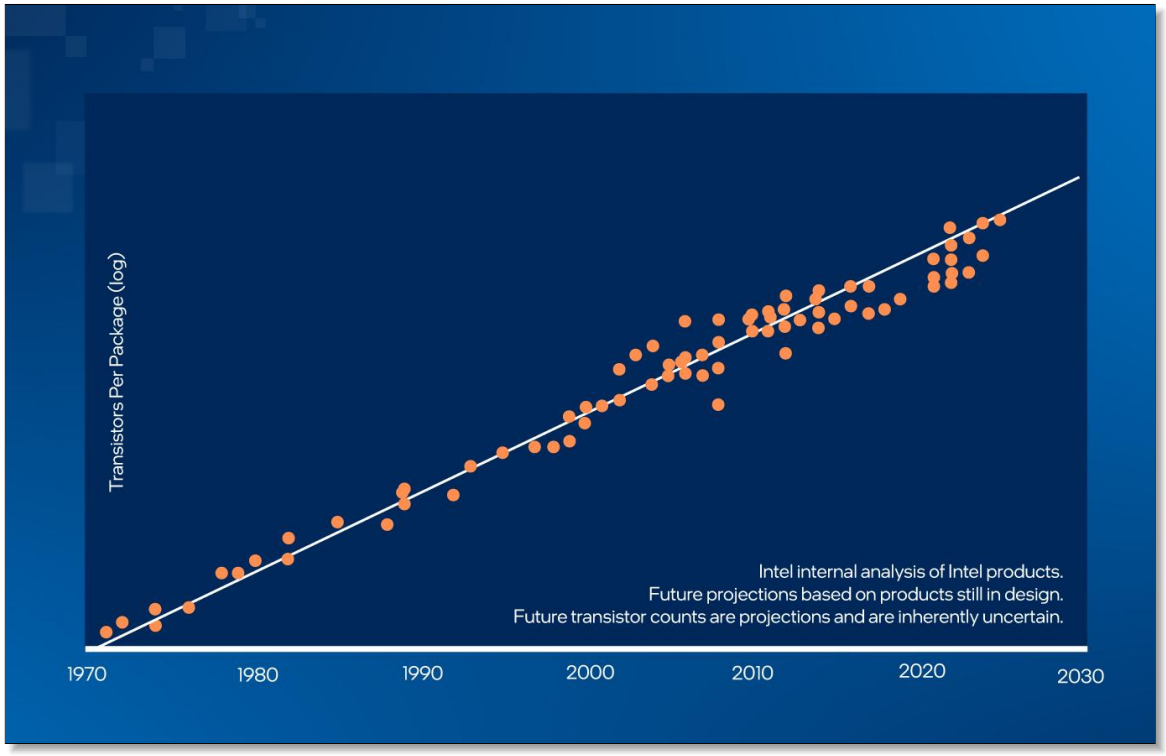


Figure 1.1: Logarithmic graph of transistor density per year, with data and future projections from Intel [8]. The logarithmic fit of the data is known as Moore's law.

Through aggressive down-scaling of transistors, Moore's law has been generally upheld to this day, despite growing challenges and issues with continued down-scaling. Today transistor dimensions have reached below 10 nm, a size at which even a small defect can cause catastrophic loss of function. Defects and faults are common occurrences during chip fabrication and they have often led to reduced yields [10] due to degradation of the transistors [11], but with the previously larger transistors, they had to be numerous or severe in order to impact the functioning of the transistors and thereby the resulting microchips. Yet with the semiconductor industry aiming to reduce transistor dimensions even further, reducing the occurrences of defects and faults are paramount to maintaining the current yield for even yet more sophisticated fabrication nodes, and also to increase the yield of the current fabrication nodes, in order to help alleviate the global semiconductor shortage crisis. To reduce the occurrence of critical defects in semiconductor fabrication, it is necessary to research and understand the defects and how they impact the functioning of the semiconductors and transistors, which is therefore also the goal of the present doctoral thesis.

1.2 Defects

Crystal defects can be split into four distinct categories based on their spatial dimensionality: Point defects (0D), Line defects (1D), Surface/interface defects (2D) and volume defects

(3D). Point defects consists of a single or a few atoms in the crystal lattice, which are missing, in surplus at interstitial sites, are the wrong atomic species, or a combination of all these three possibilities. Point defects will be further explored in Section 4.4, where examples of four selected point defects are shown in Figure 4.14. Line defects consists of dislocation lines wherein the crystal lattice abruptly shifts or "slips" by one or more atomic positions and the perfect crystal order is restored on either side of the dislocation line. Line defects often occur to relieve crystal stress, such as stress incurred from matching crystal lattices at interfaces between different materials. An example of a line defect is shown in the orange oval in Figure 3.1B. Surface defects are similar to dislocation lines, in that perfect crystal order is restored on either side of the plane which defines the defect. Surface defects include grain boundaries, which are boundaries between parts of a material with different crystallographic orientations, and twins and stacking faults, which are both related to errors in the stacking sequence of the material. Rotational twin planes, a type of twin defect that can also be considered a grain-boundary, with a relative orientation difference consisting of a rotation about the defect plane's axis by 60°, will be further explored in Chapter 3 and are indicated by red lines in Figure 3.1. Finally, volume defects consist of any remaining conceivable defects with widths in all three dimensions considerably larger than the lattice spacing, including large voids in the crystal, poly-granular precipitates or foreign molecules or large particles [12, 13].

A significant amount of research has been carried out, in order to attempt to understand and possibly avoid all of these defects in fabricated devices, as many of them degrade device performance in some way or another [11, 13], such as by acting as recombination centres [14, 15], current leakage paths [16, 17], or by causing optical birefringence [18]. In the article of Moram et al. [19], the origin and cause of threading dislocations in thin films are researched, due to their critically device performance impairing effects. Bioud et al. [20] investigated threading dislocations caused by lattice mismatch with Silicon, and suggested an elimination procedure to remove the defects by dislocation-selective etching. Chen et al. [21] studied the effect of stacking faults on Young's modulus of various nanowires and in the articles of Krishnamachari et al. [22] and Hannah et al. [23], methods to avoid stacking faults and twins in nanowires are developed, by growing in a specific crystal facet direction and through a two-temperature growth procedure, respectively. In the article of Tahini et al. [24], vacancies, a type of point defect, are studied in an array of semiconductors to ascertain their likelihood of formation and the possibility to avoid them by growing in specific element-rich conditions. Lee et al. [25] showed, using first-principles simulations, that antisites, another type of point defect, traps charges in gate-all-around Nanowire Tunnel Field-Effect-Transistors (NW TFETs), thereby limiting the performance of the transistor and of further down-scaling. Grandidier et al. [26] studied antisites and precipitates using Scanning Tunnelling microscopy (STM), and found that precipitates were avoided and antisites were reduced by doping the semiconductor with beryllium, and that the decrease in photoresponse time in the undoped semiconductor was caused by the precipitates acting as recombination centres.

This small sample of research articles studying defects in semiconductors highlights the importance and breadth of the field as a whole, and shows that all the different kinds of defects

are critically important in various different materials and devices. The study undertaken in this thesis will contribute further to this field of research, filling out important gaps in the currently established knowledge of defects.

1.3 III-V semiconductors

Besides the miniaturisation of transistors, another possibility to maintain the steady improvements of microchips, and thereby the growth of the semiconductor industry, which can progress in tandem with the usual down-scaling, is to change the transistor design, which has already happened multiple times. One such way to improve the transistor design, is to incorporate new materials into the standard silicon platform or replacing existing materials with more optimal alternatives. This type of improvement has a recent successful precedent, when the regular silicon channel of the transistor was replaced by strained silicon, owing to its increased charge carrier mobility [27]. This particular transistor material component replacement can conceivably be even further improved, by again replacing this strained silicon with a III-V semiconductor, i.e. a compound semiconductor consisting of materials from the third and fifth group of the periodic table, due to this group of materials even larger carrier mobilities [28–30]. The high mobility of III-Vs are already utilised with great success in High-Electron Mobility Transistors (HEMTs), where the current-gain and power-gain cutoff frequencies exceed 600 GHz and 1 THz respectively [31–33]. Building upon the design successes of III-V HEMTs, incorporating III-Vs into the more commonly used Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs), promises to yield great benefits for the device's operating performance [34–36]. Besides the proposed transistor improvement, III-V semiconductors are extremely relevant to the semiconductor industry in other electrical devices and in particular in optical devices, owing to their often direct and tunable band gaps [37–39]. Band gaps which can be tuned to be in the low dispersion and low loss window for fiber communication, making them useful for lasers, modulators, optical amplifiers and more [40–42]. Band gaps which can be tuned and layered in multi-junction solar cells, to achieve photoabsorption efficiencies at the current world record level of 47.6% in a quadruple junction solar cell [40, 43, 44]. Band gaps and intersubbands which can be utilised for long wavelength emitters and infrared detectors [40, 45], and which can even be tuned to the visible spectrum, for usages in high-efficiency LEDs which can possibly even bridge the notorious "green gap" in LED technology [40, 46]. III-V semiconductors also have very low exciton binding energies, making them ideal for devices utilising quantum confinement, such as colloidal nanocrystals, which find use-cases as near- or mid-infrared emitters and detectors [47, 48]. III-Vs are used in many types of optical detectors besides colloidal nanocrystals, such as depletion-mode and avalanche photodiodes [49, 50] and even in novel devices such as topological lasers [51, 52] or superconductive qubit couplers [53].

1.4 PIN photodiode

To limit the scope of research from all defects in all III-V materials, to a combination of which can be managed within the confines of this doctoral thesis, a specific semiconductor device is chosen to guide the present research. The chosen device is the PIN nanowire photodiode for Near-Infrared Region (NIR) photodetection, due to its relevance in the "Defect Simulations and Material Growth of III-V Nanostructures" (DESIGN) project [54], which this thesis is also a part of. The PIN photodiode was invented in 1950 by Jun-ichi Nishizawa [55] and is an adaptation of the simple PN junction, adapted to work better as a photodetector or high-frequency diode, by increasing the width of the photon-absorbing depletion region. In the PN junction, the size of the depletion region is determined by the diffusivity of the free carriers and the applied voltage, and as such it is never very large. In the PIN diode, an undoped, i.e. intrinsic, or lightly doped region is inserted in between the heavily n- and p-doped regions, hence the "PIN" moniker, which significantly increases the width of the depletion region. A simplified diagram showing the PIN diode operating as a photodetector is shown in Figure 1.2. As is shown, the PIN diode operates as a photodetector in an identical fashion to the PN junction: incident photons, with energies larger than the band gap of the semiconductor, are absorbed in the depletion/intrinsic region, exciting an electron-hole pair, which then flows into the contacts to produce a current. Since the photons are captured in the depletion region, the increased size of this region in the PIN diode, increases the quantum efficiency of the photodiode. The intrinsic region also has the added benefit of increasing the separation between the p- and n-regions, thereby lowering the capacitance of the diode, which in turn improves the diode response time and increases the breakdown voltage, making the PIN diode more useful for high frequency and high-voltage power applications [56, 57].

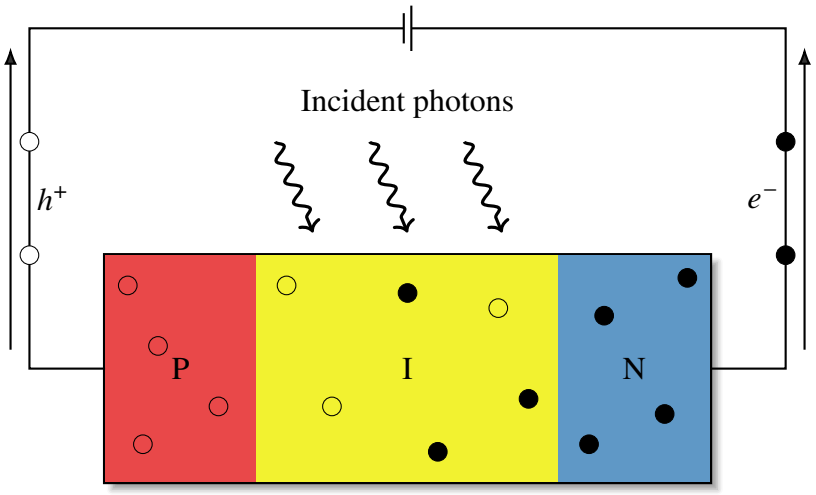


Figure 1.2: Simplified diagram of a PIN photodiode in operation. Electrons e^- and holes h^+ are indicated by filled and empty circles respectively.

Photodiodes can operate either in photovoltaic mode, i.e. without an applied bias, or in photoconductive mode, i.e. in reverse bias. When the diode is operating in photovoltaic mode, the mode of operation for solar cells, there is little to no current without illumination and a linearly increasing photocurrent with the incident photon flux, thereby giving a high

signal-to-noise ratio, making this mode useful for low-light applications. When a reverse bias is applied to the PIN diode, the width of the photon-absorbing depletion region is further increased, thereby increasing the quantum efficiency of the photodetector. The applied voltage also increases the drift velocity of carriers in the diode, which in turn reduces the diode response time. The drawback of using a photodiode in reverse bias, is the so-called "dark current" which flows even in the lack of illumination, and therefore acts as noise in the system, thereby reducing the sensitivity of the diode. While current should in general not flow in a diode under reverse bias, various mechanisms cause this small noise current, such as band-to-band tunnelling or defect-mediated trap-assisted tunnelling [56, 57].

For detection in the NIR, beyond the absorption edge of silicon, a suitable semiconductor for the PIN photodiode, is the ternary compound III-V semiconductor $In_xGa_{1-x}As$. The band gap of $In_xGa_{1-x}As$ is direct in the entire composition range and can be tuned by the indium fraction, from the band gap of InAs in the mid-infrared region (0.354 eV/3500 nm) to the band gap of GaAs in the lower NIR (1.42 eV/873 nm). Besides the suitability of the band gap, $In_xGa_{1-x}As$ also has high carrier mobilities, even among the III-V semiconductors, therefore yielding fast response times of the photodiode [35, 58, 59]. In the DESIGN project article of Oliver et al. [60], PIN photodiodes consisting of either an $In_{0.53}Ga_{0.47}As$ homojunction, an InP homojunction or a *p*-InP/*i*-InGaAs/*n*-InP heterojunction, are compared for photodetection in the NIR, using the finite-element software Sentaurus Device. It was found that the *p*-InP/*i*-InGaAs/*n*-InP heterojunction yields the optimal signal-to-noise ratio, beyond the absorption edge of silicon in the NIR. The specific indium fraction of 53% were chosen to match the lattice constant of the InGaAs with that of InP, in order to reduce the amount of defects forming from lattice mismatch strain-relaxations at the heterojunction boundaries, in an eventual fabricated device. The *p*-InP/*i*-InGaAs/*n*-InP heterojunction system outperforms the pure InGaAs diode, due to the dark current mechanisms being proportional to the band gap of the material. As such, using InP instead of InGaAs in the diode where possible, without a substantial loss of photocurrent, improves the performance of the device by lowering the signal noise.

The InP/InGaAs/InP PIN diode is a well-researched system, with much of the research investigating defects as the performance limiting factor of the photodiode [61, 62]. It was shown specifically by Wang et al. [63], that the performance-limiting factor in the diode, is the SRH recombination from point defects in the photon-absorbing intrinsic InGaAs region. For this reason, this particular dark current mechanism caused by point defects in InGaAs is studied in Chapter 4 of this doctoral thesis. But before that, the current conducting and optical properties of the InP regions are investigated in Chapter 3, following the subsequent Chapter 2, in which the theory and simulation methodology utilised in this thesis is introduced.

Chapter 2

Atomistic First-Principles Methods

There exists numerous methodologies to simulate devices and materials at the atomistic level, ranging from finite-element methods for large-scale simulations, to many-body methods, also known as GW methods, for high-accuracy simulations. These methods each have separate benefits and drawbacks, making each method ideal only for certain systems and research areas. For example, finite-element methods, such as those implemented in Synopsys Sentaurus, are well-suited for simulating devices with geometries in the nm-mm range, due to their high efficiency and accuracy at this scale, owing to their well-optimised model parameters. To simulate defects, which is the topic of this thesis, finite-element methods are insufficient, as they have no description of individual atoms, and thereby no modelling of the actual atomic defects compositions and geometries. For such a research subject, an atomistic simulation methodology is required, and among the atomistic simulation methods, Density Functional Theory (DFT) and Non-Equilibrium Green Functions (NEGF) provide the best trade-off between accuracy and simulation scale, in order to simulate the entirety of defect geometries. This chapter is therefore dedicated to the theoretical foundation and the computational details of these methods.

2.1 The many-body problem

Identical for both DFT, NEGF and many other atomistic simulation methods, is that they attempt to solve the fundamental equation of quantum mechanics, the Schrödinger equation. The full Schrödinger equation of a system includes all interactions in the system, including the complicated many-body interactions, which in most methods are approximated in order to solve the equation. From elementary quantum mechanics, it is known that the solution of the time-independent version of the Schrödinger equation yields the allowed energy levels of a system

$$\hat{H} \left|\psi_i\right\rangle = \epsilon_i \left|\psi_i\right\rangle \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian operator of the system, $|\psi_i\rangle$ is an eigenvector of the Hamiltonian and ϵ_i is its eigenvalue, which is one of the allowed energies of the system. In a system consisting of N free electrons with coordinates $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$ and M ions with coordinates

$\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M$, the full many-body Hamiltonian is written as

$$\hat{H} = \frac{1}{2}\Bigg[\underbrace{\sum_{I\neq J}^{M,M}\frac{Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|}}_{\text{Ion-ion coulomb}} \underbrace{-\sum_{i}^{N}\nabla_i^2}_{\text{Electron kinetic}} \underbrace{-\sum_{I}^{M}\frac{\nabla_I^2}{M_I}}_{\text{Ion kinetic}} \underbrace{-\sum_{i,I}^{N,M}\frac{2Z_I}{|\mathbf{r}_i-\mathbf{R}_I|}}_{\text{Ion-electron coulomb}} \underbrace{+\sum_{i\neq j}^{N,N}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}}_{\text{Electron-electron coulomb}}\Bigg] \tag{2.2}$$

where lowercase i and j denote electrons, uppercase I and J denote ions and Z_I and M_I are the charge and weight of ion I, respectively [64]. Eq. (2.2) is written in atomic units, i.e. $\hbar = m_e = e = 4\pi\varepsilon_0 = 1$, where $\hbar$, m_e, e and ε_0 are the reduced Planck constant, the electron mass, the electron charge and the vacuum permittivity, respectively. Usually the ion and electron dynamics are separated into separate problems, due to the large difference in their respective timescales. This approximation, termed the Born-Oppenheimer approximation [65], effectively allows the electronic problem to be solved using stationary ions, so that the kinetic energy term of the ions in Eq. (2.2) vanishes and the ion-ion interaction term becomes a constant, which can be removed by shifting the zero-point energy.

In the Born-Oppenheimer approximation the electronic states of the system evolves adiabatically in time with the movement of the ions, such that the electronic occupation of the states remain unchanged as the states evolve. This approximation therefore effectively neglects non-radiative transitions between electronic eigenstates, a physical phenonema of particular interest in this thesis, which thus has to be studied using special methods as is described in Section 4.5.

For crystalline solids, such as the ones to be studied in this thesis, and many other systems, another important approximation is conventionally utilised. Namely it is usually assumed that the system is periodic in one or more dimensions, which is the case in a perfect crystal, so that the ionic potential is also periodic

$$V(\mathbf{r}) = V(\mathbf{r}+\mathbf{R}) \tag{2.3}$$

where $\mathbf{R}$ is a lattice vector of the crystal lattice. This assumption is utilised in order to employ Bloch's theorem, which states that the wave function of a particle in a periodic potential, can be written as the product of a function with the same periodicity as the potential, and an imaginary phase factor stemming from the translational symmetry [66]. This is written as

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}) \tag{2.4}$$

where $u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}+\mathbf{R})$ is a periodic function characteristic of the systems unit cell, which is the part of the system that can be repeated indefinitely to produce the full system. The wavevector $\mathbf{k}$ in Eq. (2.4), also called the crystal momentum, is not free to take on any value, as it is only defined up to the reciprocal lattice vector $\mathbf{k} = \mathbf{k}+\mathbf{G}$. The advantage of utilising Bloch's theorem, is that instead of solving Scrödingers equation for the entire system, it is only necessary to solve it within the unit cell to calculate $u_{\mathbf{k}}$ for each $\mathbf{k}$ in the first Brillouin zone.

Despite the simplifications resulting from using the Born-Oppenheimer approximation and Bloch's theorem, solving the many-body Hamiltonian in Eq. (2.2) for unit cells with more than just a few atoms, remain an exceedingly demanding and untractable task, even with modern computational hardware. To exemplify how demanding the task is, one can make a rough estimate of the memory requirement necessary to simply store the many-body wave function. To make a very liberal estimate, it is assumed that only ten values of the wave function in each of its coordinates is sufficient to replicate the wave function. In such an estimate, a one-electron wave function would require $10 \times 10 \times 10 \times 8\,\mathrm{B} = 8\,\mathrm{kB}$ of memory, assuming that a single value is sufficiently represented in 8 B. Due to the interaction between the electrons, their coordinates are interdependent and the wave function of multiple electrons cannot be factorised as a product of one-electron wave functions or even a linear combination of these, i.e. a Slater determinant. Therefore, the storage required to store the many-body wave function grows as 10^{3N} with the number of electrons, quickly outgrowing any modern physical storage system.

The offending term in the many-body Hamiltonian, responsible for the inability to factorise the wave function, is the electron-electron Coulomb interaction:

$$V_C(\mathbf{r}) = -\frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{|\mathbf{r}|} \tag{2.5}$$

Here, q_1 and q_2 are the charges of the interacting particles and $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ is the vector between the particles positions. Therefore, in atomistic simulation methods, this term is approximated with models of various degree of sophistication.

2.2 Density functional theory

Density Functional Theory has its first beginnings in 1964, when Hohenberg and Kohn suggested to use the ground-state electron density as the fundamental variable to solve Shrödingers equation, instead of the complicated many-body wave function [67]. The advantage of such an approach, would be that it is much simpler to work with the only three degrees of freedom electron density, rather than the $3N$ degrees of freedom many-body wave function. However, normally the density is calculated from the wave function, rather than in vice versa:

$$n(\mathbf{r}) = n(\mathbf{r}_1) \equiv N\int d\mathbf{r}_2 \ldots \int d\mathbf{r}_N \psi^*(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \tag{2.6}$$

As such, in order to be able to use the density as the fundamental variable instead of the wave function, the relationship between these identities has to be univocal in both directions. Proof of this was given using *reductio ad absurdum* in Hohenberg and Kohns article:

2.2.1 The Hohenberg-Kohn theorems

Assume that two different potentials $v(\mathbf{r})$ and $v'(\mathbf{r})$ both yields the same density $n(\mathbf{r})$. If $v(\mathbf{r})$ and $v'(\mathbf{r})$ differs (except for a constant shift), they each give two different Schrödinger equations and thus two different set of wave functions $\psi(\mathbf{r})$ and $\psi'(\mathbf{r})$.

Using the variational principle, evaluating the ground state energy with potential $v(\mathbf{r})$ using the wave functions $\Psi'(\mathbf{r})$ yields an energy higher than the true ground state energy (since $\psi(\mathbf{r}) \neq \psi'(\mathbf{r})$)

$$\langle\psi_0|\,\hat{H}\,|\psi_0\rangle = E_0 < \langle\psi_0'|\,\hat{H}\,|\psi_0'\rangle \tag{2.7}$$

where the Hamiltonian: $\hat{H}$, in Eq. (2.7) is the Hamiltonian associated with $v(\mathbf{r})$. Using that $\hat{H} = \hat{H}' - v'(\mathbf{r}) + v(\mathbf{r})$:

$$\begin{aligned} &E_0 < \langle\psi_0'|\,\hat{H}' - v'(\mathbf{r}) + v(\mathbf{r})\,|\psi_0'\rangle \Leftrightarrow \\ &E_0 < E_0' + \int d\mathbf{r}[v(\mathbf{r}) - v'(\mathbf{r})]n(\mathbf{r}) \end{aligned} \tag{2.8}$$

The same procedure, using the wave functions $\psi(\mathbf{r})$ to evaluate the energy in $v'(\mathbf{r})$, similarly yields:

$$E_0' < E_0 + \int d\mathbf{r}[v'(\mathbf{r}) - v(\mathbf{r})]n(\mathbf{r}) \tag{2.9}$$

The densities appearing in Eq. (2.8) and (2.9) are the same due to the assumption. Adding these two inequalities together yields the absurd equation:

$$E_0 + E_0' < E_0' + E_0 \tag{2.10}$$

Since Eq. (2.10) is nonsense, the assumption must be wrong and the potentials $v(\mathbf{r})$ and $v'(\mathbf{r})$ has to yield different electron densities. According to the above proof, the wave function is indeed a unique functional of the electron density, and it can therefore be replaced as the fundamental variable of Schrödingers equation.

This proof, called the first Hohenberg-Kohn theorem, ensured the theoretical foundation of Density Functional Theory, which together with the second theorem, constitute the pillars of DFT. The second Hohenberg-Kohn theorem is in essence the recasting of the variational principle to determine the ground state energy of a system in terms of the electron density. Repeating Eq. (2.2) in terms of the electron density, with the Born-Oppenheimer approximation applied:

$$E = F[n] = \langle\psi[n]|\,\hat{H}\,|\psi[n]\rangle = \langle\psi[n]|\,\hat{T} + \hat{U}_{ee}\,|\psi[n]\rangle + \int d\mathbf{r}\, n(\mathbf{r})v_{\text{ext}}(\mathbf{r}) \tag{2.11}$$

Here $\hat{T}$ is the kinetic energy operator, $\hat{U}_{ee}$ is the electron-electron interaction operator and v_{ext} is the external potential, which usually consists of the potential generated from the lattice ions.

2.2.2 Kohn-Sham theory

As previously mentioned, solving Eq. (2.11), variationally or otherwise, is exceedingly complicated due to the Coulomb functional term $U[n] = \langle\psi[n]|\,\hat{U}_{ee}\,|\psi[n]\rangle$ and the kinetic functional term $T[n] = \langle\psi[n]|\,\hat{T}\,|\psi[n]\rangle$. Finding good approximations for these terms have proved difficult due to the sensitivity of the equation to even just small variations in $T[n]$ or $U[n]$. To circumvent this issue, Kohn and Sham reformulated Eq. (2.11) in 1965, so that only the energetically small and computationally demanding parts of $U[n]$ and $T[n]$ had to be approximated, rather than the entirety of the terms [68]:

$$E = \int d\mathbf{r}\, v_{\text{ext}}(\mathbf{r})n(\mathbf{r}) + T_s[n] + \frac{1}{2}\iint d\mathbf{r}d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\text{XC}}[n] \tag{2.12}$$

In Eq. (2.12) $T_s[n]$ is the large, but simple, bare non-interacting kinetic energy part of $T[n]$ which remains after the kinetic correlation has been moved to the functional $E_{\text{XC}}[n]$. The final term is the so-called Hartree term, which is the remainder of $U[n]$ after the exchange interaction and correlation have also been collected in the $E_{\text{XC}}[n]$ term. The collection of these complicated but energetically small terms, $E_{\text{XC}}[n]$, is called the Exchange-Correlation (XC) energy functional.

To solve Eq. (2.12), Kohn and Sham formulated a self-consistent practical approach, using an imaginary reference system. This imaginary reference system is constructed such that it has the same ground state density as the real system, without the real systems electron interactions. This is made possible by the so-called Kohn-Sham potential $v_r(\mathbf{r})$ added to the system, which is simply imagined to be, whatever it must be, to ensure the same ground state density. The Hamiltonian of the imaginary system is therefore:

$$\hat{H}_R = \sum_{j=1}^{N}\left[\hat{T}_{s,j} + v_r(\mathbf{r}_j)\right] \qquad \left[\hat{T}_{s,j} + v_r(\mathbf{r}_j)\right]\phi_i(\mathbf{r}_j) = \epsilon_i\phi_i(\mathbf{r}_j) \tag{2.13}$$

The solutions to Eq. (2.13), the single-particle wave functions $\phi_i(\mathbf{r})$, are called the Kohn-Sham orbitals. Due to the lack of electron-electron interactions, the many-body wave functions of the imaginary system can be constructed as anti-symmetric Slater Determinants (SD) of the Kohn-Sham orbitals. The electron density $n(\mathbf{r})$ in the imaginary system is therefore:

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{N_s!}}\text{SD}[\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2),\ldots,\phi_{N_s}(\mathbf{r}_{N_s})] \Rightarrow n(\mathbf{r}) = 2\sum_{i=1}^{N_s}|\phi_i(\mathbf{r})|^2 \tag{2.14}$$

Here, spin-degeneracy has been assumed, so that each single-particle orbital $\phi_i(\mathbf{r})$ is occupied by two electrons and $N_s = N/2$ is the number of occupied orbitals. Due to the first Hohenberg-Kohn theorem, the electron density which minimises the energy of Eq. (2.13), must be the ground state density. Variationally minimising the energy functionals of the real and imaginary systems, under the constraint $\int d\mathbf{r}\, n(\mathbf{r}) = N$, leads to Euler-Lagrange equations in

the density:

$$\frac{\partial T_s[n]}{\partial n} + v_r(\mathbf{r}) = \mu \tag{2.15}$$

$$\frac{\partial T_s[n]}{\partial n} + \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + v_{\text{ext}}(\mathbf{r}) + \frac{\partial E_{\text{XC}}[n]}{\partial n} = \mu \tag{2.16}$$

The Lagrange multiplier μ in Eqs. (2.15) and (2.16) is recognised as the chemical potential of the systems. Due to the reference system being constructed such that it has the same ground state density as the real system, the density which minimises the energy functional of the imaginary system, must be the same density as the one which minimises the energy functional of the real system. Setting the chemical potentials of the two systems equal, then leads to an expression for the Kohn-Sham potential:

$$v_r(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\partial E_{\text{XC}}[n]}{\partial n} \tag{2.17}$$

With this connection between the real and imaginary reference system, a self-consistent loop can be used to find the ground state electron density. The self-consistent loop is initialised by making an educated guess of the final electron density, and since the self-consistent loop converges in fewer cycles with a better starting density, the accuracy of this guess is not unimportant. One method to make a good guess of the initial density, is to construct it from a Linear Combination of Atomic Orbitals (LCAO) using Eq. (2.14). With an initial guess of the electron density, the Kohn-Sham potential is then calculated from Eq. (2.17), and afterwards the Kohn-Sham equations, Eq. (2.13), are solved to find the Kohn-Sham orbitals $|\phi_i\rangle$. These orbitals are then used to calculate a new electron density, which can then be used as the initial density of the next iteration of the loop. This loop is then continued until some convergence threshold is reached, which is often chosen to be when the difference of the total energy of the system, as calculated with Eq. (2.12), between two successive iterations, ΔE_{tot}, is below a chosen threshold, ΔE_{lim}. A more strict option for the convergence criteria, which is the one chosen in QuantumATK, is that the absolute difference between individual Hamiltonian elements of successive iterations, $\Delta \hat{H}_i$, is below the chosen threshold for all elements of the Hamiltonian. This general self-consistent loop is shown in Figure 2.1.

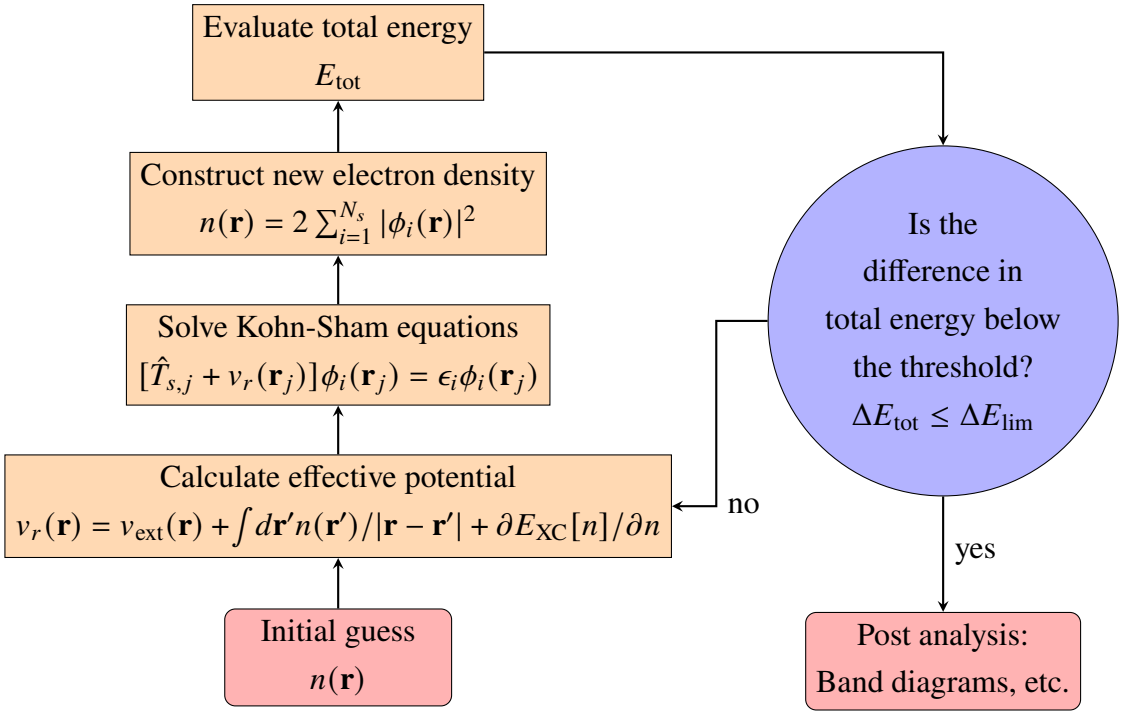


Figure 2.1: Kohn-Sham density functional theory self-consistency loop

After self-consistency have been achieved with this self-consistent loop, the ground state electron density can be used to calculate a number of system properties, such as magnetization, band diagrams, atomic forces etc. For structural relaxations, a similar self-consistent loop is used on top of the DFT loop, wherein the convergence criteria is the total forces on the atoms, which are calculated based on ground state density resulting from a DFT loop of each successive structure.

2.2.3 The exchange-correlation functional

The Kohn-Sham DFT method is in principle an exact method, if the form of the XC functional, $E_{\text{XC}}[n]$, is known. Yet since it is not known, it is always approximated by various degrees of precision. In Kohn and Shams original method, any system was approximated to being locally homogeneous. This allowed them to use the well-researched exchange-correlations of the homogeneous electron gas for the XC functional. The XC functional in this approximation, called the Local Density Approximation (LDA), is given as

$$E_{\text{XC}}^{\text{LDA}}[n] = \int d\mathbf{r}\, n(\mathbf{r}) \varepsilon_{\text{XC}}^{\text{LDA}}[n(\mathbf{r})] \tag{2.18}$$

where $\varepsilon_{\text{XC}}^{\text{LDA}}$ in Eq. (2.18) is the exchange-correlation energy density. This density is constructed as the sum of an exchange energy density, $\varepsilon_{\text{X}}^{\text{LDA}}$, and a correlation energy density, $\varepsilon_{\text{C}}^{\text{LDA}}$. The exchange energy density is known exactly and is given by the Dirac expression [64]:

$$\varepsilon_{\text{X}}^{\text{D}}[n] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} n^{1/3} \tag{2.19}$$

However, the correlation energy density must be approximated. The most popular of these approximations is the Perdew and Zunger (PZ) approximation [69]:

$$\varepsilon_{\mathrm{C}}^{\mathrm{PZ}}[n] = \begin{cases} A\ln(r_s) + B + Cr_s\ln(r_s) + Dr_s, & r_s \leq 1 \\ \gamma/(1+\beta_1\sqrt{r_s} + \beta_2 r_s), & r_s > 1 \end{cases} \tag{2.20}$$

In Eq. (2.20), $r_s = (3/4\pi n)^{1/3}$ is the mean interelectronic distance in atomic units and the coefficients A, B, C, D, γ, β_1 and β_2 are parameterised based on a quantum Monte Carlo simulation and can be found in literature such as [64].

The LDA XC functional works surprisingly well for some systems, and is still widely adopted due to its computational efficiency. In order to improve the XC functional approximation, non-locality can be included in the LDA expression. One way to include non-locality, is to modify the expression based on the gradient of the electron density. The class of such approximations to the XC functional, are called the Generalised Gradient Approximations (GGAs). Among the many GGAs developed over time, the one by Perdew, Burke and Ernzerhof (PBE) [70] is particularly successful, and its variant fitted specifically on experimental data of solids (PBES), will be utilised later in this thesis. This GGA is constructed such that it satisfies as many formal properties and physical limits as possible.

A common issue with both of these XC functionals, is that they fail to properly correct for the self-interaction error in the Hartree term of the Coulomb interaction, which appears due to the separation of the Coulomb energy in the Kohn-Sham formulation. This self-interaction error consists of the Hartree term erroneously including the interactions of the electrons with their own densities, which leads to Kohn-Sham DFT drastically underestimating band gaps of most materials. For this reason, a correction method was developed for the LDA and GGA functionals, called DFT-$^1/_2$, or LDA-$^1/_2$ and GGA-$^1/_2$ for the specific functionals [71]. This correction method consists of adding an atomic potential for each atom in the system, which is constructed as the difference between the neutral atom potential, and the potential of the atom with a small fraction of an electron missing. The DFT-$^1/_2$ method greatly improves DFTs ability to predict band gaps with the LDA and GGA XC functionals, as will be shown in Chapter 3.

The next, more sophisticated, non-local and computationally demanding, class of XC functional approximations, called the Meta-Generalised Gradient Approximations (MGGAs), include not only the gradient of the electron density, but also the Laplacian of the density. The SCAN functional is one such MGGA functional, which is built to satisfy all 17 known exact constraints that a MGGA functional can [72]. Unfortunately, SCAN suffers from the same inability as LDA and GGA to accurately predict band gaps. An alternative MGGA functional built specifically to predict band gaps, is TB09 developed by Tran and Blaha [73], in which the exchange potential is given by

$$v_x^{\mathrm{TB}}(\mathbf{r}) = cv_x^{\mathrm{BR}}(\mathbf{r}) + \frac{3c-2}{\pi}\sqrt{\frac{4\tau(\mathbf{r})}{6\rho(\mathbf{r})}} \tag{2.21}$$

where $\tau(\mathbf{r}) = {}^1\!/_2 \sum_{i=1}^{N} |\nabla\phi_i(\mathbf{r})|^2$ is the kinetic energy density, $\phi_i(\mathbf{r})$ is the i'th Kohn-Sham orbital, $v_x^{\mathrm{BR}}(\mathbf{r})$ is the Beckel-Roussel exchange potential [74] and the parameter c is calculated as

$$c = \alpha + \beta \left[\frac{1}{\Omega} \int_{\Omega} \frac{|\nabla\rho(\mathbf{r})|}{\rho(\mathbf{r})} d\mathbf{r} \right]^{\frac{1}{2}} \tag{2.22}$$

where α and β were determined by fitting a large array of experimental bandgaps and Ω is the volume of the unit cell. The correlation potential of TB09 is simply LDA correlation. TB09 accurately predicts a large array of band gaps, but is unsuitable for structural relaxations. To find an XC functional which is both suitable for geometric relaxations and for band gap predictions, it is necessary to use even more sophisticated functionals.

In principle, the accuracy of the XC functional approximations could continue to be improved in the same fashion, by including higher and higher order derivatives of the density. Yet in reality, the accuracy does not increase much by including higher order derivatives beyond the Laplacian. Instead, the next class of XC functional approximations utilises that the exchange part of the XC functional is, in fact, known exactly from Hartree-Fock theory and can be calculated from the Kohn-Sham orbitals. This class of approximations are called the Hybrid GGA functionals, due to being a hybrid between Kohn-Sham DFT and Hartree-Fock theory. However, simply including exact exchange in the DFT method turned out to reduce the accuracy of results, rather than improve it. The reason was found to be that, in less accurate XC functionals, a cancellation of errors occurred between the spatially truncated exchange and correlation terms. As such, by using exact exchange instead of a truncated exchange, this cancellation no longer occurred and the accuracy decreased. Despite this issue, an overall increase of accuracy could still be achieved by including only a fraction of the exact exchange

$$E_{\mathrm{XC}}^{\mathrm{Hybrid}} = \alpha E_{\mathrm{X}}^{\mathrm{Exact}} + (1-\alpha) E_{\mathrm{X}}^{\mathrm{GGA}} + E_{\mathrm{C}}^{\mathrm{GGA}} \tag{2.23}$$

where the fraction of exact exchange α in Eq. (2.23) depends on the specific hybrid functional. Of particular interest among the class of hybrid XC functionals, is the version developed by Heyd, Scuseria and Ernzerhof (HSE06), in which α usually takes on the value of ${}^1\!/_4$ [75]. In this hybrid functional, in order to alleviate problems of convergence in periodic systems, the exact exchange is split into a long-range and short-range part, and only the short range part is mixed into the XC functional. Hybrid functionals yield accurate band gaps due to the exact exchange, while they can still be used for structural relaxations, in contrast to the less accurate classes of XC functionals.

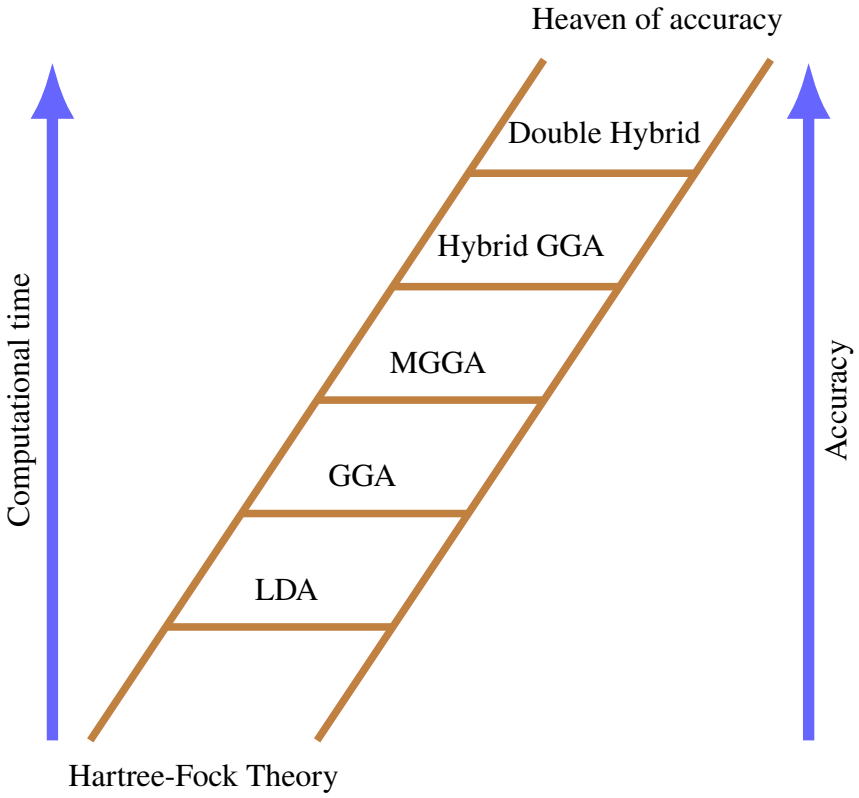


Figure 2.2: Jacobs ladder of exchange-correlation functional approximations in density functional theory.

Beyond the accuracy of hybrid functionals lies a final and most accurate class of functionals, called the double hybrid functionals. In these functionals, on top of mixing in a fraction of exact exchange, a fraction of exact partial correlation is also mixed into the functional. This class of functionals is also sometimes called "fully non-local", "generalised RPA" or simply "advanced".

This hierarchy of XC functional approximations in DFT is illustrated in the so-called Jacobs ladder in Figure 2.2. Beyond the examples of functionals in the various classes mentioned here, many more exists and can be found in the community library LIBXC [76].

2.2.4 Computational details

Besides the choice of the XC functional in a DFT calculation, many other parameters are also important, such as the k-space sampling of the Brillouin zone, the choice of pseudopotential to represent the potential of the lattice ions and the basis set in which to represent the Kohn-Sham orbitals. The most important of these parameters will be explained here.

When using Bloch's theorem, Eq. (2.4), to limit the simulation of a large system to a single unit cell, the first Brillouin zone of the system is represented on a chosen k-point grid. Due to the computational time of a DFT calculation scaling rapidly with the number of k-point grid-points, it is desirable to use as high a grid-spacing as possible without a substantial loss of accuracy. One method to heavily reduce the necessary amount of k-points, is through the utilisation of an occupation smearing function. The construction of the electron density, Eq.

(2.14), with Bloch's theorem, becomes

$$n(\mathbf{r}) = \sum_i \int_{\mathrm{BZ}} d\mathbf{k}\, f_{i\mathbf{k}} \, |\phi_{i\mathbf{k}}(\mathbf{r})|^2 \, \mathrm{d}\mathbf{k} \tag{2.24}$$

where $f_{i\mathbf{k}}$ is the occupation of the state $(i\mathbf{k})$. Since DFT is in principle strictly a ground state method, i.e. a 0 K temperature method, these occupations only take on the integer values of 0, if the state is unoccupied, and 1 if it is not. Thus, there is a sharp transition at the Fermi-level which requires an extensive amount of k-points to resolve. This issue can be alleviated, and the required amount of k-points reduced, by introducing a smearing function to smooth out the occupations near the Fermi-level. Due to its origin as the actual physical smearing function at higher temperatures, the most commonly used occupation function is the Fermi-Dirac distribution

$$f_{i\mathbf{k}} = \frac{1}{e^{(\epsilon_{i\mathbf{k}}-\mu)/\sigma} + 1} \tag{2.25}$$

where $\sigma = k_{\mathrm{B}}T$ is the thermal broadening, k_{B} is the Boltzmann constant, μ is the chemical potential and $\epsilon_{i\mathbf{k}}$ is the energy of the state. It is important to stress, that when the Fermi-Dirac function in Eq. (2.25) is used as a smearing function in DFT, the temperature T is only a fictitious temperature and not the actual temperature of the system, as DFT is still a zero-temperature ground-state method. There exist many other smearing functions, such as cold smearing [77], which is particularly useful for metals, wherein the fictitious temperature T can no longer be mistaken for a physical temperature. If features close to the Fermi level are critical for a particular system, the fictitious temperature T needs to be converged in conjunction with the k-point grid.

To calculate the reference potential with Eq. (2.17), the potential from the lattice ions in v_{ext} and the Hartree potential is firstly found by solving the Poisson equation [64]:

$$\nabla^2 v_{\mathrm{es}}(\mathbf{r}) = -4\pi n(\mathbf{r}) \qquad v_{\mathrm{es}}(\mathbf{r}) = v_{\mathrm{ion}}(\mathbf{r}) + \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{2.26}$$

Eq. (2.26) is normally solved numerically on a Cartesian grid. The fidelity of the grid $\mathrm{d}x$, is represented by an energy, $\mathrm{d}x = \pi/\sqrt{E_{\mathrm{cut}}}$, called the density-mesh cutoff E_{cut}. To ensure that the grid is fine enough for a given DFT calculation, the density-mesh cutoff should in general be converged, by running multiple DFT calculations with increasing E_{cut} until the property of interest is satisfactorily unchanged by the increase in E_{cut}.

The Poisson equation in Eq. (2.26) is a second-order differential equation and therefore requires two boundary conditions to be solved. When using Bloch's theorem to simulate a large system, the natural choice is periodic boundary conditions for both the potentials and their derivatives. Other choices include Neumman and Dirichlet boundary conditions, which are particularly useful for calculations of thin materials or slabs.

The real potential from the ions, v_{ion}, diverges close to their nuclei, this is an issue for the numerical solution of the Poisson equation, and the real potentials are therefore replaced by smooth pseudopotentials in practice. These smooth pseudopotentials are constructed such that they yield the same energy as the real potentials.

To find the single-particle Kohn-Sham orbitals with the Kohn-Sham equations, $\hat{H}_R\phi_n(\mathbf{r}) = \epsilon_n\phi_n(\mathbf{r})$, the orbitals are expanded in a chosen basis set $\{\psi_i\}$

$$\phi_n(\mathbf{r}) = \sum_i c_{ni}\psi_i(\mathbf{r}) \tag{2.27}$$

and the expansions coefficients c_{ni} are calculated. Due to their simplicity and systematic improvability, a natural choice of basis set is the plane waves: $\psi_{\mathbf{G}}(\mathbf{r}) = e^{i\mathbf{G}\cdot\mathbf{r}}$. Since the plane waves are periodic, this choice of basis set is particularly efficient for large and periodic systems. Because there exists an infinite number of plane waves, in practice such a basis set is truncated to only include plane waves with kinetic energies below a cutoff energy E_{cut} [64]:

$$\frac{\hbar^2}{2m}|\mathbf{G}|^2 < E_{\text{cut}} \tag{2.28}$$

To ensure that enough plane waves are included in the basis set to properly represent the Kohn-Sham orbitals, this cutoff energy, which is different from the one utilised for the Cartesian grid spacing in the Poisson equation, should be converged by running a convergence series until a chosen convergance criteria has been reached. In DFT, energies such as E_{cut} are conventionally given in Hartree (Ha), which is defined as:

$$1\,\text{Ha} = m_e\left(\frac{e^2}{4\pi\varepsilon_0\hbar}\right)^2 \approx 27\,\text{eV} \tag{2.29}$$

An often more computationally efficient choice of basis set, however, are non-orthogonal localised atomic orbitals. Localised basis sets are often more efficient, because fewer basis functions can be used to represent the Kohn-Sham orbitals with the same degree of accuracy. This is the case, because localised basis sets can more easily reproduce rapid fluctuations close to the nuclei and because they ignore the "empty" space far from the lattice ions. Unfortunately, the non-orthogonal localised basis sets do not form a complete set, and as such there is no systematic method to improve them, as simply adding more basis functions can lead to issues with over-completeness. Therefore, non-orthogonal localised basis sets have to be carefully constructed and validated against well-known parameters [64].

2.3 Non-equilibrium green's functions

Since DFT is strictly an equilibrium ground state method, it is inapplicable to simulate open-end non-equilibrium systems, such as systems in steady-state under an applied bias voltage. To simulate such systems atomistically, an ideal method is the Non-Equilibrium Green's Function (NEGF) method, which is based on Green's function theory [78–80]. This theory originates from the mathematical Green's function method of solving differential equations:

Suppose that $\hat{\mathcal{L}}$ is a linear differential operator, such as the Hamiltonian of a system, with corresponding differential equation:

$$\hat{\mathcal{L}} f(x) = g(x) \tag{2.30}$$

Suppose again that Eq. (2.30) needs to be solved for many different functions $g(x)$, some of which are difficult to solve, it can then prove easier to solve the following differential equation instead

$$\hat{\mathcal{L}} G(x, s) = \delta(x - s) \tag{2.31}$$

where the function $G(x, s)$ is called the Green's function. If the Green's function of this differential operator is found, then the solutions of Eq. (2.30) for any $g(x)$ can be found simply by integration:

$$f(x) = \int G(x, s) f(x) ds \tag{2.32}$$

For this reason, the Green's function is also called the fundamental solution to the differential operator.

As such, to solve the Schrödinger equation for any perturbation potential $g(x)$, which could for example be the potential from an applied bias, one simply has to calculate the Green's function, which is most simply done by inversion of the Hamiltonian

$$[E - \hat{H}(\mathbf{r})] G(\mathbf{r}, \mathbf{r}', E) = \delta(\mathbf{r} - \mathbf{r}') \qquad G(\mathbf{r}, \mathbf{r}', E) = \frac{1}{E \pm i\eta - \hat{H}(\mathbf{r})} \tag{2.33}$$

where $i\eta$ in Eq. (2.33) is a complex infinitesimal, added to circumvent issues at the poles of the Hamiltonian, i.e. the eigenenergies.

While the purpose of the Green's functions in the NEGF method is to simulate biased steady-state systems, they also form the foundation of one of the most accurate atomistic many-body simulation methods called the "GW method". In the GW method, the Green's function of the complicated many-body Hamiltonian is found through a perturbative expansion in the simple non-interacting single-particle Green's functions, with the screened Coulomb interaction as the perturbation [78, 81]. Despite the high accuracy of the GW method, it was abandoned in favour of DFT for the research in this thesis, due to its prohibitively large computational requirement.

In order to be able to calculate the Green's function with Eq. (2.33), it is necessary to know the Hamiltonian in an invertible representation. As such, the NEGF method is often used in conjunction with an atomistic method, such as tight-binding or DFT, which

can produce the necessary Hamiltonian. For the work in this thesis, the NEGF method was used with DFT, for the sake of consistency and accuracy, and the method would be aptly named NEGF-DFT. In the NEGF method, the open-ended system is modelled by splitting the system and its Hamiltonian into three parts: Left and right semi-infinite electrodes, which are assumed to be in equilibrium and to not interact with one another, and a finite central region which connects the two electrodes. In the central region, the system is in general in non-equilibrium and incoming electrons from the electrodes can be scattered on defects, interfaces or other scattering centres. The full Hamiltonian of the system can therefore be written as

$$\hat{H} = \begin{pmatrix} \mathbf{H}_L & \mathbf{V}_L & 0 \\ \mathbf{V}_L^\dagger & \mathbf{H}_C & \mathbf{V}_R \\ 0 & \mathbf{V}_R^\dagger & \mathbf{H}_R \end{pmatrix} \tag{2.34}$$

where $\mathbf{H}_L$, $\mathbf{H}_C$ and $\mathbf{H}_R$ are the Hamiltonian matrices of the left, central and right regions respectively and $\mathbf{V}_L$ and $\mathbf{V}_R$ are the coupling between the central region and the left and right electrodes. The Hamiltonians of the left and right semi-infinite electrodes, are calculated from fully periodic bulk DFT calculations of the electrode unit cells. In order for the couplings $\mathbf{V}_L$ and $\mathbf{V}_R$ to properly couple the central region and the bulk equilibrium electrodes, it is necessary to include part of the semi-infinite electrodes in the central region, such that the potential from the centre of the central region is fully screened at the interface to the electrodes. An illustration showing this splitting of the open-ended system into the three parts of the NEGF formalism, is shown in Figure 2.3.

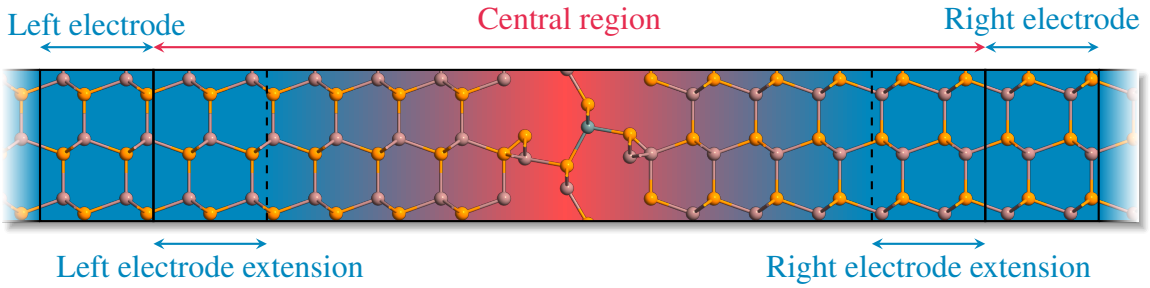


Figure 2.3: Open-ended system in NEGF formalism. The systems three parts are shown: The central scattering region and the two semi-infinite left and right electrodes. The electrode extensions in the central region are also shown and indicated by the dashed line. The background colour represents the perturbation potential from the defect in the central region, which is shown to be completely screened in the electrode extensions.

In order to find the transmission coefficients, from which the current through the system in Figure 2.3 can be calculated, and the steady-state non-equilibrium electron density, from which many other important properties can be derived, only the Green's function of the central region is necessary, which is calculated as

$$\mathbf{G}_C(E) = \frac{1}{[E \pm i\eta]\mathbf{S} - \mathbf{H}_C - \mathbf{\Sigma}_L(E) - \mathbf{\Sigma}_R(E)} \tag{2.35}$$

where $\mathbf{S}$ in Eq. (2.35) is the overlap matrix, with elements $S_{ji} = \langle \phi_j | \phi_i \rangle$, i.e. it only differs

from the identity matrix for non-orthogonal basis sets. The matrices $\mathbf{\Sigma}_L$ and $\mathbf{\Sigma}_R$ are the left and right electrode self-energies, which accounts for the coupling of the central region to the two semi-infinite electrodes, and are given by

$$\mathbf{\Sigma}_L(E) = \mathbf{V}_L \mathbf{G}_L(E) \mathbf{V}_L^\dagger \qquad \mathbf{\Sigma}_R(E) = \mathbf{V}_R \mathbf{G}_R(E) \mathbf{V}_R^\dagger \tag{2.36}$$

where the left and right electrode Green's functions $\mathbf{G}_L$ and $\mathbf{G}_R$ are calculated as in Eq. (2.33), with the left and right electrode Hamiltonians respectively.

The non-equilibrium electron density is given as the sum of the left and right contributions, where the left contribution is calculated with the central region Green's function as

$$n_L = \frac{1}{2\pi} \int \mathbf{G}_C(E) \mathbf{\Gamma}_L(E) \mathbf{G}_C^\dagger(E) f\left(\frac{E - \mu_L}{k_B T_L}\right) dE \tag{2.37}$$

and similarly for the right contribution. In Eq. (2.37), T_L and μ_L is the electron temperature and chemical potential of the Fermi function f for the electron distribution in the left electrode, $\mathbf{\Gamma}_L = [\mathbf{\Sigma}_L - \mathbf{\Sigma}_L^\dagger]/i$ is the broadening function of the left electrode and the product $\mathbf{G}_C \mathbf{\Gamma}_L \mathbf{G}_C^\dagger = \mathbf{A}_L$ is called the spectral function. The integral in Eq. (2.37) is split into an equilibrium and non-equilibrium part; the equilibrium part is calculated by the complex contour integration method, whereas the non-equilibrium part is calculated along the real axis.

The current through the system resulting from an applied bias voltage, is calculated from the transmission spectrum with the Landauer equation

$$I(V_{\text{bias}}) = \frac{2e}{h} \int T(E) \left[f\left(\frac{E - \mu_R}{k_B T_R}\right) - f\left(\frac{E - \mu_L}{k_B T_L}\right) \right] dE \tag{2.38}$$

where the chemical potential of the left electrode is shifted by the applied bias $\mu_L = E_L^F - eV_{\text{bias}}/2$, and similarly the chemical potential of the right electrode is shifted in the opposite direction $\mu_R = E_R^F + eV_{\text{bias}}/2$. The Fermi levels of both electrodes, and corresponding eigenenergies, are shifted to coincide at zero bias, to ensure electrical equilibrium. In Eq. (2.38), the factor of two accounts for both spin-channels, as spin-polarised transport is not considered in this thesis. The transmission spectrum $T(E)$ occurring in the Landauer formula, can, in the NEGF-DFT formalism, be calculated with the central region Green's function:

$$T(E) = \text{Tr}\left[\mathbf{G}_C(E) \mathbf{\Gamma}_L(E) \mathbf{G}^\dagger(E) \mathbf{\Gamma}_R(E)\right] \tag{2.39}$$

Due to the combination of the NEGF method with DFT, the computational considerations of DFT described in Section 2.2.3 and 2.2.4, such as the exchange-correlation functional or the grid finesses, carry over to the NEGF-DFT method. Due to the finite length of the central region of the NEGF system, as shown in Figure 2.3, Bloch's theorem is only applied in the directions transverse to the current transport. Additionally, in the transport direction of the system, the boundary conditions of the Poisson equation are required to be periodic in the electrodes, to simulate the semi-infinite bulks, and either Neumann or Dirichlet in the

central region, to match the potentials at the interface to the two electrodes. One important computational parameter unique to NEGF-DFT, is related to how the integral in Eq. (2.37) is carried out. For the complex contour integration, the integral can be solved using the Residue Theorem, in which the number of included poles have to be truncated to a finite value. Besides this finite number of included poles, the infinitesimal η included in the Green's function, is another important parameter which needs to be converged.

In the two following chapters of this thesis, the theory and methods described in this chapter will be utilised to study defects in the materials constituting the PIN photodetector described in the previous chapter.

Chapter 3

First-Principles Investigation of Polytypic Defects in InP

The studies and calculations presented in this chapter, build upon the research that was published in Ref. [82], but also include additional material, which were never condensed and written into its own article during the time of writing of this thesis. The chapter is split into six sections, the first of which is an introduction and motivation of the particular research of this chapter. In the second section, the mandatory convergence study of the numerical DFT parameters, described in Section 2.2.3 and 2.2.4, is carried out and the two bulk crystal phases of indium phosphide are also described and studied here. Then in the third section the polytypic defect systems are introduced and the non-analytic contribution to their phonon dispersions are studied. The fourth section deals with the optical properties of these polytypic systems, which includes calculating the complex dielectric functions of them, as well as related quantities, and investigating the treatment of polytypic defects as atomically thin quantum wells. In the fifth section, the impact on the conductivity and mobility of these polytypic defects are studied, and in the sixth and final section of this chapter, the mains results and conclusions are summarised.

For the DFT calculations of this chapter, the state-of-the-art simulation software suite QuantumATK by Synopsys version T-2022.03 was used [83]. The utilised basis set was the local atom-centred non-orthogonal PseudoDojo basis set, which was employed together with its corresponding pseudopotential [84]. For the NEGF transport calculations, the open-ended 2-probe method described in Section 2.3 was utilised, as implemented in QuantumATK [80, 85]. The remaining non-default simulation parameters will be given in the convergence study in Section 3.2.2.

3.1 Prelude

In addition to being used for the doping regions in the PIN photodiode described in Section 1.4, due to its lattice match with $In_{0.53}Ga_{0.47}As$ and higher band gap, indium phosphide is also used for many other technologically important devices. It is used in conjunction with $In_{0.53}Ga_{0.47}As$ in Heterojuction Field Effect Transistors (HFETs), in order to combine the

high mobility of $In_{0.53}Ga_{0.47}As$ at low voltage, with the high drift velocity and low impact ionisation of InP at high voltages [33, 86]. It is also used for electro-optic modulators, lasers, diodes, optical amplifiers and many more devices [40, 87], usually in conjunction with other III-V materials. InP is so extensively utilised due to being the substrate of choice for many applications, and in particular for high-speed radio-frequency applications, due do its high mobility, and optical applications above 1000 nm, owing to its large band gap.

In particular, indium phosphide is a popular advanced material platform for Integrated Photonic Circuits (IPCs). Similar to the field of integrated electronic circuits, the goal of the research field of IPCs is to make photonic systems cheap, easily available and ubiquitous. Great progress has been made towards enabling the InP platform for IPCs, as all the critical components of IPCs, including both passive and active components, such as lasers, amplifiers, photodetectors and modulators, can currently be integrated on InP [88]. The similarities between IPCs and integrated electronic circuits also extend to Moore's law, as IPCs have showed a similar miniaturisation trend in their early years. This component density growth halted to almost a complete stop, due to the low optical confinement limiting further shrinking of the waveguides [89]. A possible solution, which has recently been proposed, is the integration if InP membranes around the critical components [90]. These membranes would have a similar role to that of the silicon oxide in Silicon On Insulator (SOI) devices, which is, they would improve the confinement by isolating the device from its surroundings [91].

Evidently, indium phosphide is a very technologically relevant material, and its high-quality defect-free growth is clearly paramount to many modern applications, and in particular to IPCs. Unfortunately, high-quality crystalline growth of InP is hindered by its inherent polytypism, which causes surface defects, specifically stacking faults and Rotational Twin Planes (RTPs), to form during its growth [92, 93]. These defects are difficult to avoid, as they occur due to the low energy separation between the stable zincblende phase ground state of InP, and its meta-stable wurtzite phase. In Figure 3.1, two examples of grown InP are shown, inspected with High-Resolution Scanning Transmission Electron Microscopy (HR-STEM) from the $\langle 1\,1\,0\rangle$-direction. This particular direction of view allows visualisation of the stacking sequence along the $\langle 1\,1\,1\rangle$-direction, which is the easiest method to visualise RTPs. Figure 3.1A shows a near-pristine sparsely twinned zincblende lattice and Figure 3.1B shows a densely twinned highly intermixed zincblende/wurtzite lattice, and even a line defect appearing to be oriented along the $\langle 1\,1\,0\rangle$-direction is shown to distort the crystalline lattice. Clearly, the formation of RTPs is normally erratic and they can occur in high or low densities throughout the material.

It is possible to influence the polytypism of InP through growth control, which has allowed growth of InP in its meta-stable wurtzite phase [94]. This control is mostly limited to small dimensions, such as in nanowires, where growing InP along a specific direction has shown to completely eliminate RTPs [22]. Perfect control of InP polytypism and RTP formation, can potentially be useful for novel devices based on crystal-phase engineering, due to the specific properties of RTPs, which are not always detrimental. For example, RTPs in gallium nitride are suspected to work as atomically thin, shallow quantum wells, which could be used for

optical sensors [95–99]. Similarly, RTPs have possibly beneficial properties in InP, as Bao et al. [100] showed, using TEM and micro-photoluminescence, that RTPs in InP cause an excitation-power dependent blue-shift.

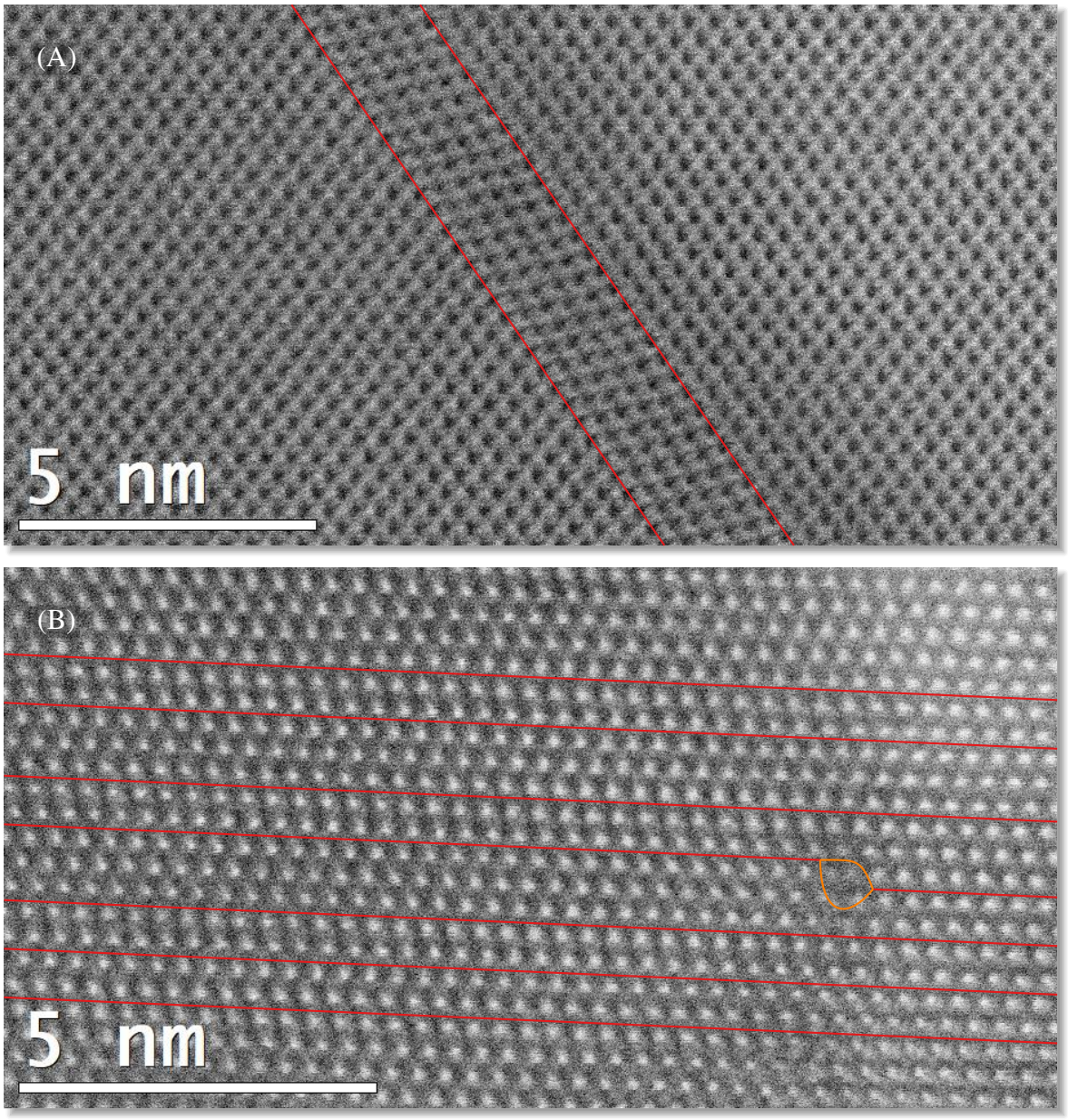


Figure 3.1: High-resolution scanning transmission electron microscopy images of bulk InP, viewed from the ⟨1 1 0⟩-direction which allows visualisation of the stacking sequence along the ⟨1 1 1⟩-direction. Rotational twin planes are indicated with red lines and the line defect by an orange oval. (A) A near-perfect crystalline zincblende lattice with only a few RTPs. (B) A highly polytypic mixed zincblende and wurtzite lattice with many RTPs/stacking faults and even a line defect going into the image.

Clearly, the study of the importance of RTPs on the performance of optical and electrical devices incorporating InP are highly relevant, not only to the InP/InGaAs/InP PIN photodiode of Section 1.4, but to many different devices and even the entire field of IPCs. It is even possible, that entirely new devices, based on crystal-phase engineering, can be discovered by the study of RTPs and polytypic defects. As such, the goal of the present thesis chapter is to study the effect of RTPs and polytypic defects on the electrical and optical properties of InP, and to discover if they can potentially improve the performance of devices.

3.2 Zincblende and Wurtize indium phosphide

In order to study polytypic defects in InP, it is first necessary to have a general comprehension of the two bulk crystal phases, which is the purpose of the present section. Firstly, the crystallographic structure of the phases will be detailed, after which the mandatory convergence study of the DFT accuracy parameters from Section 2.2.3 and 2.2.4 will be carried out. Lastly, the optical and phonon band structures of the two phases will be investigated, together with the nature of the electronic transmission.

3.2.1 Crystallographic structure

InP is a direct band gap III-V semiconductor which usually crystallises in the diamond-like Zincblende (ZB) phase, yet in recent years it has also become possible to stabilise moderately-sized bulks of InP in its meta-stable lonsdaleite-like Wurtzite (WZ) phase [94]. Zincblende, like diamond, consists of two interwoven Face-Centered Cubic (FFC) structures, which can be described by a single FCC structure with a basis of $(0, 0, 0)$ and $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and as such has a minimal unit cell consisting of two atoms. Zincblende differentiates from the diamond structure by having the two FCC lattices occupied solely by elements of opposite polarity, which are indium and phosphor in the case of InP. Zincblende InP has a direct band gap of 1.34 eV, a lattice constant of 5.87 Å and symmetry space group 216, i.e. $T_d{}^2$ or F-43m symmetry, in Schoenflies or Hermann-Mauguin notation respectively [101, 102]. It is illustrated from various directions in Figure 3.2, where Figure 3.2a shows its ABC stacking sequence along the $[1\,1\,1]$-direction, as viewed from the $[1\,\bar{1}\,0]$-direction, which is the same direction by symmetry as the STEM image in Figure 3.1, and direct comparison can therefore aid to analyse the STEM image. In Figure 3.2b Zincblende InP is shown from the $[1\,1\,1]$-direction, which shows its similarity to a hexagonal structure. The unit cell is also displayed in this figure, illustrated by the dotted line. In Figure 3.2c, the 6-atom $[1\,1\,1]$-stacking ABC supercell is shown, rather than the conventional cubic cell or minimal unit cell, for easier comparison with the minimal unit cell of Wurtzite.

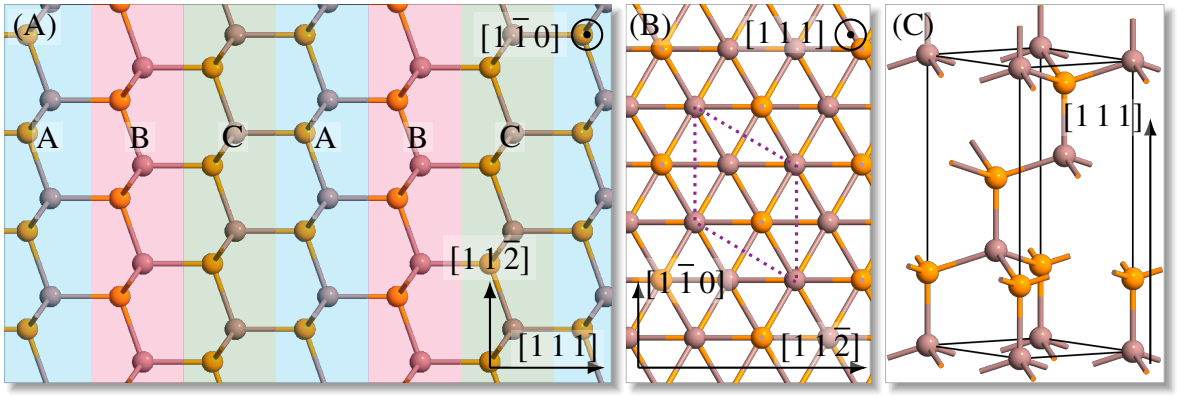


Figure 3.2: Zincblende structure of InP viewed from corresponding directions to the Wurtzite structure in Figure 3.3. (A) Viewed from the $[1\,\bar{1}\,0]$-direction, which reveals the ABC stacking sequence along the $[1\,1\,1]$-direction. (B) Viewed from the $[1\,1\,1]$-direction, where the unit cell is shown by the dotted line. (C) The 6-atom ABC stacking sequence supercell.

In the same way that Zincblende is similar to the diamond structure, Wurtzite is similar to the lonsdaleite structure, i.e. hexagonal diamond, with the elemental species of opposite polarity occupying separate sub-lattices. The sub-lattices in question are here Hexagonal Close-Packed (HCP) structures, rather than FCC as for Zincblende, and they are interwoven such that the second sub-lattice is occupying half the tetrahedral voids, i.e. interstitials, of the first sub-lattice. Wurtzite InP also has a direct band gap, which is slightly larger than that of Zincblende, at 1.41 eV [94], and the hexagonal lattice constants $a = 4.14$ Å and $c = 6.80$ Å [103]. Since Wurtzite consists of two HCP lattices, the minimal unit cell contains four atoms and the structure has symmetry space group 168, i.e. $C_{6v}{}^4$ or $P6_3mc$ in Schoenflies or Hermann-Mauguin notation, respectively. The Wurtzite structure is shown in Figure 3.3 from the hexagonal directions corresponding to the directions for the Zincblende structure in Figure 3.2. In Figure 3.3a is shown its ABAB stacking sequence, it can be seen here that the B-layer in Wurtzite is identical to the layers of the Zincblende structure, shown in Figure 3.2a, whereas the A-layer is dissimilar. In Figure 3.3b, the Wurtzite structure is shown from the $[0\,0\,0\,1]$-direction, where the hexagonality of the structure is clearly illustrated. The minimal unit cell is also shown by the dotted line, and is seen to be identical, from this direction, with the 6-atom unit cell of Zincblende shown in Figure 3.2b. In Figure 3.3c is shown the minimal 4-atom unit cell of the Wurtzite structure, which by comparison with the 6-atom unit cell of Zincblende in Figure 3.2c, is seen to be identical in the first layer, but different in the second layer and lacking a third layer.

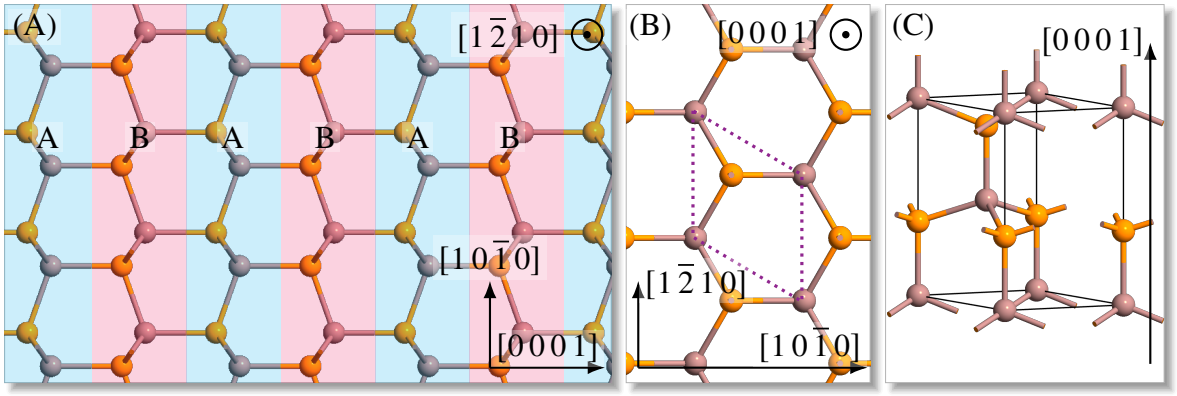


Figure 3.3: Wurtzite structure of InP viewed from corresponding directions to the Zincblende structure in Figure 3.2. (A) Viewed from the [1 $\bar{2}$ 1 0]-direction, wherein the ABAB stacking sequence along the [0 0 0 1]-direction can be seen. (B) Viewed from the [0 0 0 1]-direction, where the minimal 4-atom unit cell is shown by the dotted line. (C) The minimal 4-atom unit cell.

3.2.2 Convergence study

In order to use DFT to study the properties of these two phases of InP, it is first required to find the optimal DFT parameter set with which to simulate them, and to check that the most critical parameters are sufficiently converged.

The two DFT parameters which usually have the largest impact on the simulation results, are the chosen exchange-correlation functional and the basis set utilised to represent the Kohn-Sham orbitals. The parameters whose convergence have to be checked, in order to avoid nonsensical results, are the Brillouin zone k-point sampling, the occupation function temperature and the real-space density-mesh cutoff. These parameters were all explained in detail in Sections 2.2.3 and 2.2.4. To determine successful convergence, not only the total energy per atom is checked, but also the lattice constant and band gaps, and these are all checked for both crystallographic phases. For both phases, the minimal unit cell, with the experimental lattice constants, is utilised and relaxed, until the interatomic forces are lower than 0.05 eV Å^{-1} and the stress is lower than 0.1 GPa. The relaxations are constrained to maintain the Bravais lattice, but are otherwise free.

The first and most basic parameter to converge, is the Brillouin zone k-point sampling density. The remaining convergence parameters are initially kept at their default values, which is an occupation function temperature of 1000 K and a density-mesh cutoff of 60 Ha. The default exchange-correlation functional, PBE, is replaced with PBES, which is merely a more appropriate version of PBE, as it is fitted specifically to solids. For the basis set, the more accurate "High" version of the PseudoDojo basis set is used, which is merely an extension of the default "Medium" set by a few extra basis functions which tends to increase the accuracy of the DFT calculation. This assumption will be validated after the convergence checks, to also ensure that the larger basis set does not lead to issues with over-completeness.

The convergence is plotted in Figure 3.4, where the absolute values are shown with empty markers, and the values relative to the most accurate calculation are shown with filled markers.

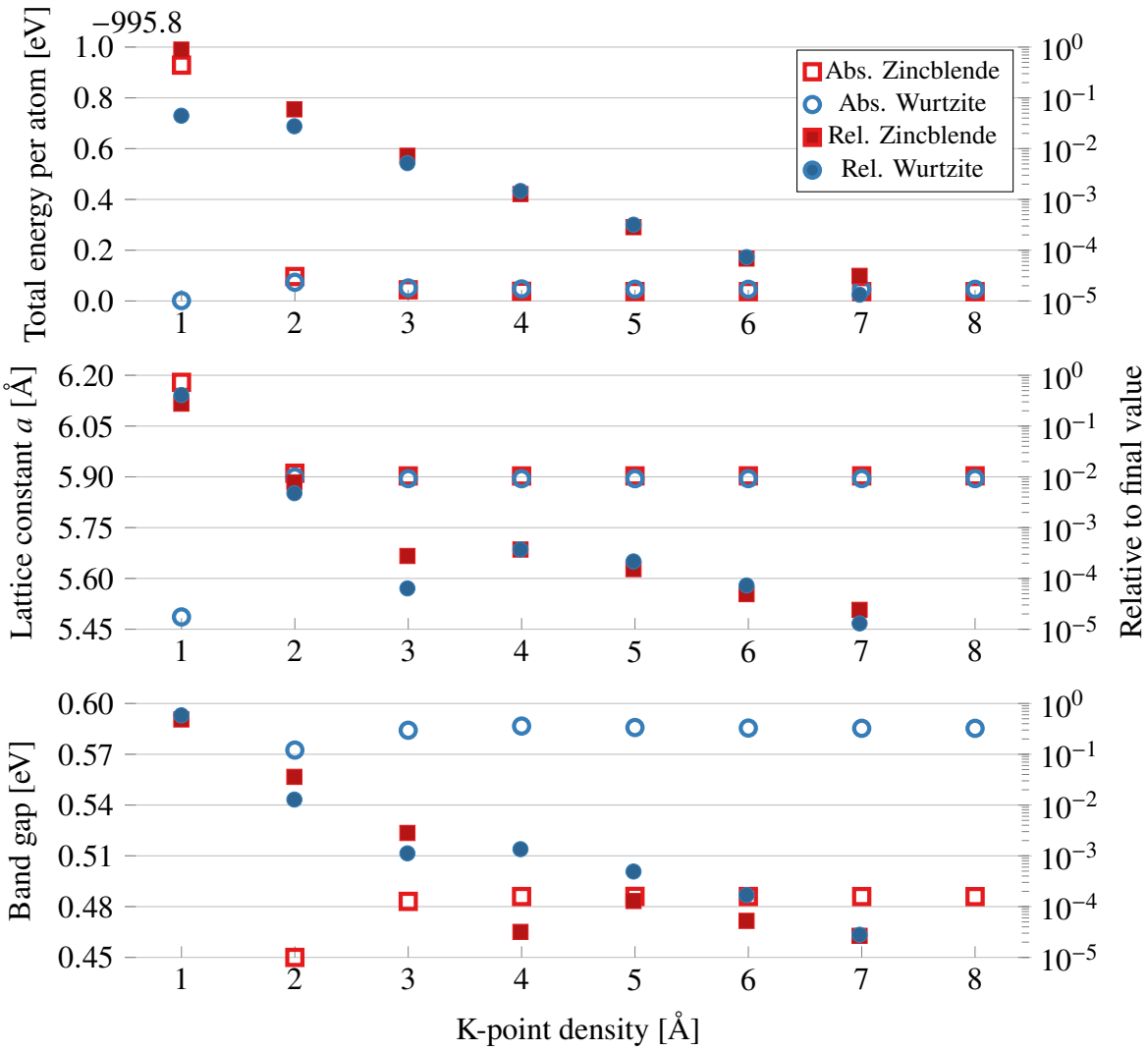


Figure 3.4: Convergence test of Brillouin zone k-point density sampling in both Zincblende InP and its meta-stable Wurtzite phase. Occupation function temperature was 1000 K. Filled symbols indicate values taken relative to the final most accurate value, whereas empty symbols represent the absolute values. The absolute values of the total energies were shifted by −995.8 eV to facilitate easier reading. The Wurtzite lattice constant is the FCC-equivalent distance, rather than an actual lattice constant, in order to allow for direct comparisons.

It can be seen that a total energy or band gap convergence tolerance in the meV range, requires a k-point sampling density of between 4 Å and 6 Å. Kohn-Sham DFT without any correction methods is notoriously bad at reproducing exact band gaps, as such, the band gaps shown in Figure 3.4 lies well within expectations for the GGA exchange-correlation functional. While still being far from the experimental values, the difference between the Wurtzite and Zincblende band gaps are still fairly well reproduced, at a 100 meV rather than the experimental 70 meV. To produce more accurate band gaps, one of several correction methods can be used, or a more accurate exchange-correlation functional. This will be explored in the end of this convergence section. Care should be taken when considering the lattice constants in Figure 3.4, as the Wurtzite lattice constants are the FCC-equivalent distances, rather than actual lattice constants, for a more direct and straightforward comparison with the Zincblende lattice constants.

The next parameter to converge is the occupation function temperature, which is not a real temperature, but rather just the width of a smearing function used to avoid numerical issues with the otherwise sharp occupation shift at the Fermi level. Since the occupation function temperature is related to the k-point sampling, a high k-point sampling density of 8 Å is utilised, in order to ensure that it remains converged with decreasing occupation function temperatures. This second convergence study is shown in Figure 3.5, where the remaining parameters were left unchanged in the calculations. Unlike the k-point sampling density, the total energy is not a monotonically decreasing function of the occupation function temperature. In fact, despite that some convergence behaviour can be seen from the plots relative to the final value, considering the scale of the values, the system appears converged at all the considered parameter values. Despite this apparent small effect of the occupation function temperature on bulk properties, certain properties which depends critically on the structure close to the Fermi level can be significantly affected. As such, to err on the side of caution, an occupation function temperature of 300 K is therefore used for future calculations.

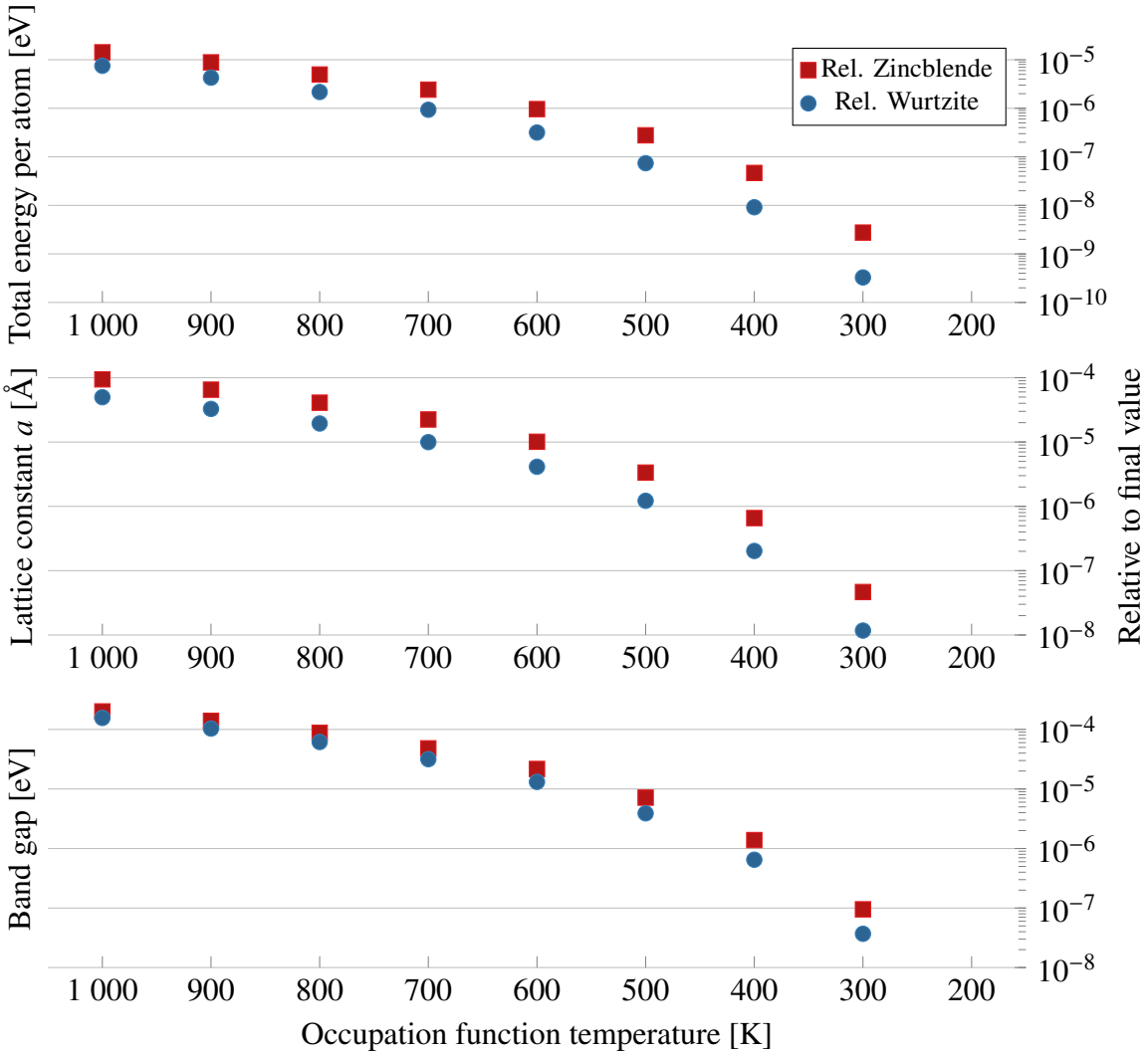


Figure 3.5: Convergence test of occupation function temperature in both Zincblende InP and its meta-stable Wurtzite phase. The Wurtzite lattice constant is the FCC-equivalent distance, rather than an actual lattice constant, in order to allow for direct comparisons. The absolute values are not plotted due to their indiscernible flatness for the small variations here.

The use of a lower occupation function temperature leads to a sharper occupation at the Fermi level, which in turn usually requires a higher k-point sampling to properly resolve. This relationship necessitates another convergence study of the k-point sampling density, using the new occupation function temperature of 300 K. Inspecting this convergence study in Figure 3.6, it is seen that despite this usual relationship, the k-point sampling convergence is unchanged by the change in the occupation function temperature, except for the band gaps where a k-point sampling density of 8 Å is now required to reach a tolerance of 0.1 meV rather than 6 Å. Thus the k-point sampling density of 8 Å remains sufficient.

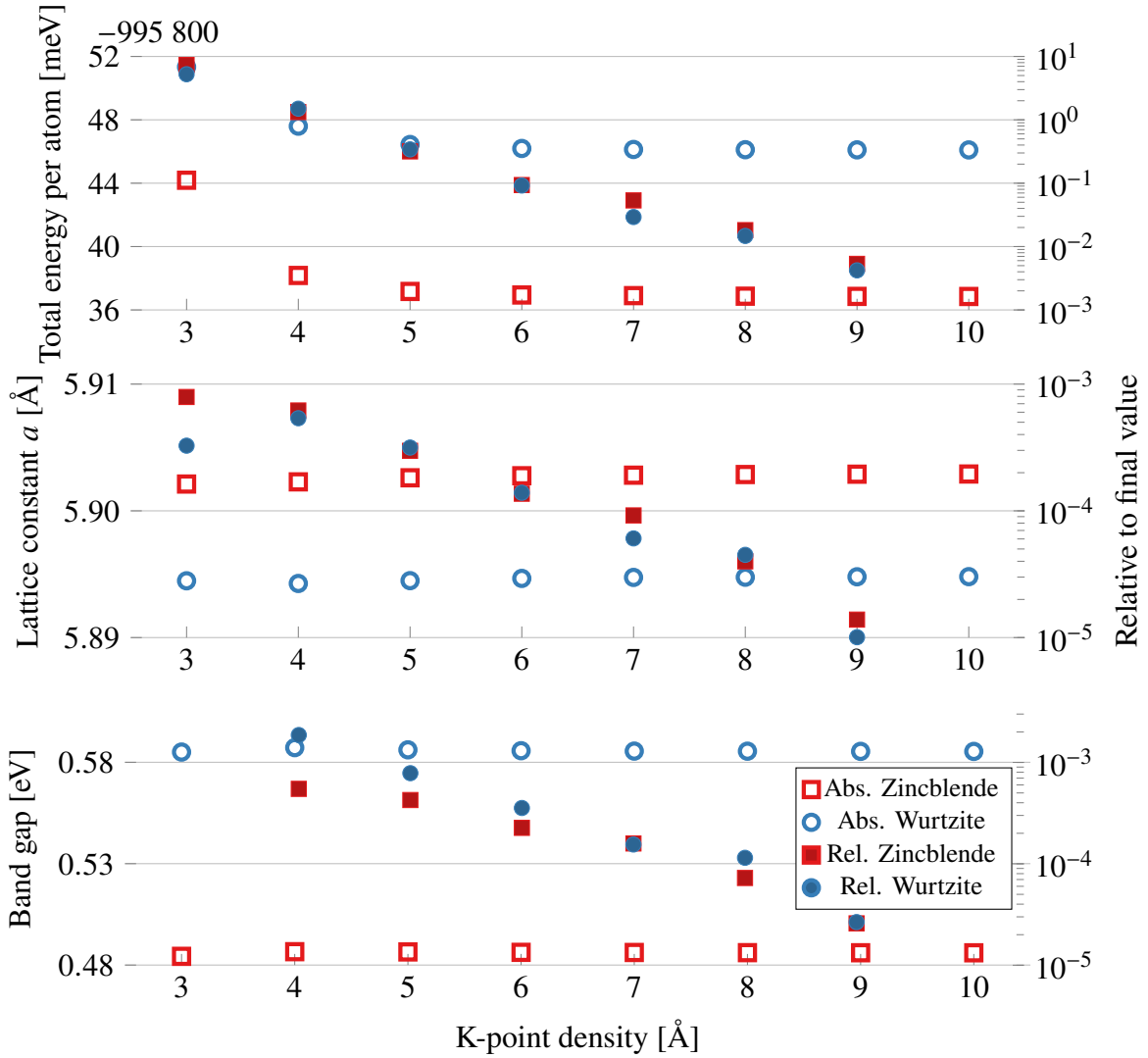


Figure 3.6: Convergence test of Brillouin zone k-point density sampling in both Zincblende InP and its meta-stable Wurtzite phase. Occupation function temperature was 300 K. Filled symbols indicate values taken relative to the final most accurate value, whereas empty symbols represent the absolute values. The absolute values of the total energies per atom were shifted by −995.8 eV to facilitate easier reading. The Wurtzite lattice constant is the FCC-equivalent distance, rather than an actual lattice constant, in order to allow for direct comparisons.

The final parameter to converge is the density-mesh cutoff, whose convergence is shown in Figure 3.7. It is seen that the previously used default value of 60 Ha is insufficiently converged, when requiring a tolerance of 1 meV in the total energy and band gaps. Instead a minimum of 75 Ha is needed, and preferably 95 Ha, to achieve an accuracy of 0.1 meV.

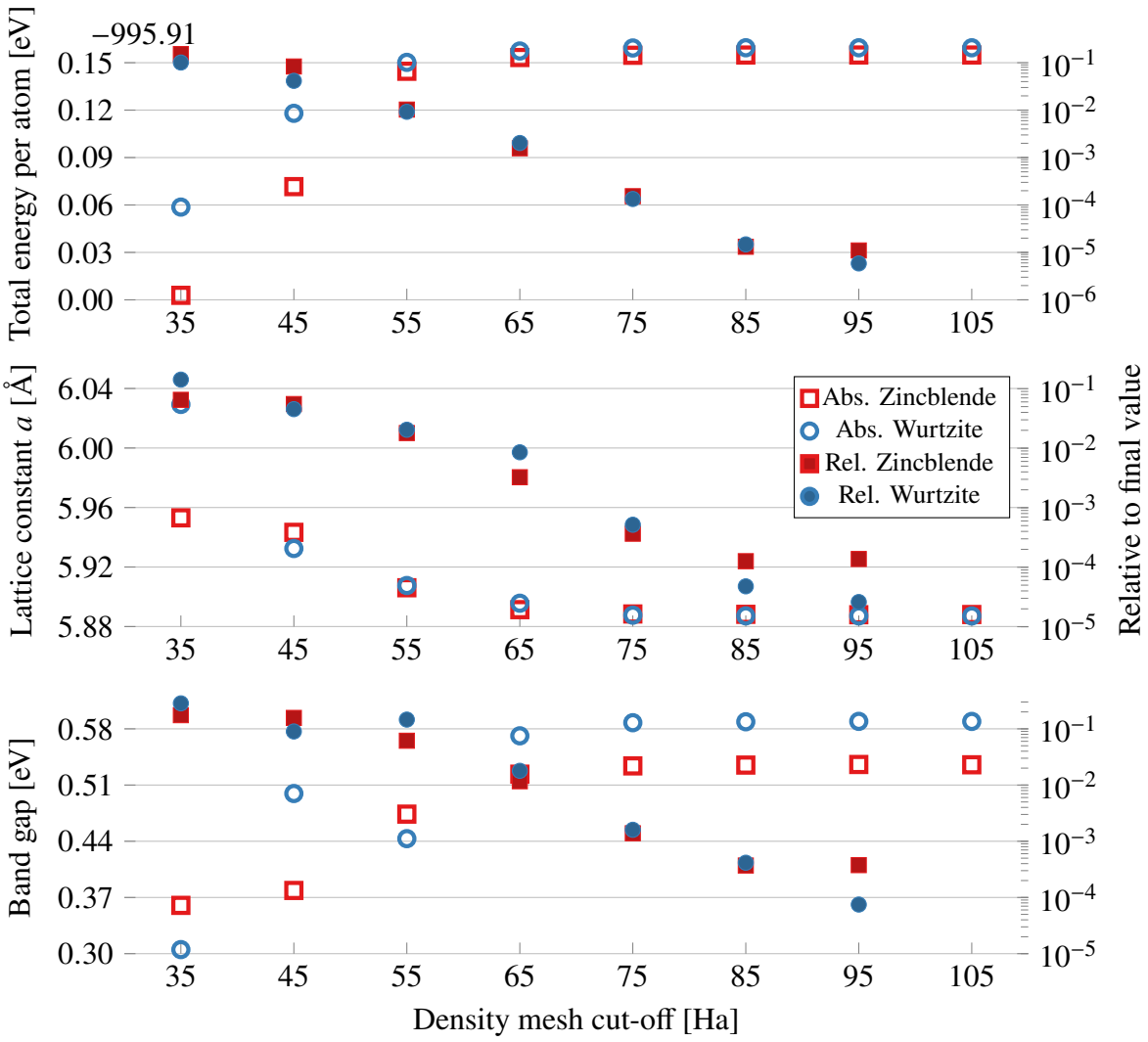


Figure 3.7: Convergence test of density-mesh cutoff in both Zincblende InP and its metastable Wurtzite phase. Filled symbols indicate values taken relative to the final most accurate value, whereas empty symbols represent the absolute values. The absolute values of the total energies were shifted by −995.91 eV, to facilitate easier reading, and the Wurtzite lattice constant is the FCC-equivalent distance, rather than an actual lattice constant, in order to allow for direct comparisons.

Having converged the final accuracy parameter for the systems of interest in this chapter, the next step is testing other combinations of exchange-correlation functionals and basis sets. For the convergence studies carried out thus far, a specific combination was chosen based on *a posteriori* knowledge, but to ensure that this choice was a good one for the systems of interest, other combinations will now also be tested. These combinations include an exchange-correlation functional from every step of the Jacobs ladder in Figure 2.2 up to the hybrid functionals. For the LDA functional, the usual PZ version is used, as given in Eq. (2.20), for the GGA functional, the PBES version is used again, for the MGGA functional, the SCAN functional is used, for the hybrid functional, the HSE06 functional is used, which was shortly described in Section 2.2.3. All these functionals are both tested with the "High" and "Medium" version of the PseudoDojo basis sets and pseudopotentials.

The comparison of both Zincblende and FCC-equivalent Wurtzite lattice constants for all these combinations against the experimental values, are shown in Figure 3.8, together

with their respective calculation times in identical conditions. For the band gap comparisons in Figure 3.8, the LDA and GGA methods have used the DFT-$^1/_2$ correction scheme on the structures relaxed with regular LDA and GGA, since the DFT-$^1/_2$ correction yields wrong energies and thereby forces and stresses. Similarly, in the MGGA case the TB09 functional was used to calculate the band gaps with the SCAN relaxed structures, since TB09 cannot be used for structural relaxations. In Figure 3.8, the patterned bars represent the Wurtzite structures, whereas the solid bars represent the Zincblende structures. Considering the lattice constants, it appears that all the functionals are equally good at replicating the experimental value of Zincblende InP with the "High" PseudoDojo basis set, and that none of them correctly predicts the difference between Zincblende and Wurtzite. This is not necessarily an issue with the DFT calculations, but could also be due to the data on Wurtzite InP only coming from nanowires due to its meta-stability, and thus not being representative for a true bulk system as the one simulated with DFT here. The shift to the smaller "Medium" basis set worsens the lattice constants for all the functionals, but even more so for the GGA and MGGA functionals.

Considering the band gaps, MGGA drastically overshoots the band gaps for both basis sets whereas LDA-$^1/_2$ and GGA-$^1/_2$ underestimates it slightly. HSE gets the band gap almost exactly correct and with the "High" basis set it also gets the Zincblende/Wurtzite band gap difference almost exactly correct. Unfortunately using HSE is prohibitively expensive for the large basis set, which can be seen in the time per iteration test. As such, the optimal combination of exchange-correlation functional and basis set, providing the highest level of accuracy without being too computationally demanding, turns out to be the GGA with the "High" basis set, as was previously merely assumed. While slightly underestimating the absolute value of the band gaps, this choice yields the correct Zincblende/Wurtzite band gap difference, and as such can be comfortably utilised when only the differences are of importance. While the LDA functional performs equally well, and even slightly better for the band gaps, than GGA, the GGA functional should in principle provide a more accurate electron density than LDA, which could be important for some of the later calculations of the present chapter. For this reason, and since LDA and GGA are equally computationally demanding here, GGA is favoured for the calculations of this thesis.

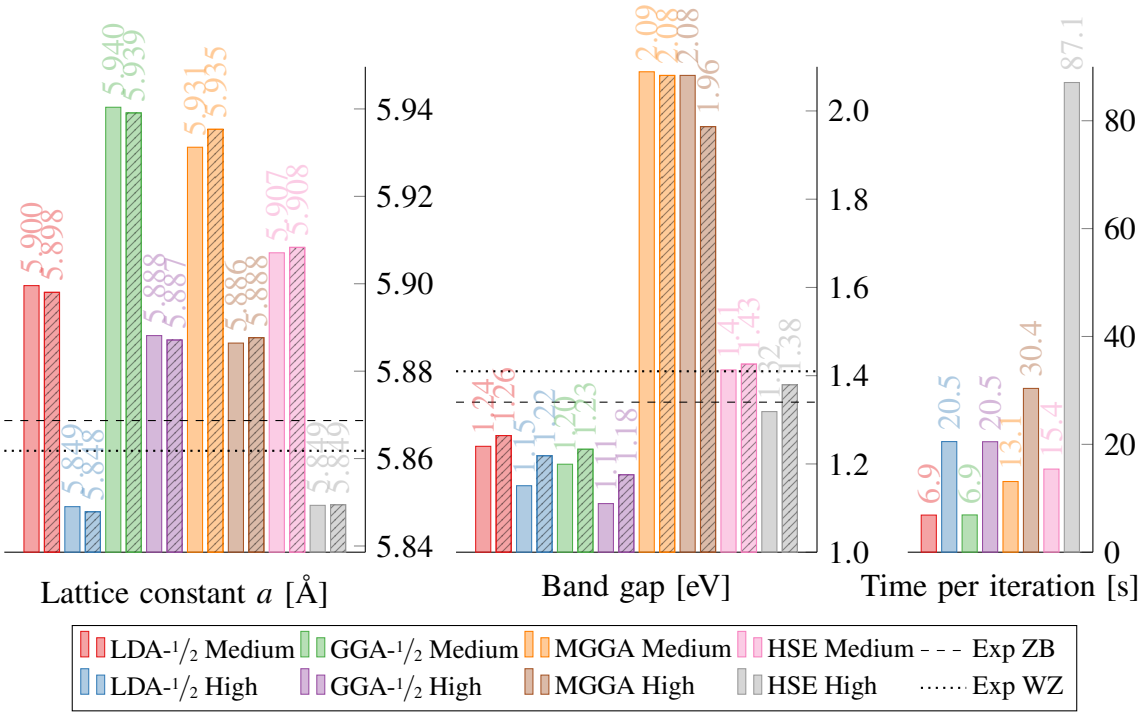
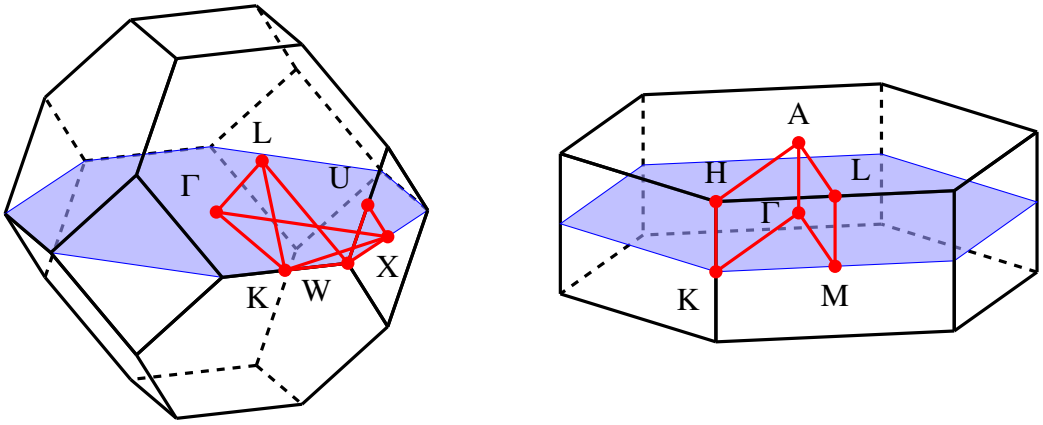


Figure 3.8: Comparison of combinations of exchange-correlation functionals and basis sets to simulate Zincblende InP and its meta-stable Wurtzite phase. Wurtzite results are shown by the patterned bars whereas the bare bars represent Zincblende results. For the DFT-$^1/_2$ functionals, the relaxations were performed with regular LDA and GGA. For the MGGA functional, relaxations were performed with SCAN while band gaps were calculated with TB09. Experimental values from: [94, 101–103].

3.2.3 Band structures and transmission spectra

Having performed a full convergence study on both crystallographic phases of InP, DFT calculations of the difference in the properties between the two phases can now be carried out. The first of such calculations to be carried out, are the band structures of the two phases. Due to the difference in the two phases Bravais lattice types, FCC for Zincblende and Hexagonal for Wurtzite, their band structures are plotted differently due to the different symmetries of the two phases. In Figure 3.9 are the first Brillouin zones of Zincblende and Wurtzite illustrated, with all their named high-symmetry points, i.e. critical points, shown, as well as the path between them which is usually traversed in band structures. Because of the symmetries of the crystal lattices, the crystals are fully described in a small subset of the first Brillouin zone defined by the allowed symmetry operations, aptly named the irreducible Brillouin zone. It is this irreducible Brillouin zone which is explored by traversing the path between the critical points shown in Figure 3.9. Despite that the critical point names "L" and "K" are shared between the two lattice types, these points are not equivalent in the two lattices and neither are the paths connecting them. Yet there are equivalent paths, for example the L-point in Zincblende lies along the [1 1 1]-direction, which was shown to be similar, in Section 3.2.1, to the [0 0 0 1]-direction of Wurtzite, on which its A-point lies, thus the Γ-L path of Zincblende is equivalent to the Γ-A path of Wurtzite.



Zincblende Wurtzite

Figure 3.9: First Brillouin zones of Zincblende and Wurtzite. Critical points are shown in red and labelled according to nomenclature, and the normal path for Brillouin zone exploration between these critical points, used in band structures, are also shown. The blue shaded area is the $z = 0$ plane, drawn for reference.

Since electron densities for both Zincblende and Wurtzite InP were already calculated with the accurate HSE functional and the "High" basis set in the convergance study, these were used to calculate the InP band structures which are shown in Figure 3.10. Since spin-orbit effects were not included in these calculations, the split-off valence band has not split off from the heavy hole valence band in the band structures. It can be seen that the Wurtzite band structure has an additional conduction band called the "dark" conduction band. This band originates from the 4-atom unit cell of Wurtzite, which can be considered an unfolding of the 2-atom unit cell of the Zincblende structure along the [1 1 1]-direction. This unfolding in real-space leads to a folding in reciprocal space, where the L-point in Zincblende is folded into the Γ-point in Wurtzite, yielding this new conduction band. This "dark" conduction band has Γ_8^c symmetry in WZ double-group notation with spin-orbit coupling [39], in contrast to the regular "bright" conduction band which has Γ_7^c symmetry. Similarly the split-off valence band and light hole valence band have Γ_7^v symmetry, whereas the heavy hole valence band has Γ_9^v symmetry. The new conduction band is called "dark" because optical transitions associated with this band is only allowed through spin-orbit coupling, and then only with the heavy hole valence band. Since the transitions are mediated by spin-orbit coupling, the transition oscillator strength is weak and the expected photoluminescence is therefore low. In certain Wurtzite materials, this dark conduction band is lower in energy than the regular bright conduction band, leading to a so-called pseudo-direct band gap. In other materials, such as Gallium Arsenide, the two conduction bands are so close energetically, that which one is lower can be controlled by strain [104].

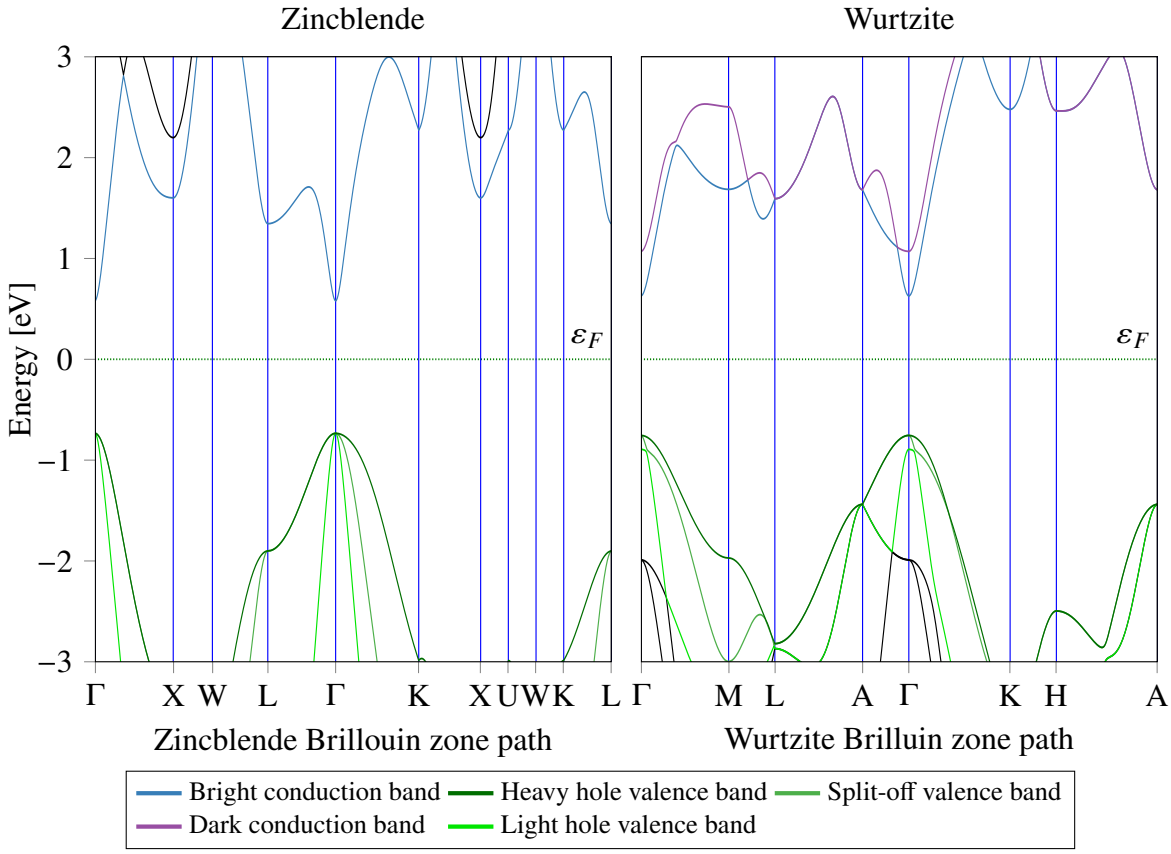


Figure 3.10: Zincblende and Wurtzite InP band structures calculated with HSE and the "High" Pseudodojo basis set. Bands are coloured based on their type to ease identification. The band colourisation was automatic and sometimes fail at band crossings.

Since the important energy differences, such as the dark conduction band offset, are difficult to quantify from Figure 3.10, they are collected for convenience in Table 3.1 and compared to pure GGA results, i.e. without the DFT-$^1/_2$ correction. The energy difference between the dark and bright conduction bands were found to be 0.495 eV and 0.441 eV, for the GGA and HSE functionals, respectively. Considering the results from Gallium Arsenide [104], transitioning from the direct to pseudo-direct band gap in Wurtzite InP, would therefore require a large strain of at least 2%-3%.

Returning to the differences between the GGA and HSE functional in Table 3.1, as was mentioned in the convergence study, the GGA functional correctly predicts the band gap difference between Zincblende and Wurtzite, despite underestimating the absolute band gap values. This is even the case without the DFT-$^1/_2$ correction, since regular GGA generally gets most of the band structure right and the DFT-$^1/_2$-correction usually just shifts the conduction bands upwards in energy. As can be seen from the values in Table 3.1, the GGA functional not only correctly determines the band gap difference between Wurtzite and Zincblende, but more or less also its distribution between conduction and valence band edge differences. It even fairly well replicates the energy difference between the dark and bright conduction bands. Since the DFT-$^1/_2$ correction is semi-empirical, as the charge fraction which is removed is based on fits to experimental data, this is a positive result, since it means that the correction can be avoided and calculations be made more "*ab initio*", when specifically studying differences

between Zincblende and Wurtzite, such as is the case in this thesis. Avoiding the DFT-$^1/_2$ correction has the added benefit, of allowing analysis and geometrical relaxations to be done with the same functional, making the studies more scientifically robust.

Table 3.1: Band structure parameters from band structures of Wurtzite and Zincblende InP calculated with either the PBES GGA functional or the HSE06 hybrid functional. The HSE06 band structures are shown in Figure 3.10. All energies are given in eV.

XC Functional	GGA			HSE		
Crystal Phase	ZB	WZ	Diff.	ZB	WZ	Diff.
Conduction band edge	0.242	0.283	0.041	0.655	0.690	0.035
Valence band edge	-0.288	-0.318	0.030	-0.655	-0.690	0.035
Band gap	0.531	0.601	0.070	1.310	1.379	0.069
Dark conduction band offset	-	0.495	-	-	0.441	-

Staying within band structures, the next bulk crystal property which will be examined in the two crystal phases of InP, is the phonon band structures, i.e. the phonon dispersion. Like how the electron density is the fundamental property of DFT, the dynamical matrix is the fundamental property of lattice dynamics, it gives the force response of an atom from the displacement of another atom from its equilibrium. Given the dynamical matrix of a system, the frequencies of the phonon modes of the system are calculated as the eigenvalues of the dynamical matrix. Due to the usage of the Born-Oppenheimer approximation, lattice ions are stationary in DFT calculations, and thus the dynamical matrix has to be calculated from finite differences in a repeated supercell, which is also termed the frozen phonon method. For dynamical matrix calculations, the number of unit cell repetitions to include in the finite differences, is an important convergence parameter which has a large impact on the computational requirements of the calculation. The necessary amount of repetitions can be drastically reduced by employing the Wigner-Seitz method, which basically consists of decomposing the contributions from atoms based on a Wigner-Seitz cell centred at the displaced atom. For the calculations in this thesis, convergence with repetitions was carefully checked both with and without the Wigner-Seitz method, as well as with regular accuracy parameters. The resulting phonon band structures for both Zincblende and Wurtzite InP, calculated with PBES and 5x5x5 repetitions (5x5x3 for Wurtzite) in the Wigner-Seitz method, is shown in Figure 3.11.

In the Zincblende phase, the Γ-point optical phonon frequencies are calculated to be at 37.1 meV for the Transverse Optical (TO) mode and 41.7 meV for the Longitudinal Optical (LO) mode. This matches fairly well the experimental values of 38.1 meV and 42.6 meV from the Ioffe database [102], thereby validating the utilised calculation method and parameters. Like for the optical band structures, it can be seen that the folding of the Wurtzite unit cell again folds a new set of bands into the first Brillouin zone, shown in Figure 3.11 by the dashed lines. These phonon modes consists of the Wurtzite unit cell atoms of the same species oscillating against each other, rather than together, as in the regular phonon modes.

The folding of the acoustic modes leads to new optical modes with frequencies far from the frequencies of the regular optical modes. Unfortunately, these new optical modes are not Raman active and thus require more advanced experimental techniques to measure, such as multiphonon Raman scattering or Inelastic Neutron Scattering (INS).

In the Wurtzite band structure, a small discontinuity of 0.525 meV in the dispersion of the LO phonon mode is seen near the Γ-point. This discontinuity is caused by the generation of a macroscopic electric field, coming from the change in covalency of displaced ions in polar materials, such as InP. This effect, called polar phonon splitting, is exclusive to non-cubic polar materials and also causes a discontinuity in the TO phonon dispersion (shown in Figure 3.11, if zoomed in enough). Polar phonon splitting is included in the dynamical matrix through a separate so-called non-analytic contribution, which will be described and studied further in Section 3.3.2. The apparent unphysicallity of these discontinuities can be explained by phonon-polariton coupling, as the macroscopic electric field from the ionic displacement couples to light in the long wavelength limit [105].

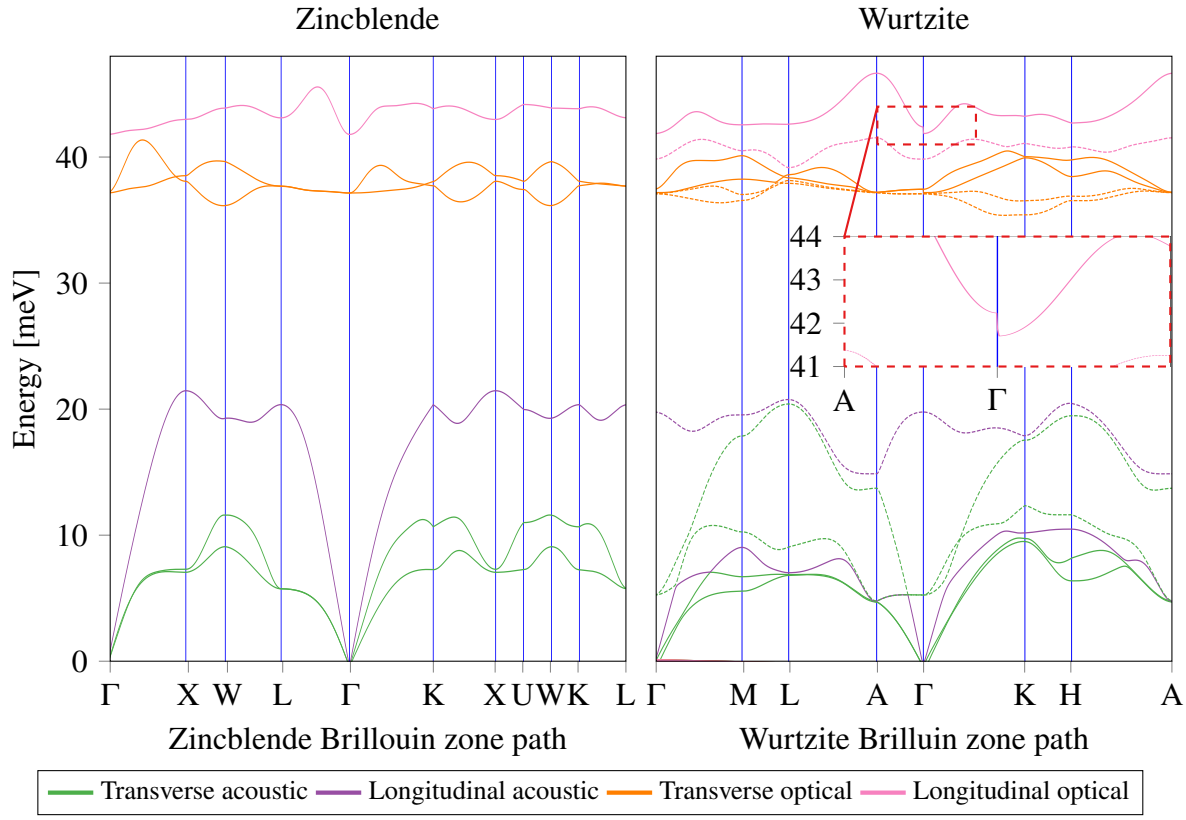


Figure 3.11: Zincblende and Wurtzite InP phonon band structures calculated in the frozen phonon scheme with the Wigner-Seitz method. Super cell repetitions were 5x5x5 and 5x5x3 for Zincblende and Wurtzite, respectively. Bands are coloured based on their type to ease identification. The band colourisation was automatic and fail at band crossings, as seen in the longitudinal acoustic mode of Wurtzite. The inset in the Wurtzite band structure shows the discontinuity in the dispersion coming from the non-analytic contribution.

The final property to investigate in the pristine Zinblende and Wurtzite bulk crystals, is the electrical transport in the phases. The transport direction is chosen along the similar [1 1 1]- and [0 0 0 1]-directions. To calculate the electrical transport, the NEGF method described

in Section 2.3 is utilised, with the central region identical to the electrodes and without an applied bias voltage. NEGF-DFT calculations require a much denser k-point sampling than regular DFT, especially in the transport direction, and as such an extremely high k-point density of 27 Å was utilised in these calculations, with even higher central region densities of 100 Å in the transport directions. The calculation of the transmission spectra with Eq. (2.39), requires an even denser transverse k-point sampling than the self-consistent NEGF calculation, it was therefore increased to 48 Å for these transport calculations. These k-point samplings were all chosen based on similar convergence studies as the ones shown in Figure 3.4 and 3.6. The numerical comparison of the electrical transport and conductivity of the two crystallographic phases will be given in Section 3.5.1 along with other transport calculations. In the present section, the qualitative difference in the two phases transport will be investigated, by comparing their reciprocal space transmission coefficients at a suitable energy. These transmission coefficients are shown in Figure 3.12 at −0.66 eV, due to the differences being large enough at this energy to be clearly visible. The real-space lattices are overlaid for better identification of directions. It is clear that the transmission in Wurtzite simply follows the hexagonal symmetry of the Wurtzite lattice. In the Zincblende transport along [1 1 1], there is the same hexagonal symmetry as in Wurtzite, yet there is also an additional weaker hexagonal symmetry with a phase-angle difference of 30°. Considering the real-space Zincblende lattice, this corresponds to the directions of the second nearest neighbours in-plane, which thereby explains the second hexagonal symmetry of the transmission coefficients. Evidently, Zincblende InP more strongly favours off-Γ transport along the [1 1 1]/[0 0 0 1]-direction, as compared to Wurtzite InP.

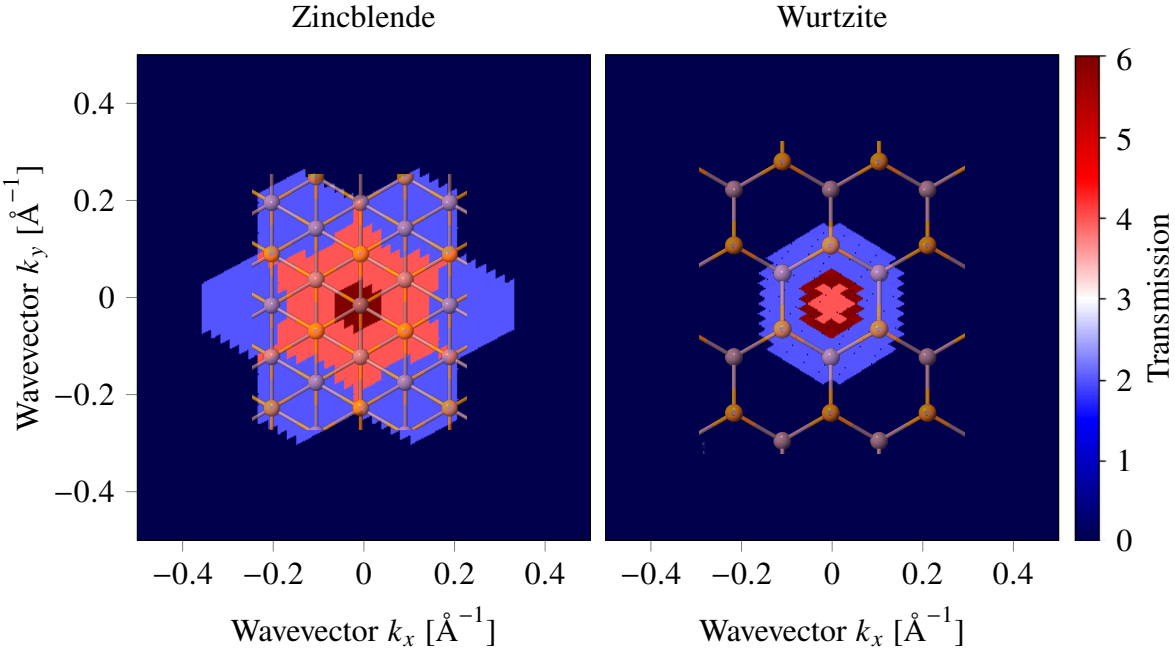


Figure 3.12: InP [1 1 1]/[0 0 0 1]-transport transmission coefficients at −0.66 eV in reciprocal space, with the real space lattice overlaid for comparison of directions. The NEGF-DFT transport calculation was unbiased, and the central regions were kept identical to the electrodes.

3.3 Indium phosphide polytypism - phonon dispersions

Having done a convergence study on the pristine polytypes of InP and investigated the differences and similarities in their bulk crystal properties, the next step of this chapter is the investigation of phase mixing, i.e. the mixing of these two polytypes and in particular RTPs. This study is split into three sections, the first and present section will present the systems to be studied and investigate their phonon dispersions, the following section deals with optical properties of the systems, including the dielectric function and phase-boundary quantum wells and the final section studies the conductivity and mobility of the defect systems.

3.3.1 Rotational twin planes and phase mixing

Examples of experimental polytypic defects were shown in Figure 3.1, seen with HR-STEM. The goal of this polytypism study is to emulate such polytypic defect systems. Unfortunately, DFT relies on Bloch's theorem to simulate bulk structures, and as such, the simulated systems will necessarily have to be fully periodic, rather than random, such as the shown experimental examples. In Chapter 4, a method to properly emulate randomness within Bloch's theorem, using so-called special quasi-random structures (SQSs), will be utilised. Using this method with extended planar defects, such as the ones to be studied here, would require unfeasible large super cells, and was therefore not done.

A Rotational Twin Plane (RTP) is the smallest possible polytypic defect, it consists of a sudden rotation of the Zincblende lattice by 60° around any of the ⟨1 1 1⟩-axes. The resulting structure, viewed from the ⟨1 1 0⟩-directions, such as in Figure 3.1, appears as a mirror image in the defect plane with its atomic species swapped, which is why the defect is called a twin. RTPs are often also called stacking faults, since they are an interruption of the regular ABC-stacking sequence of Zincblende along the ⟨1 1 1⟩-directions.

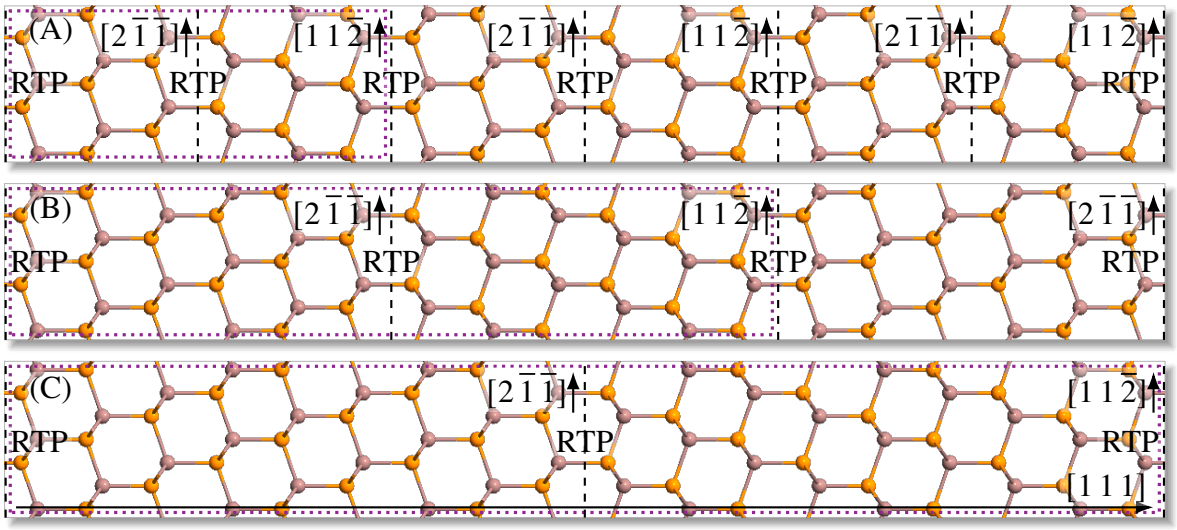


Figure 3.13: The Rotational Twin Plane (RTP) InP systems to be simulated. The purple dotted boxes indicate the utilised simulation unit cells, RTPs are indicated by dashed lines and the systems are extended to equal length only for easier comparison. (A) 1 RTP every 3 layers of Zincblende. (B) 1 RTP every 6 layers of Zincblende. (C) 1 RTP every 9 layers of Zincblende.

The two layers on either side of the defect plane, the last un-rotated layer and the first rotated layer, together forms a 2-layer Wurtzite segment. Since both layers can either be attributed to this Wurtzite segment, or to their own respective ZB lattices, this will not be considered a true Wurtzite segment in this thesis. Thus the highest density of RTPs possible in a system, is one RTP every three layers of ZB, in order for the ZB segments to be unquestionably ZB rather than WZ. This maximal RTP density system, together with of half and a third of its density will be studied in this dissertation and are illustrated in Figure 3.13.

Due to the above discussion about RTPs being two layer WZ segments, the minimal WZ segment size, in order to have a segment which is unquestionably Wurtzite, is three layers. Since a layer is shared between a ZB and a WZ segment at their interface, as was discussed in regards to the RTPs, this physical three-layer Wurtzite segment is actually two-layers, as it shares a layer equally with ZB to either side. To accommodate for Bloch's theorem, the phase mixing systems super cells have to be fully periodic, and as such the ZB segments have to be a multiple of three layers long, and the WZ segments a multiple of two layers long, in order to return to the original layer before the next periodic repetition of the lattice. For this reason, the maximally intermixed phase mixing system consists of two WZ layers every three layers of ZB. This system is shown in Figure 3.14, together with systems with double the WZ (four WZ layers every three layers of ZB) or double the ZB (two WZ layers every six layers of ZB) which are the systems which will be studied in this dissertation. Note that the simulation unit cell in Figure 3.14 is indicated by the purple dotted line, as in Figure 3.13.

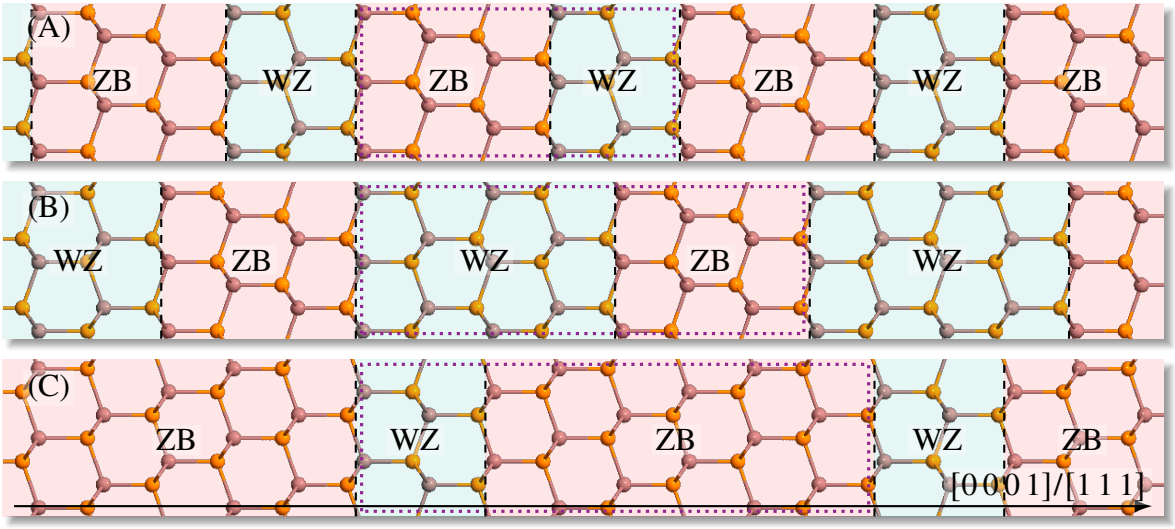


Figure 3.14: The phase mixing InP systems to be simulated. The purple dotted boxes indicate the utilised simulation unit cells and the systems are extended to equal length only for easier comparison. For illustration, the Zincblende and Wurtzite segments coloration terminates in the middle of layers, to indicate that these layers are shared between both segments. (A) 2 Wurtzite layers every 3 layers of Zincblende. (B) 4 Wurtzite layers every 3 layers of Zincblende. (C) 2 Wurtzite layers every 6 layers of Zincblende.

Besides the fully periodic systems shown in Figure 3.13 and 3.14, it is possible in the NEGF formalism to construct perfectly pristine ZB and WZ systems, with only a single or a few RTPs or phase segments in the central regions. This will be utilised in Section 3.5.1 to simulate the effects of single RTPs or WZ segments.

3.3.2 Phonons in polar materials

Due to the macroscopic electric field generated from optical phonons in polar materials such as InP, many material properties, including the dielectric function, have substantial contributions from phonon effects. For this reason, the first study of the polytypic defect systems in Figure 3.13 and 3.14, will be on the polar optical phonons of the systems. As was briefly mentioned in Section 3.2.3, the effect of polar optical phonon splitting is included in the dynamical matrix through the so-called non-analytic contribution [106], which is given as

$$\mathbf{D}^{\mathrm{NA}}_{i\alpha j\beta}(\mathbf{q}) = \frac{e}{\epsilon_0 V\sqrt{M_i M_j}} \frac{[\mathbf{Z}(i)\cdot\hat{\mathbf{q}}]_\alpha [\mathbf{Z}(j)\cdot\hat{\mathbf{q}}]_\beta}{\hat{\mathbf{q}}\epsilon^\infty\hat{\mathbf{q}}} \tag{3.1}$$

where V in Eq. (3.1) is the volume of the unit cell, α and β are Cartesian directions, M_i is the mass of atom i, ϵ_0 is the vacuum permittivity, e is the electron charge, ϵ^∞ is the high-frequency dielectric tensor, $\hat{\mathbf{q}} = \mathbf{q}/|\mathbf{q}|$ is a unit vector in reciprocal space and $\mathbf{Z}(i)$ is the Born effective charge tensor, which is simply the derivative of the systems polarisation with regards to the atoms coordinates. This contribution is called non-analytic because it approaches different values for $q \to 0$ from different directions in reciprocal space, i.e. it is responsible for the discontinuity shown in Figure 3.11.

To calculate the non-analytic contribution to the dynamical matrix in Eq. (3.1), the Born effective charges of the systems are required. During the calculation of the Born effective charges used to calculate the phonon band structures in Figure 3.11, it was found out, that in Wurtzite there is an anisotropy between the Born effective charges along the $[0\,0\,0\,1]$- and $[1\,0\,\bar{1}\,0]$-direction, which is not present in the Zincblende structure. To investigate how this anisotropy evolves in the defect systems, as the RTP or phasemixing density increases, this Born effective charge anisotropy is plotted in Figure 3.15 as a function of the atoms fractional z-coordinate, since the z-axis lies along the $[1\,1\,1]$/$[0\,0\,0\,1]$-direction.

In Figure 3.15, positive anisotropies refer to higher Born effective charges along the $[0\,0\,0\,1]$-direction as compared to the $[1\,0\,\bar{1}\,0]$-direction. The positive anisotropies belong to the Indium atoms, which also have positive Born effective charges, and the negative anisotropies belong to the Phosphorus atoms, which also have negative Born effective charges, i.e. large absolute anisotropies refer to larger Born effective charges along the $[0\,0\,0\,1]$-direction.

Looking at the RTP systems, it is clear that a large anistropy is induced just around the RTP, which is then compensated for by the next few atoms having a lower than regular anisotropy. Interestingly, the anisotropy converges to a constant value of about $0.05\,e$ in the Zincblende regions, where the anisotropy should be zero. Thus the symmetry-breaking of even a single RTP must perturb the entire ZB segment, yielding a net anisotropy even in bulk Zincblende. This behaviour is not seen in the system with 1 RTP every 3 layers of Zinblende, which is likely because the RTPs are so close together that they interact with each other, forcing an even stronger compensation from the layers which does not belong to any RTP.

Looking at the phase mixing systems, the same convergence behaviour is seen in the Zincblende segments, with the anisotropy reaching a constant value of $0.05\,e$ even in the

smallest Zincblende segment. It is also seen that the Wurtzite segments reach the bulk Wurtzite anisotropy value even in the smallest 2-layer segment, although only 1 of the layers reach this value. Comparing the bulk WZ (or WZ segment) anisotropy value with the RTP-induced anisotropy, it is clear that phase mixing induces larger Born charge anisotropies than RTPs.

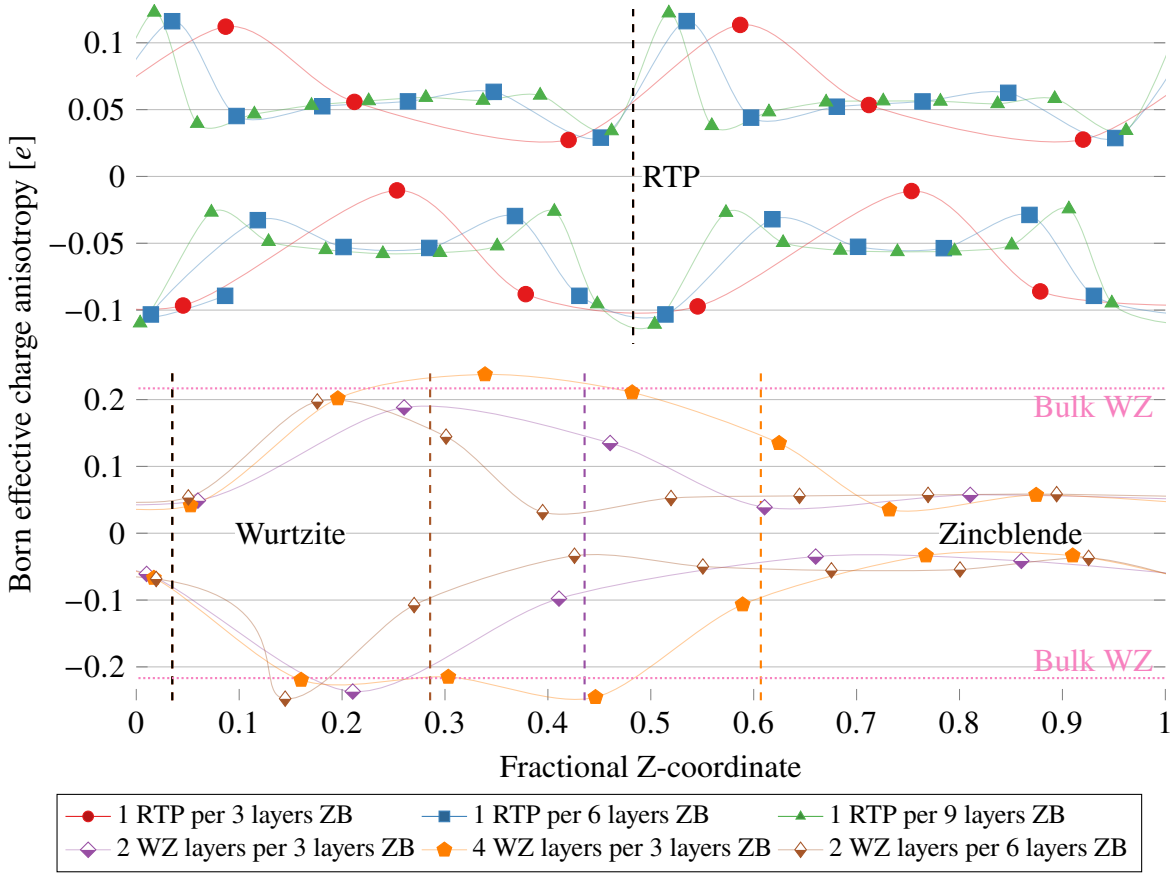


Figure 3.15: Born effective charge anisotropies of RTP and phase mixing systems plotted along the fractional z-coordinate, which is the axis of the stacking sequence. Lines between points are only added for visual guidance and reflect no physical effects. Large absolute anisotropies indicate larger Born effective charges along the $[0\,0\,0\,1]$-direction as compared to the $[1\,0\,\bar{1}\,0]$-direction. For the RTP systems, the systems are shifted so that their RTPs are matched to coincide, and their location is marked by the dashed line. For the phase mixing systems, the systems are shifted so that one interface is matched across the systems, marked by the dashed black line. The other interfaces are marked by dashed lines in their systems respective colours. The anisotropy of bulk Wurtzite is also plotted by the dotted pink line for reference.

Having calculated the Born effective charges of the polytypic systems, the non-analytic contribution in Eq. (3.1) can be calculated and added to the systems dynamical matrices. For the calculation of the regular dynamical matrix of the polytypic systems, the Wigner-Seitz method was used with 5x5x3 repetition super cells for all the systems. The phonon Density Of States (DOS) calculated from the phonon dispersions of the systems, are illustrated in Figure 3.16. It can be seen that all the polytypic systems DOSs look almost identical, yet with some clear differences from both pristine Zincblende and Wurtzite. Considering that the polytypic systems, and the phase mixing systems in particular, are systems mixed of the

two phases, it would be expected that their DOSs would be somewhere in between those of Wurtzite and Zincblende. Yet at various energies, and in particular at 43 meV, the DOSs of the polytypic systems are all lower or higher than those of Zincblende and Wurtzite. It can be seen though, that the DOSs of the polytypic systems more closely resembles Wurtzite than Zincblende, considering that they miss the shoulder from 12 meV to 15 meV, have an additional side peak at 11 meV, have their peak at 20 meV moved to 18 meV and lack the high (0.9 meV^{-1}) narrow peak at 37.4 meV.

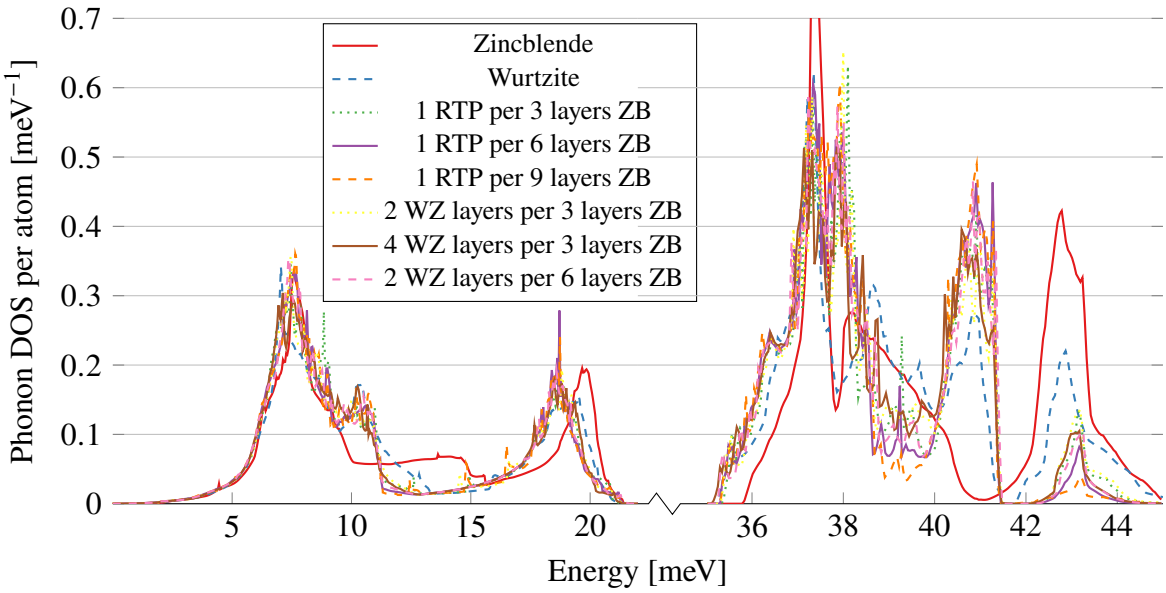


Figure 3.16: Phonon DOS per atom of InP polytypic systems. The non-analytic contribution is included in the dynamical matrix, which was calculated with the Wigner-Seitz method using 5x5x3 repetition super cells.

The polar phonon splitting discontinuity in the optical phonon dispersions shown in the inset back in Figure 3.11, is an inherent effect of non-cubic lattices and was caused by the non-analytic contribution to the dynamical matrix. To investigate the effect of the non-analytic contribution to the phonon dispersion of the polytypic systems, the phonon energies of the regular LO and TO phonon modes, as $\mathbf{q} \to \Gamma$ from both the A- and K-directions, are given in Table 3.2. It can be seen, that the TO-mode discontinuity is far smaller than the LO-mode discontinuity in all the systems, except for Wurtzite where it is still substantial. Unsurprisingly, the Wurtzite system has the largest discontinuities out of all the systems, followed by the system with 2 layers of Wurtzite every 3 layers of Zincblende and then the other medium sized systems. There is clearly a relationship between the size of the discontinuities and the amount of Wurtzite in the system, as is evident from the much smaller discontinues in the system with 1 RTP every 9 layers of Zincblende and the phase mixing system with 6 layers of Zincblende. The relationship is not simple or linear, as the system with 4 layers of Wurtzite every 3 layers of Zincblende should then be the polytypic system with the largest discontinuities.

Table 3.2: Polar phonon splitting discontinuities at the Γ-point of LO and TO modes in polytypic InP systems. Phonon mode energies were evaluated as **q** approaches zero from either the A- or K-directions. The second regular TO mode showed no discontinuity and was therefore not included in this table. The phonon dispersions were calculated from a dynamical matrix including the non-analytic contribution, and was calculated with the Wigner-Seitz method using 5x5x3 repetition super cells. All energies are given in meV.

System	LO mode			TO mode		
	Γ (K)	Γ (A)	Diff	Γ (K)	Γ (A)	Diff
Pristine Zincblende	41.66	41.66	0	37.07	37.07	0
Pristine Wurtzite	41.72	42.23	0.53	37.10	37.36	0.26
1 RTP per 3 layers ZB	41.67	41.99	0.32	37.23	37.16	-0.07
1 RTP per 6 layers ZB	41.61	41.91	0.30	37.15	37.25	0.10
1 RTP per 9 layers ZB	41.62	41.79	0.16	37.16	37.12	-0.04
2 WZ layers per 3 layers ZB	41.67	42.03	0.37	37.24	37.15	-0.09
4 WZ layers per 3 layers ZB	41.65	41.98	0.33	37.11	37.17	0.06
2 WZ layers per 6 layers ZB	41.65	41.92	0.27	37.24	37.17	-0.07

Considering that the mobility of polar materials, such as InP, is limited by polar optical phonon scattering, the discontinuities shown in Table 3.2 should yield anisotropic mobilities. This topic will be investigated in Section 3.5.2 with classical mobility models.

3.4 Indium phosphide polytypism - optical properties

In the semiconductor device guiding the research in the present thesis, the PIN photodiode shown in Figure 1.2, the most critical material property for its operation is the photon absorption of its constituents. For this reason, the dielectric function of the polytypic systems is calculated in Section 3.4.1, which is then used to derive the absorption coefficients of the systems. In Section 3.4.2 the polytypic defects are studied as atomically thin quantum wells, since quantum wells can be utilised to enhance photon absorption at specific wavelengths.

3.4.1 The complex dielectric function

The key material parameter in solid state optics is the complex dielectric function $\epsilon_r(\omega)$, which is related to the more physical electric susceptibility $\chi(\omega)$ through the simple relation $\epsilon_r(\omega) = \chi_e(\omega) + 1$. The electrical susceptibility quantifies a materials polarisability in response to an electric field, such as the one in an electromagnetic wave at a frequency ω. It can be calculated with the Kubo-Greenwood linear-response formula [78]

$$\chi_{ij}(\omega) = \frac{e^2}{\hbar m_e^2 V} \sum_{nm\mathbf{k}} \frac{f_{m\mathbf{k}} - f_{n\mathbf{k}}}{\omega_{nm}^2(\mathbf{k})[\omega_{nm}(\mathbf{k}) - \omega - i\Gamma/\hbar]} p_{nm}^i(\mathbf{k}) p_{nm}^j(\mathbf{k}) \tag{3.2}$$

where m_e is the electron mass, Γ is an energy broadening, $p^i_{nm} = \langle n\mathbf{k}|\mathbf{p}^i|m\mathbf{k}\rangle$ is the i'th electron component of the momentum operator between states n and m, $\hbar\omega_{nm} = E_n - E_m$ is the energy difference of the states and $f_{n\mathbf{k}}$ is the Fermi function at energy $E_n(\mathbf{k})$ and k-point $\mathbf{k}$. The Kubo-Greenwood formula in Eq. (3.2) only gives the electrical contribution to the dielectric function. At low frequencies, where the ions have time to respond to the electrical field, there will also be an ionic contribution from the phonons, in a polar material such as InP. The ionic contribution is simply given by a sum of Lorentz oscillators over each phonon mode [107]

$$\epsilon^{\text{ionic}}_{r,\alpha\beta}(\omega) = \frac{1}{\epsilon_0 V}\sum_{\lambda}\frac{S^{\lambda}_{\alpha\beta}}{\omega^2_{\lambda} - \omega^2 - i\Gamma\omega} \tag{3.3}$$

where ω_λ is the frequency of phonon mode λ and $S^{\lambda}_{\alpha\beta} = \xi^{\dagger\lambda}_{\alpha}\xi^{\lambda}_{\beta}$ is the Lorentz oscillator strength, with ξ^{λ}_{α} the infrared mode polarity, given as

$$\xi^{\lambda}_{\alpha} = \sum_{i,\beta} Z_{\alpha\beta}(i)e^{\lambda}_{\beta}/\sqrt{M_i} \tag{3.4}$$

where e^{λ}_{β} is the eigendisplacement of the phonon mode and $Z_{\alpha\beta}(i)$ is the $\alpha\beta$'th Cartesian component of the Born effective charge tensor for atom i.

Using the already calculated Born effective charges and dynamical matrices, the Kubo-Greenwood formula in Eq. (3.2), together with the ionic contribution in Eq. (3.3), is used to calculate the complex dielectic function in all the polytypic systems, with a much denser Brillouin-zone k-point sampling of 28 Å. Just like the Born effective charges, there is an anisotropy in the dielectric function between the A-direction and the K-direction, coming from the Wurtzite parts of the polytypic systems. The A-direction component of the dielectric tensor of all the systems are plotted in Figure 3.17 and the K-direction component in Figure 3.18. It can be seen, in either direction, that the polytypic systems high-energy behaviour most closely resembles Zincblende, with a degree of resemblance depending on their Wurtzite content. At these high energies Wurtzite and Zincblende behaves quite differently, with the dielectric function of Wurtzite dropping earlier in energy than that of Zincblende. Especially at 4 eV there is a large contrast between the phases, as Zincblende has an additional top in its dielectric function, whereas the one of Wurtzite drops to zero. This stark contrast is attributed to electronic resonances between deep valence bands and high-energy conduction bands in both phases, as it only begins to occur for energies over 1.0 eV over the band gap of InP.

It is important to note here, that the dielectric function in Figure 3.17 was calculated with regular PBES GGA, so that the achieved band gap is only around 0.6 eV and all energy values should be shifted accordingly. The sharp resonances at very low energies come from the ionic contribution Lorentz oscillators, whose contributions to the dielectric function ends at around 0.1 eV. Since the dielectric function of Zincblende is isotropic, the largest differences between the K- and A-direction components are seen in the Wurtzite system. Most notable is the lack of a resonance at approximately 3 eV in the K-direction, which causes the Wurtzite dielectric function to flatten out until the resonances starting to occur at 3.5 eV.

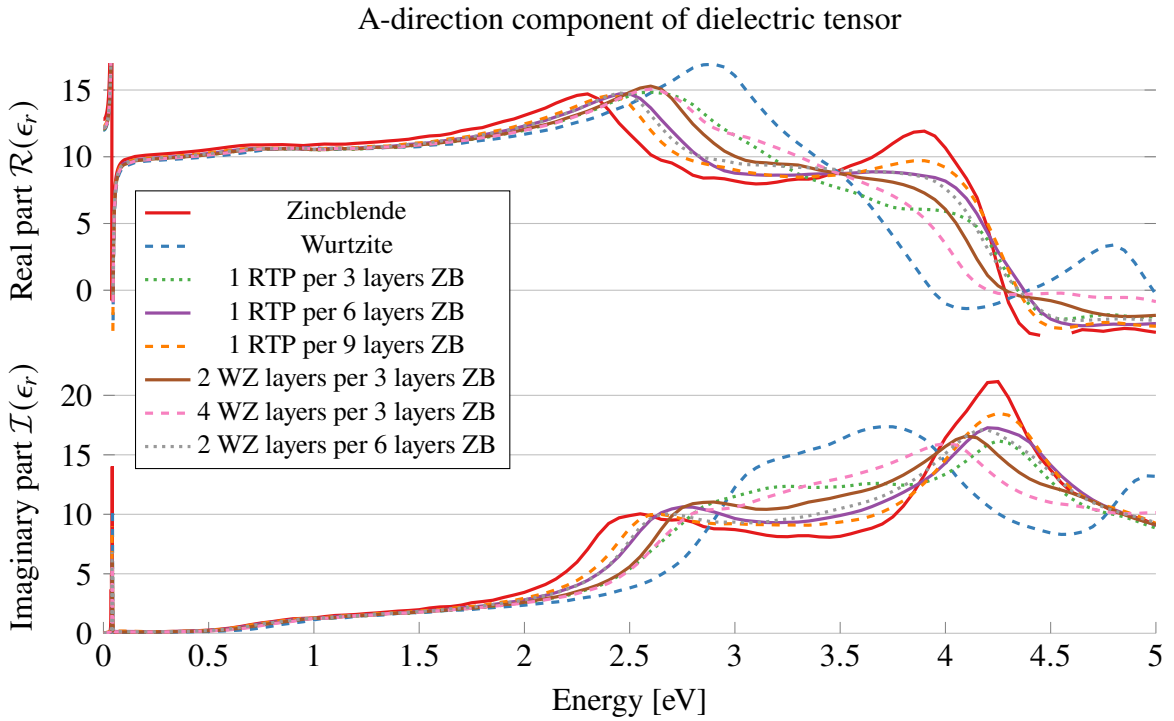


Figure 3.17: Real and imaginary part of the A-direction component of the dielectric tensor in polytypic InP systems. The electronic dielectric function is calculated with the Kubo-Greenwood formula, Eq. (3.2), whereas the ionic dielectric function is calculated as a sum of Lorentz oscillators, Eq. (3.3). The full dielectric tensor is the sum of both contributions and it is sampled at a Brillouin-zone sampling density of 28 Å.

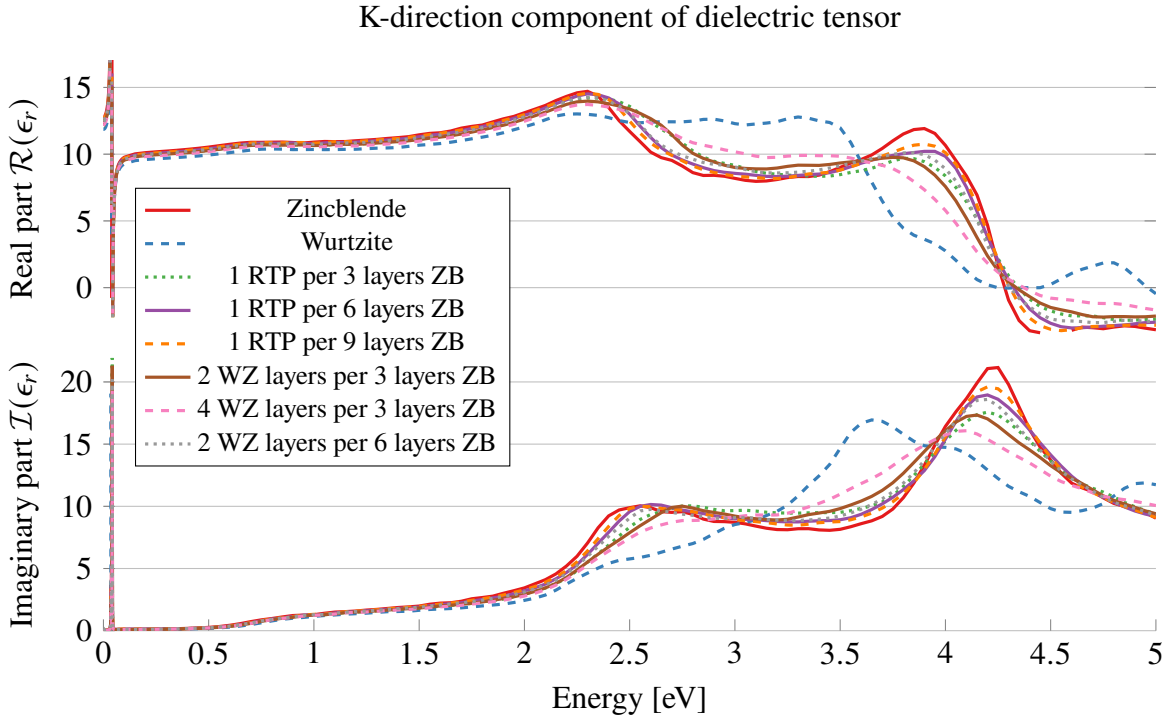


Figure 3.18: Real and imaginary part of the K-direction component of the dielectric tensor in polytypic InP systems. The electronic dielectric function is calculated with the Kubo-Greenwood formula, Eq. (3.2), whereas the ionic dielectric function is calculated as a sum of Lorentz oscillators, Eq. (3.3). The full dielectric tensor is the sum of both contributions and it is sampled at a Brillouin-zone sampling density of 28 Å.

In both directions, the real part of the dielectric function is approximately constant in-between the ionic contributions and the high-energy resonances starting at around 1.5 eV. In this energy-range the frequency is too fast for the ions to respond, so that only the electrons contribute to the dielectric function. This constant value, called the dynamic or high-frequency dielectric constant, is an important material parameter together with the low-frequency or static dielectric constant, which is the full dielectric function evaluated at zero field, i.e. with the ionic contributions. These values are tabulated in Table 3.3 in both the K- and A-directions for all the studied systems. As can be seen, the calculated dielectric constants for Zinblende InP fairly closely matches the experimental values taken from Ioffe [102]. Wurtzite shows a noticeable difference in the static dielectric constant to Zincblende, with the dynamic dielectric constant surprisingly showing a smaller difference. This means that the difference in the static dielectric constants originate not only from the phonon dynamics, but also from the differences in the electronic states between the two phases. Another surprising revelation in Table 3.3, is that Wurtzite is among the systems showing the smallest anisotropies in the dielectric constants, with the larger anisotropies appearing in the defect systems most closely resembling pristine Zincblende. If it is assumed that Zincblende actually has the highest anisotropy, once its symmetry along the A-direction has been broken, the results in Table 3.3 can be explained.

Table 3.3: A- and K-direction components of the static and dynamic dielectric constants in InP polytypic systems. Static dielectric constants were found by taking the full dielectric functions in Figure 3.17 and 3.18 and evaluating them at zero energy. The dynamic dielectric constants was found in a similar fashion, except that the dynamic contributions to the dielectric functions were not included. Experimental InP values are taken from the Ioffe database [102].

System	ϵ_r^0			ϵ_0^∞		
	K	A	Diff	K	A	Diff
Experimental Zinblende [102]	12.5	12.5	0	9.61	9.61	0
Pristine Zincblende	12.80	12.80	0	10.28	10.28	0
Pristine Wurtzite	11.86	11.99	0.13	9.70	9.76	0.06
1 RTP per 3 layers ZB	12.42	12.06	-0.36	10.06	9.94	-0.12
1 RTP per 6 layers ZB	12.63	12.21	-0.42	10.20	10.11	-0.09
1 RTP per 9 layers ZB	12.69	12.29	-0.39	10.24	10.18	-0.06
2 WZ layers per 3 layers ZB	12.38	12.09	-0.29	10.04	9.96	-0.08
4 WZ layers per 3 layers ZB	12.26	12.16	-0.10	9.96	9.98	0.02
2 WZ layers per 6 layers ZB	12.54	12.19	-0.35	10.14	10.07	-0.07

Having calculated the dielectric function along both the K- and A-direction of the defect materials, the absorption coefficients can finally be calculated. The absorption coefficient is calculated from the complex refractive index, which is simply the square root of the dielectric function [107]

$$n + i\kappa = \sqrt{\epsilon_r} \tag{3.5}$$

where n in Eq. (3.5) is the real refractive index, which is another important optical material property and κ is the extinction coefficient, which is related to the optical absorption via

$$\alpha = 2\frac{\omega}{c}\kappa \tag{3.6}$$

where c is the speed of light. The absorption coefficients of all the polytypic systems are calculated using Eq. (3.5) and (3.6), and the results are plotted in Figure 3.19 for both the A- and K-directions. Just as with the dielectric functions, the polytypic systems resemble Zincblende, with a degree of resemblance depending on their Wurtzite content, as was to be expected since the absorption coefficients are calculated from the dielectric functions. The differences are harder to distinguish in Figure 3.19 than in the dielectric functions, due to the scaling of the absorption coefficients with the electromagnetic frequency demanding a log scale to properly view them.

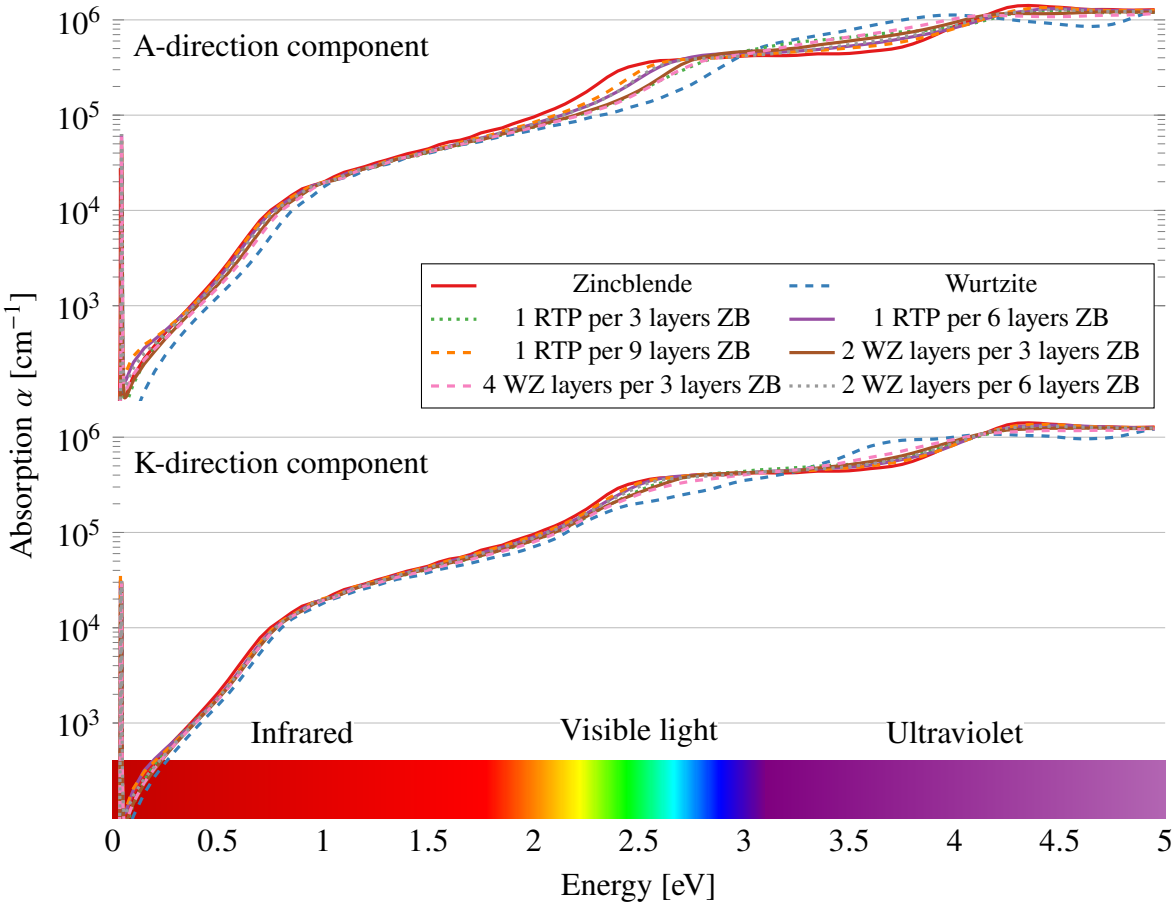


Figure 3.19: Photoabsorption of polytypic InP systems along both the A- and K-directions. The sharp peaks at low energies are due to low broadenings of the ionic contributions to the dielectric functions used to calculate the absorption coefficients. The electromagnetic spectrum is illustrated for easier comprehension of the energy to wavelength relation. DFT calculations were performed with PBES, which drastically underestimates the band gaps, and as such the absorption region up to approximately 0.53 eV should be stretched up to approximately 1.34 eV to compare with experimental absorptions.

3.4.2 Polytypism as atomically thin quantum wells

It is a common concern, that the randomly occurring polytypic defects may function as small 70 meV deep quantum wells, despite their sizes being at the few-atoms scale, or in the extreme case of a RTP, even a single atomic layer. As an example, there are theoretical calculations and experimental data, suggesting that RTPs in Gallium Nitride do indeed work as atomically thin, shallow, quantum wells [95–99]. Polytypic defects acting as quantum wells can either be noise in photodetectors, such as the PIN photodiode in Figure 1.2, or potentially be utilised as optical sensors themselves at very specific wavelengths.

The transition between Zincblende and Wurtzite in polytypic defects modelled as quantum wells, is commonly modelled with an abrupt band gap shift at the interfaces between the two phases. Since the band gap is technically speaking a bulk crystal property, in reality there will be a charge transfer between the two phases, which in turn will cause a smooth transition between the band gaps. To investigate how gradual or sudden this transition is, and thereby the justifiability of treating RTPs or other small polytypic defects as quantum wells, a Zincblende/Wurtzite interface super cell is constructed. This interface system is constructed with a total of 127 layers of InP, 64 layers of Wurtzite and 63 layers of Zincblende, to ensure that the individual phases are bulk-like in their centres, far away from the interfaces. To quantify the "band gap" of each separate layer in the system, the Projected Density Of States (PDOS) is calculated, projected onto the individual atoms of the system. A cutoff of $0.1\ \mathrm{meV}^{-1}$ is used, to determine the Conduction Band Minimum (CBM), and Valence Band Maximum (VBM), from the calculated PDOS of the individual atoms. This specific cutoff was chosen to replicate the bulk band gap values in the centres of the two InP phases. The "band gap" per layer, i.e. CBM-VBM difference, is plotted in Figure 3.20, as a function of the atoms' z-coordinates, which lie along the [1 1 1]/[0 0 0 1]-direction. It is plotted for a few different cutoff values, to show the sensitivity of the results to its choice, with the system super cell overlaid on top.

Looking at Figure 3.20, it is evident that the change in the layer band gap from Zincblende to Wurtzite, is far from instantaneous. The transition between the band gaps of the two phases, going from Zincblende to Wurtzite, is seen to require approximately 7 nm or 20 atomic layers to stabilise. Based on this result, there will be no visible band offset or misalignment in the 1-layer thin RTPs, or in phase mixing segments of a few atomic layers, such as the polytypic systems studied thus far. It is also clear, that the abruptness of the transition of the local band gap depends critically on the chosen PDOS cutoff value, with higher (lower) cutoffs yielding sharper (smoother) transitions. Comparing the local band gaps in the middle of the Wurtzite and Zincblende phases, to the bulk crystal bands gaps plotted for reference, it is evident that the chosen cutoff value of $0.1\ \mathrm{meV}^{-1}$ is correct for this study.

Another fascinating result extracted from Figure 3.20, is that the layer band gap transition is anisotropic: The band energy difference going from Zincblende to Wurtzite is different in comparison to the transition going from Wurtzite to Zincblende. The Wurtzite to Zincblende transition only needs approximately 3.5 nm or 10 atomic layers. This is likely due to the polar nature of InP causing in-layer dipoles between Indium and Phosphorus atoms in each

layer, which means that the two sides of an InP layer are electronically different. A fact which is well-known in the field of semiconductor growth. Based on these transition widths, a well-defined polytypic quantum well structure, made from the two crystal-phases of InP, will therefore require a minimum width of 10.5 nm, in order to reach the bulk band gap value in the centre of the well. Similar results were also achieved in the work of L. Dacal and A. Cantarero [108], although with shorter and smoother transitions. This difference is attributed to their usage of MGGA and plane waves to describe local properties, in contrast to the local basis sets and regular GGA/HSE utilised in this thesis.

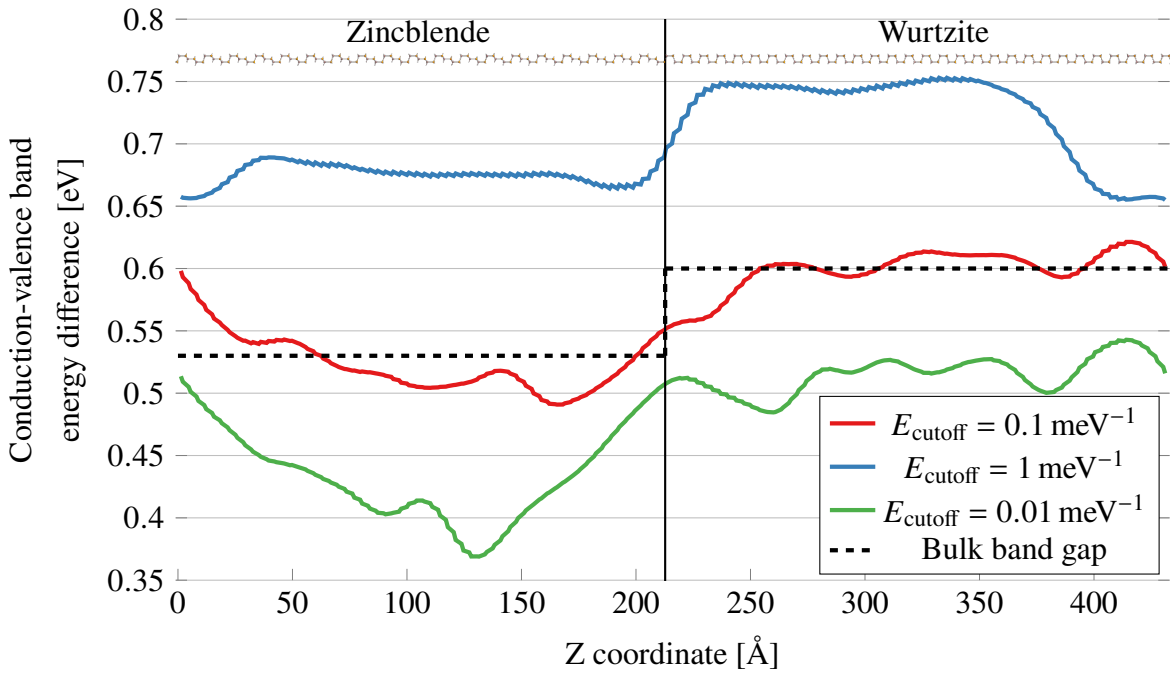


Figure 3.20: Conduction-valence band energy differences, i.e. "local band gaps", calculated from a PDOS, in a 127 atomic layer long Zincblende/Wurtzite interface system of InP, for three different PDOS cutoff values. The used exchange-correlation functional was GGA PBES and the bulk band gaps achieved with the same functional are also plotted for comparison. The super cell of the system is overlaid in the top of the figure for illustrative purposes.

To build confidence in the results gathered from Figure 3.20, the same calculation is carried out for a smaller system, consisting of only 86 layers of InP, with the same PDOS cutoff of 0.1 meV^{-1}. The calculation is performed both with GGA PBES again and with the hybrid HSE functional, which was shown back in Figure 3.8 to produce more accurate band gaps. The resulting local band gaps are plotted in Figure 3.21, together with their respective bulk crystal band gaps. Inspecting the transitions in Figure 3.21 for both the PBE and HSE functionals, confirms the results obtained above for the larger system, yet with slightly smaller transition widths (approximately 6 nm and 5 nm). These shorter local band gap transition widths can be explained by the amplified interaction with the systems periodic image, due to the reduced system length. Another possible explanation, at least for the HSE functional, could be the reduced tunnelling due to the higher band gaps.

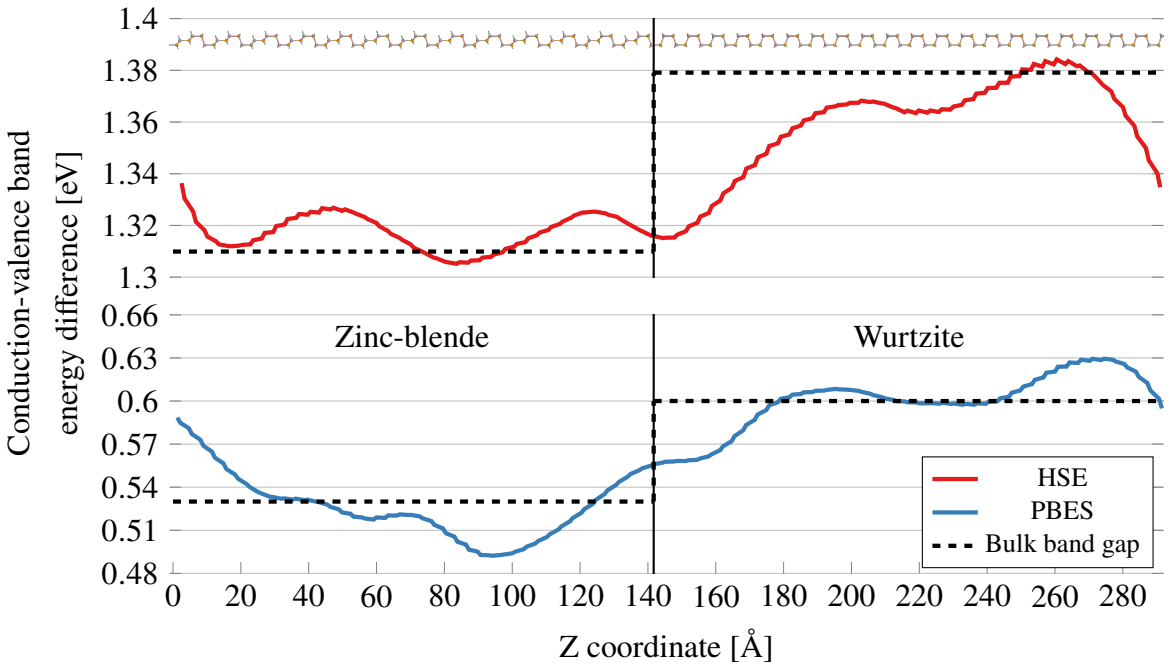


Figure 3.21: Conduction-valence band energy differences, i.e. "local band gaps", calculated from a PDOS, in a 86 atomic layer long Zincblende/Wurtzite interface system of InP, for three different PDOS cutoff values. The used exchange-correlation functional was GGA PBES and the bulk band gaps achieved with the same functional are also plotted for comparison. The super cell of the system is overlaid in the top of the figure for illustrative purposes.

As was already mentioned with the smaller systems in Figure 3.21, one weakness in this study is the effect of the periodic image on the local band gap transition of the central interface. This issue was the main reason that such a long system was used for the study in Figure 3.20, as the shorter system in Figure 3.21 still shows signs of this periodic interaction. To completely eliminate this conceivable error, the NEGF-DFT method described in Section 2.3 can be utilised. In this method, there is no periodic image in the transport direction due to the semi-infinite electrodes. For this reason, the Zincblende/Wurtzite interface system used in Figure 3.20, is converted to a NEGF system, and the Local Density Of State (LDOS) is calculated and shown in Figure 3.22.

Upon inspection, the advantage of the semi-infinite electrodes of the NEGF method becomes immediately apparent, as the local band gap transition has become less irregular and more smooth. In the NEGF method, both electrodes are true bulk crystal systems, which means they exhibit the correct band gap, which they can then in turn enforce to the left and right side of the interface in the central region. This method has its own inherent issues, since the transition to the electrode band gaps is strictly enforced by the calculation method, no matter the length of the central region. For this reason, the length of the central region needs to be carefully converged, just as the length of the systems in Figure 3.20 and 3.21 needed to be. Convergence with central region length was checked for the system in Figure 3.22. Inspecting the width of the local band gap transition, it can be seen that the transition width is notably longer, approximately 15 nm (from flat band to flat band) as compared to the 7 nm of the periodic DFT calculation. At the interface between the phases, there is also a slight

increase in the local band gap which did not appear in the periodic DFT calculations.

Another important point to notice, is that the local band gap transition is entirely determined by the valence bands as the conduction bands are entirely flat across the interface. This result stands in direct contrasts to the models commonly used to describe interfaces, wherein the CBM and VBM utilised for both the materials of the interface, are taken directly from bulk calculations. Doing so for InP would predict band-misalignment at both the CBM and VBM, according to the values presented back in Table 3.1. As such, the commonly used model for band gap transitions at interfaces, would predict a categorically different transition for the interface of the InP phases, than the one shown in Figure 3.22.

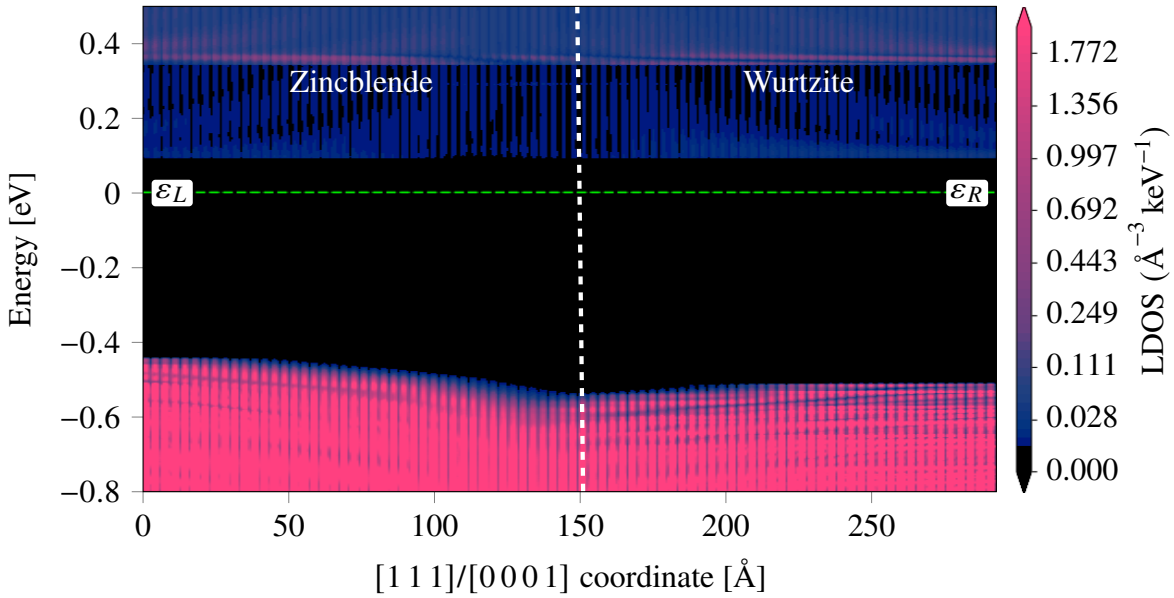


Figure 3.22: Local Density Of States (LDOS) of 127 atomic layer long Zincblende/Wurtzite interface system of InP. The used exchange-correlation functional was GGA PBES and the k-point sampling density along the transport direction was increased to 150 Å. The LDOS was sampled with a k-point density of 14 Å in the transverse directions and an energy spacing of 2.5 meV.

It is obvious from the VBM-bending in Figure 3.22, that the hole transport in InP must be drastically impacted by the formation of Zincblende/Wurtzite interfaces, such as those in RTPs. Even with the flatness of the CBM across the interface, some reallocation of higher-energy states still occur due to the interface, as seen in Figure 3.22, which could, potentially, give rise to a change in the electron transport in InP. This topic will be the subject of the following section, together with the polar optical phonon limited mobility and its anisotropy due to the non-analytic contribution to the dynamical matrix, which was studied in Section 3.3.2.

3.5 Indium phosphide polytypism - electrical properties

One of the most common concerns with defects of any kind in semiconductor devices, is how much they degrade the current conducting abilities of the device. This is therefore the topic of the current section, where firstly the impact of the polytypic defects on the ballistic transport in InP is studied using NEGF. Following this study, the effect of the polar optical phonon anisotropy, found in Section 3.3.2, on the mobility of InP, will also be investigated.

3.5.1 Ballistic transport

To calculate the electrical transport through the high-density polytypic defect systems, the NEGF-DFT formalism described in Section 2.3 is utilised. In order to simulate the high-density of polytypic defects studied thus far with the NEGF formalism, the electrodes of the NEGF systems are set up to be identical to the central regions, which are simply the DFT unit cells used so far, which were illustrated back in Figure 3.13 and 3.14. Pristine Zincblende and Wurtzite NEGF systems were also created for comparison. To ensure a current flow, a small n-type background doping of $1 \times 10^{17}\,\mathrm{cm}^{-3}$ was added to the systems, which is comparable to the experimental background doping from impurities in grown InP [109]. The simulated doping is however not associated with any particular atoms or impurities, but is added as a background charge to the electron density. The transmission spectra for the unbiased systems were then calculated using Eq. (2.39), with the transport direction along the [1 1 1]/[0 0 0 1]-directions, after which the currents for increasing applied bias voltages were calculated with Eq. (2.38). Since the interest in studying transport properties through defects, lies in how they impact the transport of an otherwise pristine semiconductor, the conductivities relative to the pristine Zincblende InP system are plotted in Figure 3.23.

It can be seen that, in the Ohmic region, i.e. the low-bias region where I-V curves are linear, pristine Wurtzite exhibits a 5% higher conductivity than pristine Zincblende along the [1 1 1]/[0 0 0 1]-direction. As expected, it is also seen that the polytypic defects generally decrease the conductivity in this region. Surprisingly however, the largest reduction is seen in the systems with the fewest defects, in particular the system with a single RTP every 9 layers of Zincblende, which displays a 32% reduction in conductivity relative to pristine Zincblende InP. Based on this, it is likely that there is a trade-off between the negative effect of the polytypic defects perturbing the systems, and the positive effect of the systems becoming closer to being Wurtzite, This assumption also explains why the system with the highest conductivity, after the pristine Wurtzite system, is the system with 4 layers of Wurtzite every 3 layers of Zincblende, as this system has the highest content of Wurtzite segments out of all the studied systems.

At voltages above approximately 30 mV, it is seen that the conductivities relative to pristine Zincblende starts to increase, in particular for the pristine Wurtzite system, the relative conductivity reaches a 13% increase in the 0.1 V to 0.7 V range, before falling below the conductivity of Zincblende at 0.9 V. Due to the lower band gap in these calculations, stemming from the use of the GGA PBES functional, both Zincblende and Wurtzite are

beyond their electrical breakdown at this voltage. At voltages above the electrical breakdown of a system, the current is no longer linear with the applied voltage and the conductivity starts to increase rapidly. Since the band gap of Zincblende is lower than that of Wurtzite, it reaches electrical breakdown at a lower voltage, and its conductivity thus starts to increase earlier than for Wurtzite. This is the reason for the sharp drop in relative conductivity of Wurtzite at 0.9 V, and by extension also for the polytypic systems, as their conductivities resemble that of Wurtzite, with their similarity depending on their Wurtzite content.

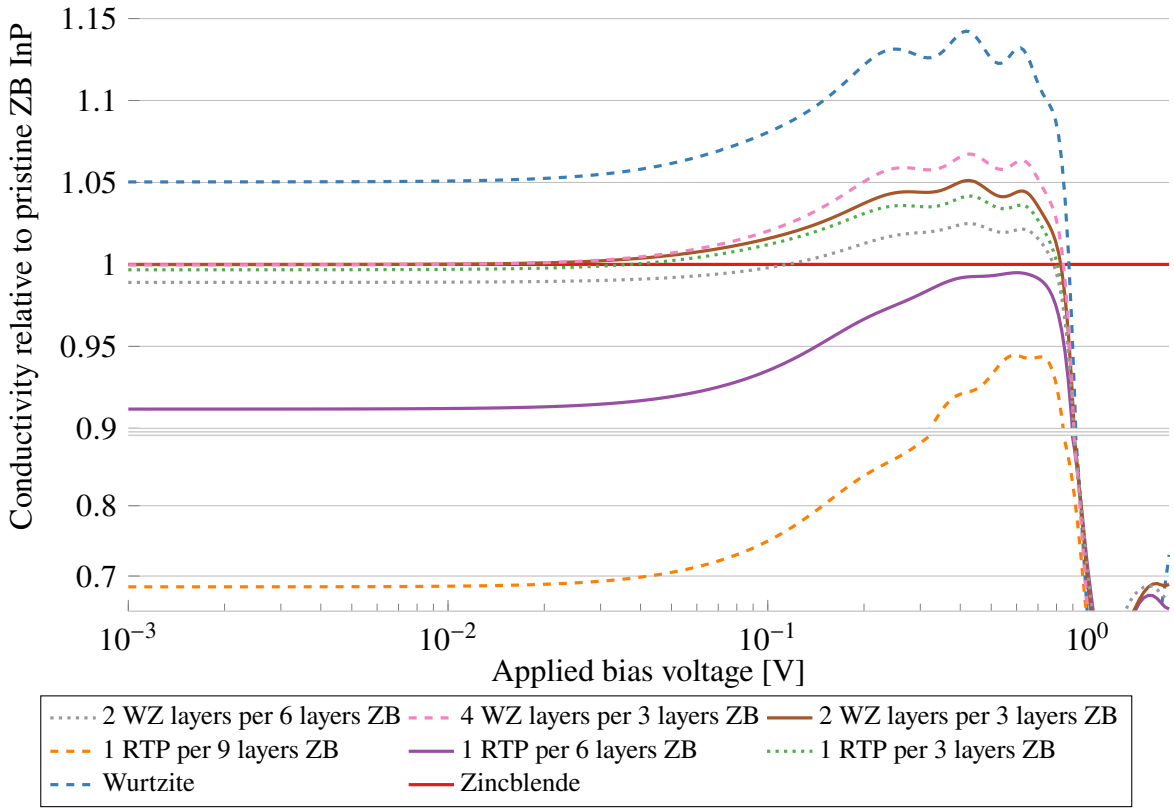


Figure 3.23: Differential conductivity of InP polytypic defect systems relative to pristine Zincblende, at an intrinsic n-type background doping of 1×10^{17} cm^{-3}. The ballistic transport transmission spectra are calculated from unbiased NEGF-DFT calculations with GGA PBES and with the transport direction along the [1 1 1]/[0 0 0 1] crystallographic direction.

Since InP is used for the n- and p-regions of the PIN photodiode of Section 1.4, it is highly relevant to investigate how the Ohmic conductivity of the polytypic systems change with various doping levels of either n- or p-type. Since various doping concentrations are simulated by shifting the Fermi level, this study is also relevant for InP as a transistor channel material, since the Fermi level is in this case controlled by the gate voltage. The Ohmic conductivity, which is the constant conductivity in the low-bias range shown in Figure 3.23, is calculated at zero bias voltage for the all the polytypic systems, with various Fermi level shifts corresponding to certain doping levels. The resulting conductivities relative to the pristine Zincblende InP system are plotted in Figure 3.24.

It is seen that, the conductivities for increasing n-type doping levels behave identically to the conductivities for increased bias voltage in Figure 3.23, with an increase in the relative conductivity prior to electrical breakdown depending on the systems Wurtzite content. For

the increasing p-doping the relative conductivities slowly decrease, until a doping level of around 4×10^{19} cm^{-3}, after which the conductivities fall sharply. For Wurtzite the relative conductivity falls to a mere 5% of the conductivity of pristine Zincblende, and for the polytypic systems the conductivities fall similarly depending on their Wurtzite content. Yet this is untrue for the system with 1 RTP every 9 layers of Zincblende, the system with the lowest Wurtzite content, for which the conductivity falls to a level in-between the other polytypic systems. This would suggest that there exists a lower limit in the Wurtzite content, in this doping region, after which the conductivity does not approach that of pristine Zincblende and instead becomes further perturbed by the defects. This limit appears to have been reached in the systems with 6 layers of Zincblende in-between the polytypic defects, where the relative conductivity at high p-doping levels reach a value of 38% of the pristine Zincblende system.

This discussion, along with other conclusions drawn from the results of the study in Figure 3.24 at high doping levels, should only be taken tentatively. This is the case, because the underlying assumption that the structure and energy levels of the systems are unchanged with the doping level, breaks down at high doping levels. How high a doping level needs to be for this assumption to break down, depends not only on the system but also on the specifications of the simulation and the implementation of the doping.

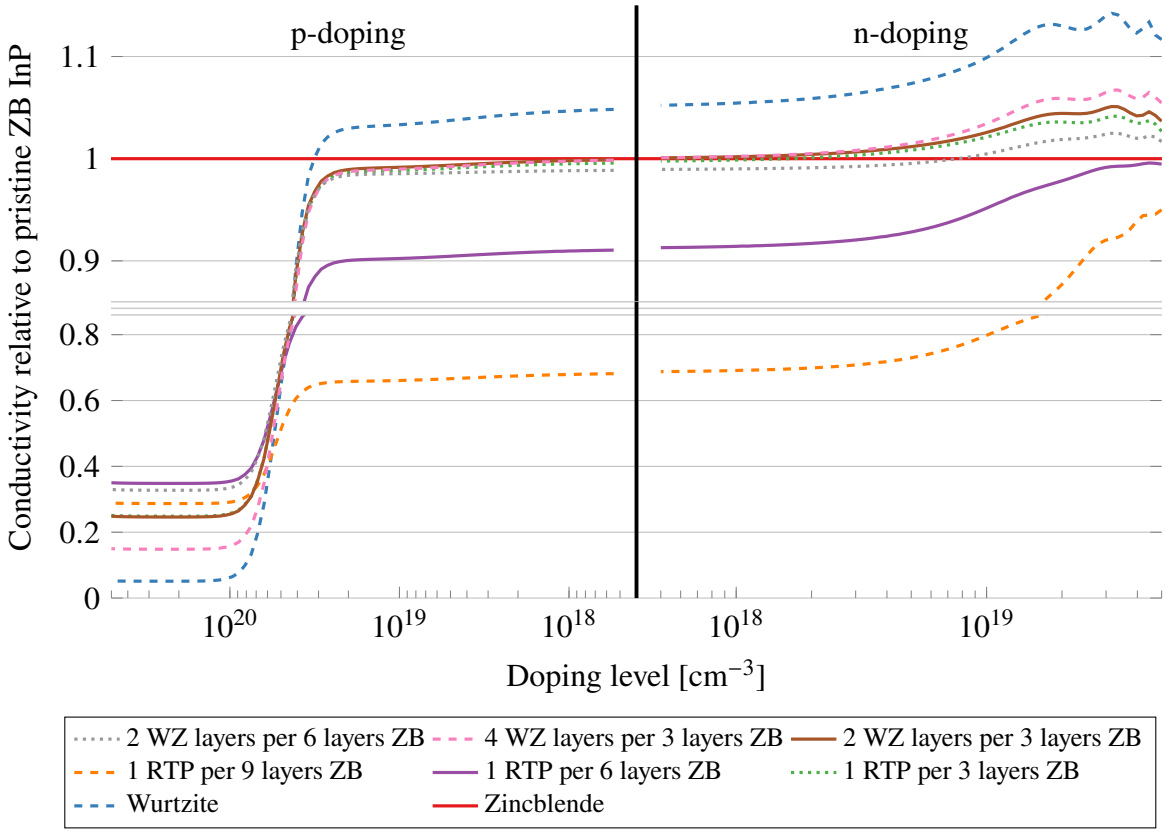


Figure 3.24: Differential Ohmic conductivity of InP polytypic defect systems relative to pristine Zincblende, calculated at zero bias voltage and with various doping levels. The ballistic transport transmission spectra are calculated from unbiased NEGF-DFT calculations with GGA PBES and with the transport direction along the [1 1 1]/[0 0 0 1]-direction.

Based on the results from Figure 3.23 and 3.24, the highest conductivities are reached in systems with the highest contents of Wurtzite and having fewer polytypic defects in Zincblende

only worsens the conductivity. The conductivity of pristine Zincblende is however higher than the defect systems, for most doping regions and applied voltages, and as such the conductivity of a Zincblende system with polytypyc defects should eventually increase again, as the density of defects approach zero. To explore this hypothesis, NEGF systems are set up with pristine Zincblende InP in both electrodes and singular polytypic defects in the central regions, rather than periodically repeating defects. The studied systems contain one, two and three RTPs, with a separation of 3 Zincblende layers between each one, as well as a Wurtzite region of two and three layers thickness. The relative conductivities are calculated again and plotted versus the applied voltage in Figure 3.25.

As in the dense polytypic defect systems, the relative conductivities are seen to increase after the Ohmic region, and to fall sharply at the electronic breakdown of about 0.9 V. Surprisingly, this behaviour is observed even in the extreme case of a single RTP in an otherwise pristine Zincblende lattice, where the Ohmic conductivity is reduced by 15% compared to pristine Zincblende. The three small peaks in the region between the Ohmic region and the electrical breakdown, characteristic of the Wurtzite system, are however not present in these systems. The sharp drop in relative conductivity at 0.9 V is also less severe, showing that the systems does behave closer to pristine Zincblende than the dense polytypic defect systems. For this same reason, the trade-off between the increased conductivity from the Wurtzite segments, and the perturbation of the defects, is also not present in Figure 3.25, as the conductivity simply falls with an increasing number or width of defects. This inevitably leads to the system with the lowest relative conductivity being the system with a single 3-layer Wurtzite segment, at a relative conductivity of slightly under 60% compared to pristine Zincblende.

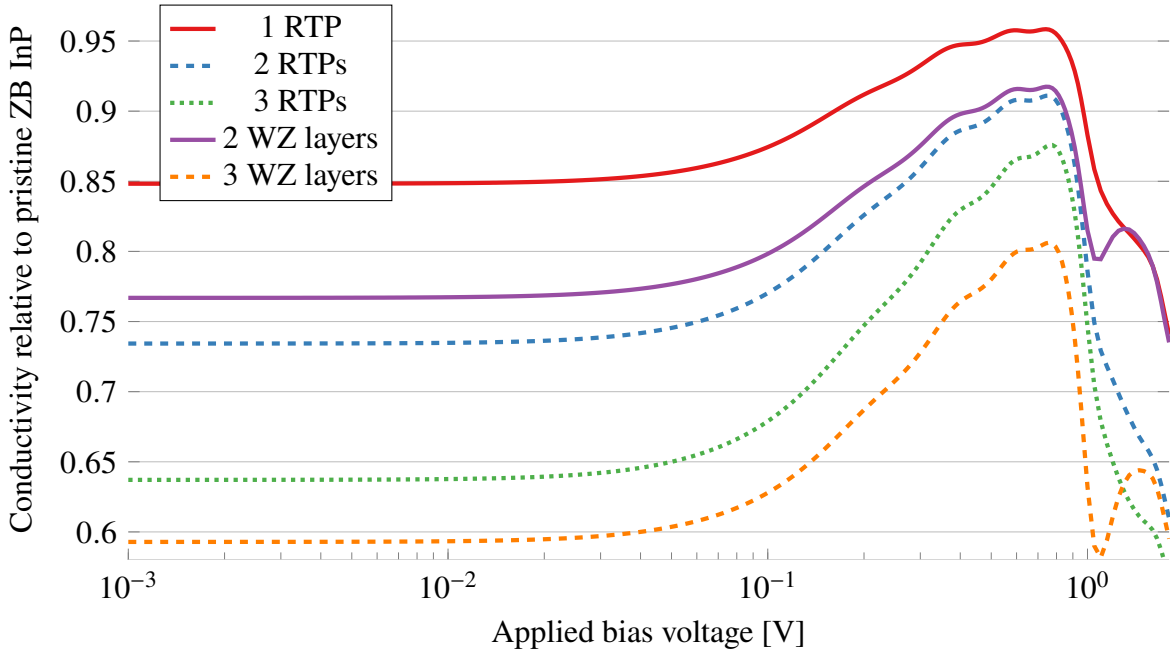


Figure 3.25: Differential conductivity of InP systems with singular polytypic defects relative to pristine Zincblende, at an n-type background doping of 1×10^{17} cm^{-3}. The ballistic transport transmission spectra are calculated from unbiased NEGF-DFT calculations with GGA PBES and with the transport direction along the [1 1 1]/[0 0 0 1] crystallographic direction.

The relative conductivities of these singular polytypic defect systems for various doping levels are also calculated, and are plotted in Figure 3.26. These conductivities illustrate mostly the same results as in Figure 3.25, with a surprisingly Wurtzite-like behaviour of increased relative conductivitities at high n-type doping levels and a sharp conductivity drop at a p-type doping level of $4 \times 10^{19}\,\mathrm{cm}^{-3}$. The large p-type doping conductivity drop is however far less severe in these systems, as seen in the system with a single RTP, which only drops from 85% to 79% conductivity relative to pristine Zincblende. The reduced conductivity drop quickly reaches the level of the dense polytypic defect systems though, as the system with a single 3-layer Wurtzite segment drops to only 40% relative conductivity. Unmistakably, in the extrema of singular defects in Zincblende InP, the polytypic defects severely hinder current transport along the $[1\,1\,1]/[0\,0\,0\,1]$-direction. This stands in contrast to the opposite extrema, where a high density of Wurtzite segments causes little reduction to the Ohmic conductivity and can actually increase the conductivity at high voltages/n-type doping levels.

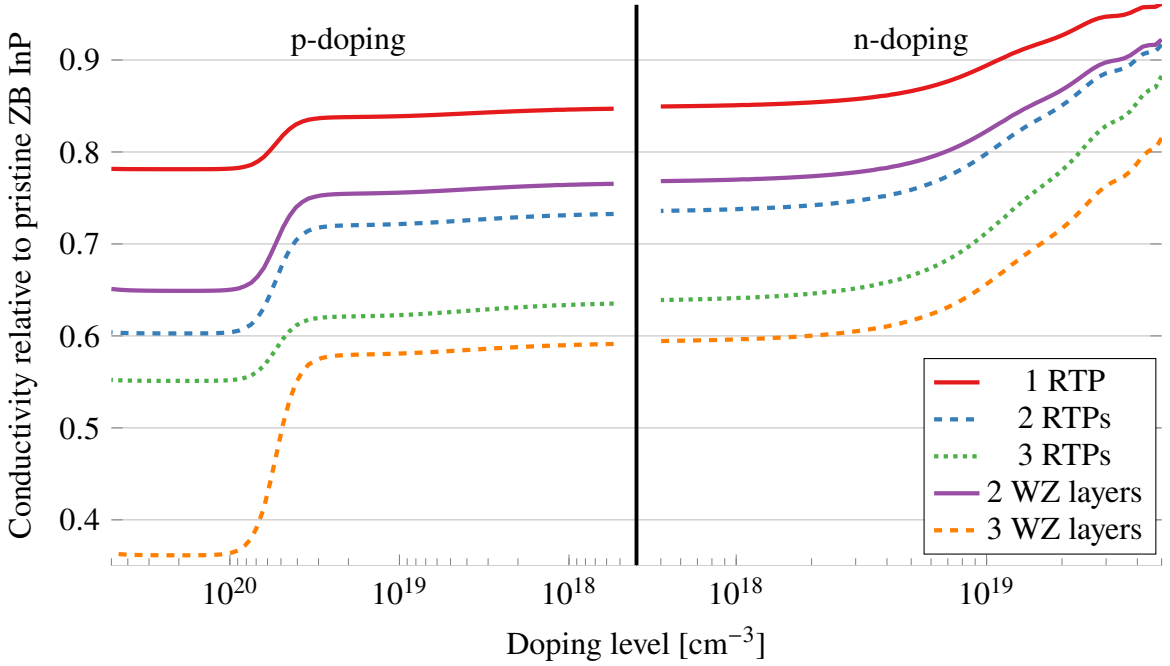


Figure 3.26: Differential Ohmic conductivity of InP systems with singular polytypic defects relative to pristine Zincblende, calculated at zero bias voltage and with various doping levels. The ballistic transport transmission spectra are calculated from unbiased NEGF-DFT calculations with GGA PBES and with the transport direction along the $[1\,1\,1]/[0\,0\,0\,1]$-direction.

Similarly to the the K- and A-direction anisotropy explored in Sections 3.3 and 3.4, it is likely that there is also an anisotropy in the conductivity in the polytypic defect systems. While the NEGF calculations of the present section lacks the electron-phonon interaction, which induced the anisotropy in the earlier studies, an anisotropy should also arise in these calculations, due to the difference between conducting current across the polytypic defects, along the $[1\,1\,1]/[0\,0\,0\,1]$-direction, and conducting current along the polytypic defects, for example along the $[1\,1\,\bar{2}]/[1\,0\,\bar{1}\,0]$-direction. The conductivity calculations of the dense polytypic defect systems are therefore repeated, with the transport in the $[1\,1\,\bar{2}]/[1\,0\,\bar{1}\,0]$-direction, and the resulting conductivities relative to pristine Zincblende are plotted in Figure

3.27 against the applied bias voltage.

Since Zincblende is isotropic, the relative conductivities shown in Figure 3.27 can be readily compared with the conductivities along the [1 1 1]/[0 0 0 1]-direction in Figure 3.23. A number of differences are immediately apparent upon inspection of the relative conductivities, for one the pristine Wurtzite system exhibits a slightly lower Ohmic conductivity than the pristince Zincblende system in this transport direction, in contrast to the increased conductivity along the [1 1 1]/[0 0 0 1]-direction. Another difference is the increased conductivity in the region between the Ohmic region and the electrical breakdown, where the Wurtzite system only shows two conductivity peaks along the $[1\,1\,\bar{2}]/[1\,0\,\bar{1}\,0]$-direction, and with a maximum height of only approximately 8%, as compared to the three peaks and 13% along the [1 1 1]/[0 0 0 1]-direction. Yet the most interesting result in Figure 3.27, is the increased relative Ohmic conductivity of most of the polytypic defect systems, despite the Wurtzite system having a lower conductivity in this transport direction. The system displaying the highest conductivity is the system with 1 RTP every 6 layers of Zincblende, which exhibits a conductivity approximately 9% higher than the pristine Zincblende system. There appears to be little systematic relationship between the increased conductivity and the density of the polytypic defects, except that the systems with 6 Zincblende layers in-between the polytypic defects have higher conductivites than systems with 3 or 9 layers of Zincblende.

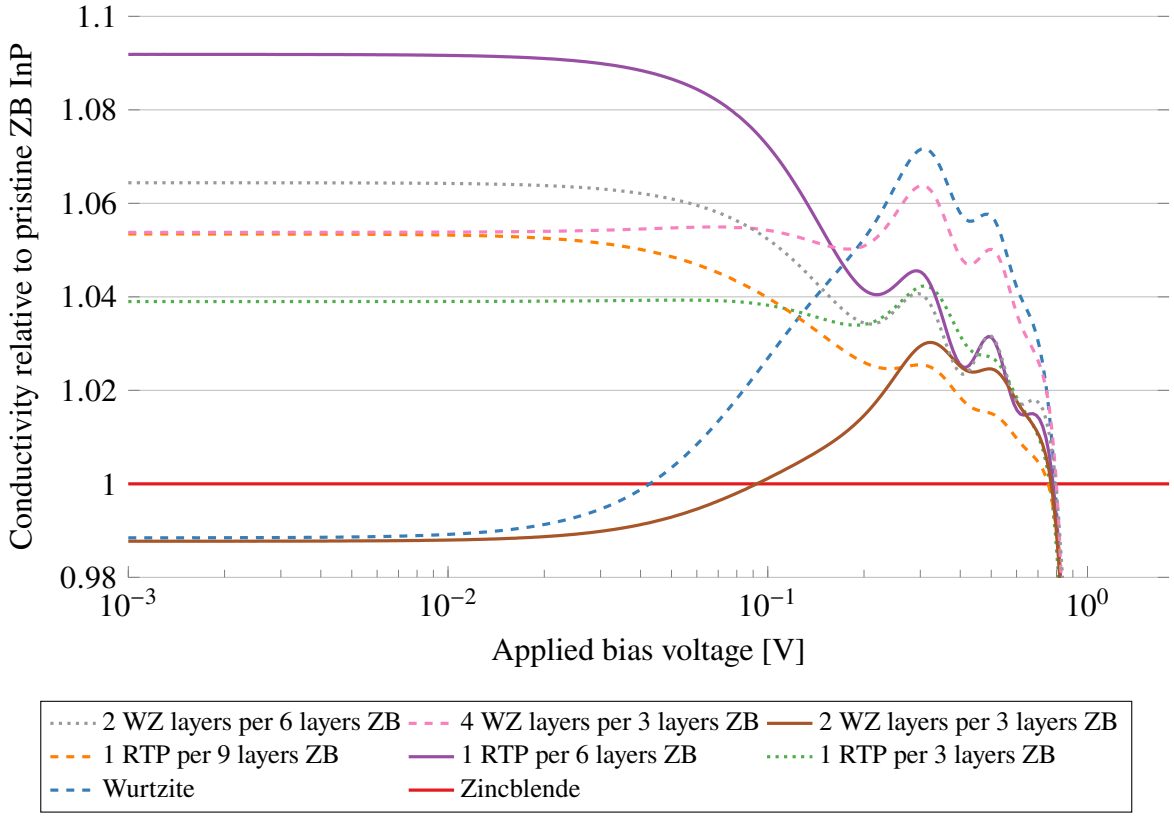


Figure 3.27: Differential conductivity of InP polytypic defect systems relative to pristine Zincblende, at an intrinsic n-type background doping of 1×10^{17} cm^{-3}. The ballistic transport transmission spectra are calculated from unbiased NEGF-DFT calculations with GGA PBES and with the transport direction along the $[1\,1\,\bar{2}]/[1\,0\,\bar{1}\,0]$ crystallographic direction.

Similarly to the previous calculations, the Ohmic conductivities of the systems are also plotted against different doping levels in Figure 3.28. Interestingly, in contrast to the transport along the [1 1 1]/[0 0 0 1]-direction, the relative conductivities of the defect systems in this direction increase with increasing p-doping level, until the sudden drop at $4 \times 10^{19}\,\mathrm{cm}^{-3}$. At the high p-doping levels beyond this sudden drop in relative conductivity, the systems behave similarly to their [1 1 1]/[0 0 0 1]-direction behaviour. That is, the Wurtzite system exhibits the lowest relative conductivity at around 5% of pristine Zincblende and the defect systems appear in order of their Wurtzite content. Similarly, at high n-type doping levels, or bias voltages between the Ohmic region and electrical breakdown, the systems begins to act like in the [1 1 1]/[0 0 0 1]-direction, with the Wurtzite system exhibiting the highest relative conductivity and the defect systems following after depending on their Wurtzite content.

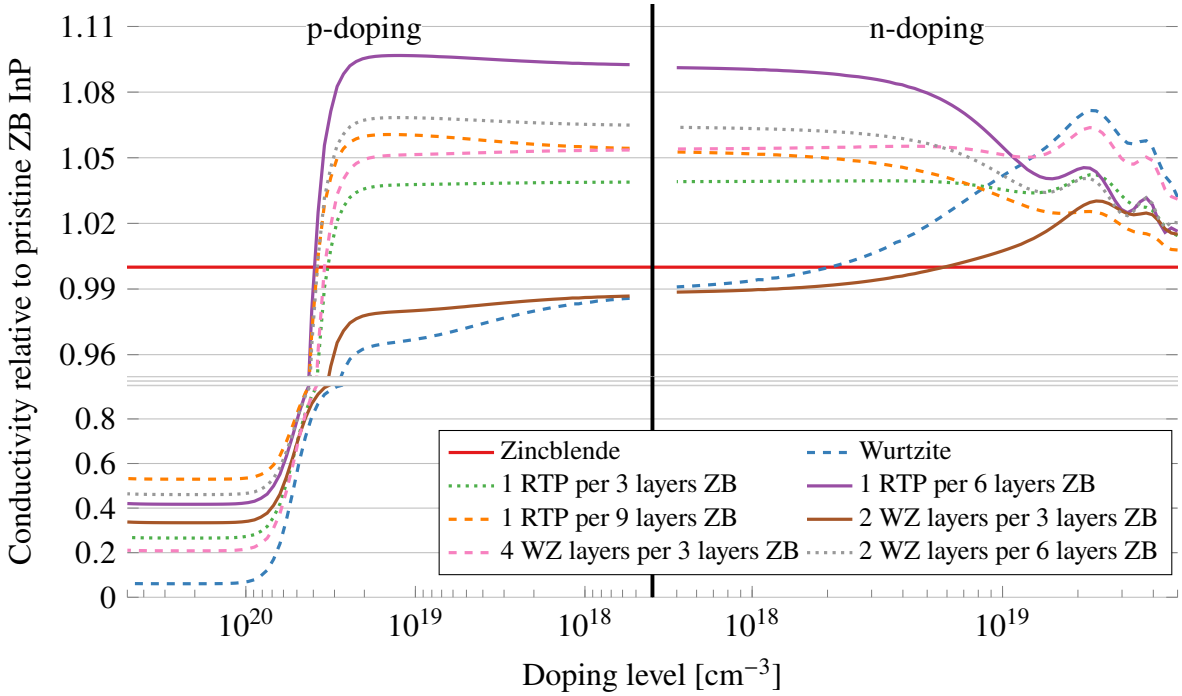


Figure 3.28: Differential Ohmic conductivity of InP polytypic defect systems relative to pristine Zincblende, calculated at zero bias voltage and with various doping levels. The ballistic transport transmission spectra are calculated from unbiased NEGF-DFT calculations with GGA PBES and with the transport direction along the $[1\,1\,\bar{2}]/[1\,0\,\bar{1}\,0]$-direction.

From these calculations of the transport along the $[1\,1\,\bar{2}]/[1\,0\,\bar{1}\,0]$-direction in the polytypic systems, it can be concluded that the defects act as pathways of lesser resistance, thereby increasing the conductivity of the systems along this direction. Based on this, the defect system with the highest conductivity should be the system with the most polytypic defects, rather than the system with 1 RTP every 6 layers of Zincblende. This deviation can be explained from the calculated strain in the polytypic defects, as reducing the periodicity of RTPs from 1 RTP every 3 layers, to 1 RTP every 6 layers, induce an additional 0.07% strain in the system. This should be considered relative to the total strain between Wurtzite and Zincblende of 0.37%, i.e. it is a relatively large increase in strain. As was mentioned in Section 1.3 in the case if silicon, strain can increase the conductivity of a semiconductor by changing the energy bands, and as such the increased conductivity from the larger strain

in the lower density defect system, apparently outweighs the loss in conductivity due to reducing the number of polytypic defects. This effect has a clear optimum at 1 RTP every 6 layers of Zincblende, as further reducing the periodicity of RTPs to 1 RTP every 9 layers of Zincblende, only increases the strain by an additional 0.003%.

3.5.2 Mobility

In the classical drift-diffusion model of current transport, charge carriers move unperturbed through a material, until they scatter from collisions with various scattering centres into new directions and energies. As such, the main parameter which determines the current transport of a material, is the average time between scattering events, called the relaxation time, and its inverse, the scattering rate. The scattering rate of a material depends on its relevant scattering mechanisms, which can include among others carrier-carrier scattering, ionised impurity scattering, acoustic phonon scattering, piezoelectric scattering and Polar Optical Phonon (POP) scattering. The dominant scattering mechanisms in a polar semiconductor such as InP, is piezoelectric scattering at very low temperatures in a very pure sample, which is simply acoustic phonon scattering with polar acoustic phonons, and POP scattering at room temperature and above [110]. A simple phenomenological model of the POP scattering rate, based on Fermi's golden rule, is given in Eq. (3.7), where the terms containing $(N_0 + 1)$ are phonon emission terms, where the carriers scatter to lower energies by emitting a phonon, and the terms containing N_0 are phonon absorption terms [110]:

$$\begin{aligned}\frac{1}{\tau_p(m)} = \frac{q^2\omega_0\left(\frac{\epsilon_0}{\epsilon_\infty}-1\right)}{4\pi\epsilon_0\varepsilon_0\hbar\sqrt{2E(p)/m^*}}\Bigg[N_0\sqrt{1+\frac{\hbar\omega_0}{E(p)}}+(N_0+1)\sqrt{1-\frac{\hbar\omega_0}{E(p)}} \\ -\frac{\hbar\omega_0 N_0}{E(p)}\sinh^{-1}\left(\frac{E(p)}{\hbar\omega_0}\right)^{1/2}+\frac{\hbar\omega_0(N_0+1)}{E(p)}\sinh^{-1}\left(\frac{E(p)}{\hbar\omega_0}-1\right)^{1/2}\Bigg]\end{aligned} \tag{3.7}$$

In Eq. (3.7) τ_m is the POP momentum relaxation time, which is larger than the associated energy relaxation time τ, since POP scattering is a non-isotropic scattering mechanism, i.e. it favours small angle scattering events, which has less of an impact on the momentum of the carrier and thereby the current. For the remaining parameters, $E(p)$ is the electronic energy at momentum p, ω_0 is the frequency of the polar optical phonon and N_0 is its Bose-Einstein occupation, at energy $\hbar\omega_0$ and a temperature T. The parameters q and m^* are the charge and effective mass of the carrier, ε_0 is the vacuum permittivity and ϵ_0 and ϵ_∞ are the static and high-frequency dielectric constants respectively, which were described in Section 3.4.1.

Despite its simplicity, the model in Eq. (3.7) catches the anisotropy in the POP scattering rates of the polytypic defect systems, due to the anisotropy in the dielectric constants and the phonon frequencies. To visualise this anisotropy, the electron transport POP scattering rates of Wurtzite along the K- and A-directions are plotted in Figure 3.29, along with the POP scattering rate of Zincblende. To better compare the Wurtzite scattering rates with those of Zincblende, the Zincblende energies are shown shifted by the difference in Conduction Band Minima (CBM) between the systems, so that their conduction onsets are aligned. It is

readily apparent, that the scattering rates are mostly constant in the energy range where only phonon absorption is allowed. This energy range starts at the CBM, which is approximately at 0.31 eV for Wurtzite, and ends at 0.35 eV, which is approximately the energy of the CBM plus the energy of the polar longitudinal optical phonon. At this energy and above, phonon emission is also allowed, since the conduction band electrons now have high enough energies to emit an optical phonon and still remain in the conduction band. This yields a sharp and sudden increase in the scattering rate, which slowly falls off with the energy. Looking at the scattering rates in the phonon absorption plus emission region, it can be clearly seen, that there is a sizeable difference in the scattering rates between the Zincblende and Wurtzite phases, which originates from the differences in their calculated dielectric constants. Considering the dielectric constants of the polytypic defect systems, and their anisotropies, in Table 3.3, this would suggest considerable differences in the systems electronic transport properties, as well as a notable anisotropy. Due to the low anisotropy of the calculated dielectric constants of Wurtzite, the difference between its K- and A-direction scattering rates in Figure 3.29 are small, yet still distinctly visible.

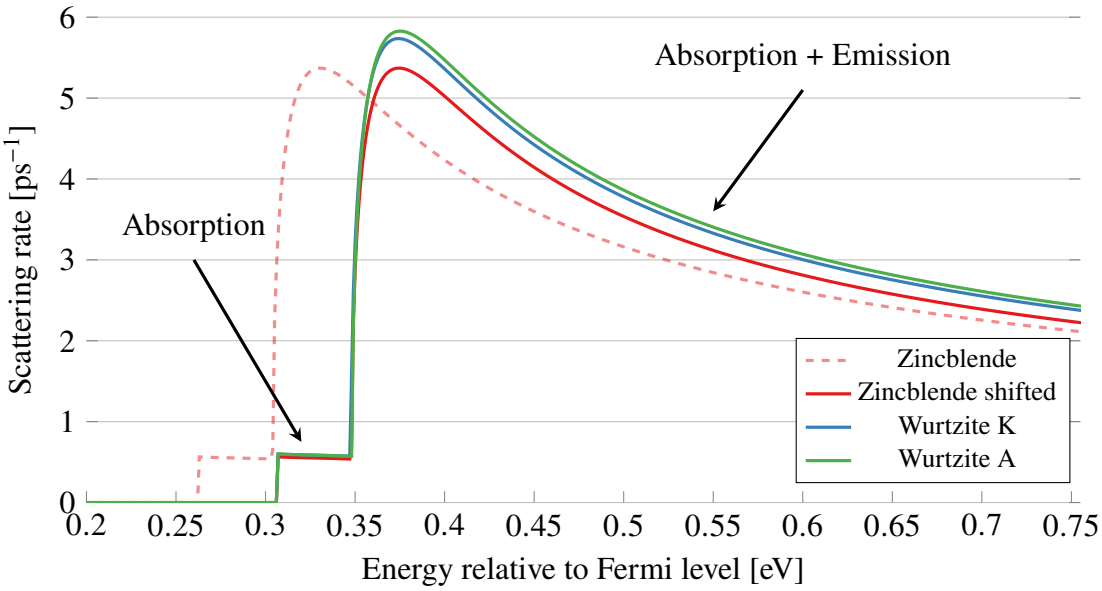


Figure 3.29: Polar Optical Phonon (POP) scattering rates of Zincblende and Wurtzite InP, calculated with Eq. (3.7) at 300 K using parameters from Table 3.2 and 3.3. Wurtzite scattering rates are calculated along the K- and A-directions, using the appropriate parameters. The Zincblende scattering rate is also shown with its energies shifted to align with the Wurtzite conduction band minima.

The POP scattering rate in Eq. (3.7) is calculated at 300 K for all the polytypic defect systems, using the parameters from Table 3.2 and 3.3. The scattering rates along the A- and K-directions of the systems are then used to compute the electron mobilities, using the semi-classical Boltzmann transport equation method as implemented in QuantumATK T-2022.03, which is based on the implementation in Ref. [111]. These mobilites are illustrated in Figure 3.30, along with the resulting anisotropies and the experimental Zincblende InP electron mobility, which is taken from the Ioffe database [102].

It is immediately seen that all the calculated mobilities far exceeds the experimentally

observed value. This is due to the simple scattering rate model in Eq. (3.7) being a gross oversimplification of the real scattering rates in the physical systems, which underestimates the POP scattering rate by approximately a factor of 10. The model is however mostly phenomenologically correct, meaning that it replicates the correct physical behaviour, despite the wrong absolute values. For this reason, the model is still usable to draw qualitative conclusions about systems, such as trends and behaviours.

In the mobilities in Figure 3.30, it is seen that along both the A-direction, i.e. [1 1 1]/[0 0 0 1], and the K-direction, the mobilities of the defect systems increase with the systems Zincblende content. Along the A-direction, the mobilities of the defect systems surpasses the mobility of the pristine Zincblende system, whereas along the K-direction, they only approach it. This result can be explained by the changes to the energy of the LO phonon mode in the defect systems, shown in Table 3.2. Along the K-direction, the LO phonon mode is mostly unchanged in the defect systems, meaning that the mobility is simply lowered due to the incorporation of the lower mobility Wurtzite segments. Along the A-direction on the other hand, the LO phonon mode is shifted upwards in energy, which reduces the scattering rate in the phonon absorption region, and thereby increases the mobility. The mobility depends significantly more on the scattering rates at lower energies, compared to those at higher energies, due to the exponential decay of the electron occupations in the Fermi function. This explains the increased mobility along the A-direction in the systems, despite the higher scattering rates at high energies shown in Figure 3.29.

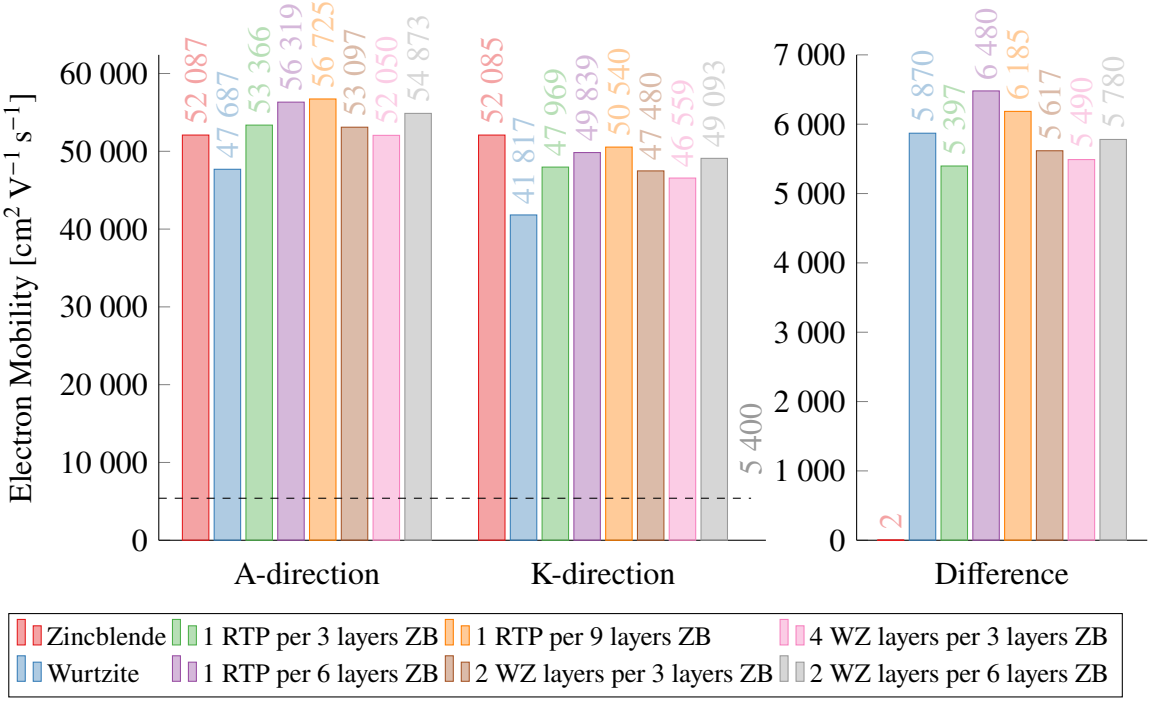


Figure 3.30: Electron mobility of pristine InP and polytypic defect systems along the A- and K-directions at 300 K and a carrier density of approximately 1×10^{19} cm^{-3}, corresponding to a Fermi-level 0.1 eV below the conduction band minimum. Mobilities are calculated with the method described in [111], using the Polar Optical Phonon (POP) scattering rates of Eq. (3.7) and the parameters from Table 3.2 and 3.3. The experimental value of Zincblende InP, taken from Ioffe [102], is shown by the dashed line.

The defect-induced increase in mobility along the A-direction due to POPs shown in Figure 3.30, stands in contrast to the reduced conductivities from the NEGF calculations shown in Figure 3.23. The reduced mobilities along the K-direction similarly contrasts the increased conductivities in the NEGF calculations shown in Figure 3.27. Neither the mobility or NEGF calculations here yield a complete picture of the transport in InP, as both methods lack critically important transport physics. The NEGF calculations lack the critical POP scattering mechanism which dominates the scattering rate and the mobility calculations lack the information about the complicated anisotropic energy landscape of the systems. Evidently, in order to get a full understanding of the transport properties of polytypic defects in InP, it is crucially necessary to include both POP scattering and a full electrical description of the system, to be able to quantify which of these competing defect-related effects have a larger impact on the electrical transport.

3.6 Conclusions

To summarise, in this chapter the research published in Ref. [82], on the electrical and optical properties of InP polytypism, was presented with new discussions and results deriving from phonon dispersion calculations of the same systems. First, the two crystal phases of InP, Zincblende and Wurtzite, were simulated and compared by their electrical bandstructure, phonon bandstructure, and electrical transmission. It was found that the Wurtzite structure exhibits specific properties which are not found in Zincblende, such as a second "dark" conduction band and a discontinuity in its optical phonon dispersion at the Γ-point.

Then the study of polytypic defect systems, i.e. systems which is a mix of Zincblende and Wurtzite, were started out by calculations of their Born effective charges and phonon dispersions. It was found that Wurtzite exhibits an anisotropy in its Born effective charges, in contrast to pristine Zincblende, which induces the Γ-point phonon dispersion discontinuity. It was also discovered, that just a small amount of Wurtzite in a Zinblende system also induces this anisotropy in the Zincblende section of the system, though to a lesser degree than in the Wurtzite section, thereby breaking its isotropy.

After this, the complex dielectric functions of the systems were calculated, and it was found that the polytypic defect systems containing the least Wurtzite exhibited the highest anisotropy in their dielectric constants, despite pristine Zincblende being isotropic. A hypothesis to explain this was proposed, which was that Zincblende actually has a higher anisotropy in its dielectric constants than Wurtzite, once its symmetry along the A-direction is broken.

Following this study, the validity of polytypic defects acting as atomically thin quantum wells were investigated, using both periodic DFT and semi-infinite NEGF methods. These calculations revealed that, in order for a crystal-phase quantum well to have the well-phase band gap in the centre of the well, a width of no less than 10 nm is required. This result excludes random polytypic defects, such as rotational twin planes and small Wurtzite segments, as crystal-phase quantum wells, although they could possibly still be created through crystal-phase engineering.

Finally, the electrical transport through and along the polytypic defects were studied in both the classical collision and the ballistic transport regimes. These studies yielded conflicting results, with the ballistic transport study suggesting that the defects decrease conductivity across the defect plane and increase the conductivity along it, whereas the classical collision study suggested the inverse.

In the ballistic transport study, it was found that for conducting current along the [1 1 1]/[0 0 0 1]-direction, systems containing more Wurtzite segments had higher conductivities than systems with fewer segments, due to the higher conductivity of Wurtzite compared to Zincblende. For transport along the polytypic defect planes, it was found that the defects act as pathways of lower resistance, thereby increasing the electrical transport. This increased transport was found to have an optimum in the polytypic defect content, due to the strain on the defects, and thereby their current conducting abilities, being increased by lowering their occurrence. This optimum in number of polytypic defects versus higher transport per defect, was discovered to be a single rotational twin plane every six layers of Zincblende.

In the classical collision study, it was found that the Γ-point splitting of the longitudinal polar optical phonon dispersion in Wurtzite increases the mobility along the A-direction, due to the increase in phonon frequency and hence relaxation time. This effect was found to compete with the general lower mobility of the Wurtzite phase, as compared to the Zincblende phase, yet an optimum Wurtzite fraction was not found. This is the case, due to the fact that the highest mobility was discovered in the system containing the least amount of Wurtzite out of the simulated systems, the system with a single rotational twin plane every nine layers of Zincblende.

Chapter 4

SRH Recombination and Trap Levels in $In_{0.53}Ga_{0.47}As$ Point Defects from First Principles

This chapter is based on research that was published in Ref. [112], but is expanded with additional findings and discussions that were deemed unfit for the aforementioned publication. The chapter is split into six sections, the first section motivates the specific research of the chapter and introduces the two different methods to simulate random compounds in first-principles calculations. The second and third sections goes in-depth with these methods and analyses their respective strengths and weaknesses. The second section additionally includes the obligatory first-principles convergence study. In the fourth section, transition levels of various point-defects are investigated and the most important ones are selected. The selected defects are then further investigated in section five, by calculations of their device-impairing properties and in section six the main results and conclusions of the chapter are summarised.

For the DFT simulations in this chapter, version U-2022.12 of QuantumATK by Synopsys was used [83]. For the DFT basis set, the non-orthogonal atom-centred local basis set, PseudoDojo, was used together with its corresponding pseudopotential [84]. The remaining non-default simulation parameters will be given in the convergence study in Section 4.2.1.

4.1 Prelude

4.1.1 Motivation

Besides being the proposed material for the photon-absorbing intrinsic region of the PIN photodetector described in Section 1.4, InGaAs finds usage in many other existing applications, due to its high electron mobility, and the versatility of its direct band gap and lattice constant being tunable, by controlling the indium to gallium ratio of the compound. It is favoured over Ge for all kinds of infrared photodetectors in the same spectral region, due to its faster response time, higher quantum efficiency and lower dark current [113]. These traits make InGaAs not only useful for the PIN heterojunction photodiode, but also for other photodetec-

tors such as avalanche photodiodes [50]. Due to the tunability of the direct InGaAs band gap, and its low exciton binding energy, InGaAs is also used for long wavelength lasers based on quantum confinement, such as quantum wells and quantum dots [40, 114]. $In_xGa_{1-x}As$ is also used as the last of the photon-absorbing layers in the current state-of-the-art quadruple solar cell junctions [43, 44, 115], due to having a compatible direct band gap, lattice constant and absorption window with the other complementary III-V semiconductor compounds, which have absorption windows at higher energies.

Apart from its optical properties, InGaAs also has extremely useful electrical properties, such as an incredibly high electron mobility, approximately six times larger than bulk silicon. This extremely high electron mobility makes InGaAs an ideal candidate as a channel-material replacement for n-type metal-oxide semiconductor (nMOS) transistors, and in particular for low-power applications, in order to continue the steady improvement promised by Moore's law [29, 30, 35]. In fact, the particularly high electron mobility of InGaAs is already utilised in HEMTs [31] and HFETs, in conjunction with InP, [33], thereby validating the possibility and usability of using InGaAs as a channel material in transistor devices.

The principal issue which prohibits the incorporation of InGaAs as a channel material in MOSFETS, is the lack of a high-quality intrinsic passivation oxide, such as silicon dioxide for silicon. The poor quality of the interface between InGaAs and its native oxide, causes surface strain, imperfect passivation and a particularly large amount of interface defects, which in turn reduces current transport through trap-assisted recombination [116]. Several methods have been developed and investigated in order to improve the interface quality and thereby reduce the number of defects. These methods include a deposition of an additional passivation film, many different kinds of sulphur-passivation, plasma-passivation and regular temperature annealing [117, 118]. While these methods have yielded different degrees of success, none have yet adequately removed the defects caused by the interface, thereby impelling research into the understanding and origin of the measured interface defect state densities.

The origin of these defect state densities have been researched and recently a hypothesis has been proposed [119, 120]. It is hypothesised that the InGaAs/oxide interfaces' strain relaxation induces intrinsic bulk defects in the InGaAs below the interface, such as vacancies and antisites, in addition to the dangling bonds from the imperfect passivation at the interface [121–123]. This hypothesis is based on comparisons of accurate *ab initio* calculations of defects, with the interface defect state densities of different III-V/oxide interfaces subjected to the various passivation techniques. This origin of the defect state densities would explain the limited success of the different passivation methods, since most methods simply attempt to passivate the dangling bonds at the interface.

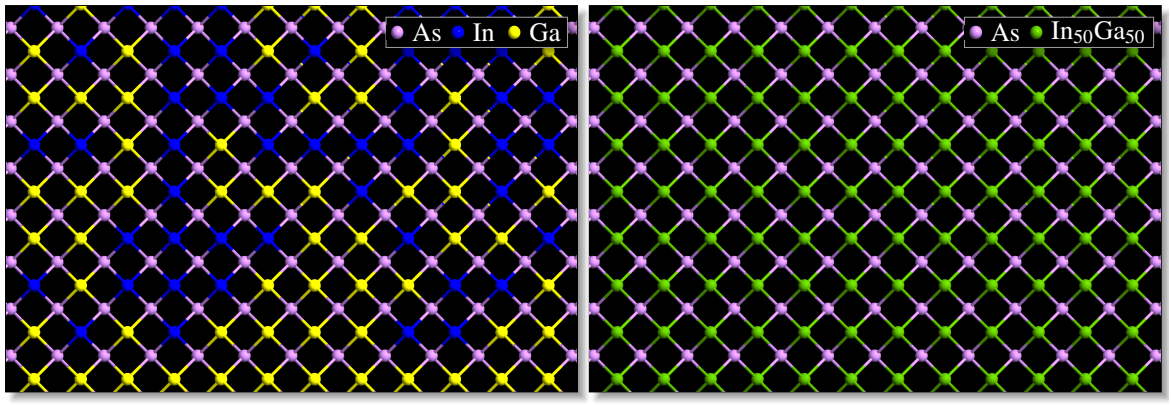
Evidently, both the performance of the InP/InGaAs/InP PIN photodiode of Section 1.4 [63] and the adaptation of InGaAs as a channel material substitute in modern MOSFETS, are hindered by bulk point defects in InGaAs. As such, in order to guide the future passivation research efforts of InGaAs, in order to remove these device-impairing defects, the goal of the present thesis chapter is to investigate various point defects in InGaAs and their impact on device performance.

4.1.2 Atomistic simulation methods of random alloys

To simulate random alloys, such as compound semiconductors like $In_{0.53}Ga_{0.47}As$, in atomistic simulations, two distinct approaches can be utilised. The first is to use a truly random alloy configuration, wherein each atom appears fractionally, according to the elemental composition. Herein lies the first issue with this method, which is that any computer-generated random number is not truly random, but rather the result of a sophisticated pseudo-random number generator. This remains to be an issue with any stochastic simulation method, despite significant progress in the field [124]. With the advent of quantum computers, this issue can finally be truly addressed, by using the inherent randomness of quantum mechanics, to derive truly random numbers [125]. But until such a solution exists, this will continue to be an issue, albeit a minor one in the use-case discussed here. A more significant issue with truly random alloy configurations, are the relatively large system sizes required, sometimes prohibitively large, to accommodate the required mole fractions and to avoid spurious effects of the periodic images of the structure. Additionally, one should in principle use an ensemble of such random systems and average out the random fluctuations, to avoid emphasising statistical outliers such as large conglomeration of identical species.

The latter of these issues can be alleviated to some extent by using Special Quasi-random Structures (SQS) [126], instead of the truly random alloy configurations. SQSs are structures that encompasses the important interactions of the truly random ensemble, while excluding the effects from the rare outliers such as systems with heavy conglomeration. Therefore their usage enables a heavily reduced ensemble size for truly random alloy simulations.

Despite the reduction of the ensemble size by using SQSs, random alloy simulations are still very computationally demanding due to the large system size requirement. An alternative compound semiconductor simulation method is the so-called Virtual Crystal Approximation (VCA). This simulation method resembles how compound semiconductors are simulated in finite element simulations. In the VCA the alloy atoms of species which can occupy the same lattice sites, are replaced with a virtual atom, which is a mix of the species in the same ratio as they appear in the alloy. This is shown in Figure 4.1, where a VCA version and a random-alloy SQS version of $In_{50}Ga_{50}As$ is shown. Using VCA enables the use of minimal system sizes, as the alloy mole fraction is represented in each individual virtual atom. This advantage comes at the cost of random alloy disorder effects, as the alloy is made isotropic in the VCA. Since the virtual atom does not actually exist in nature, there is no guarantee that the results obtained with VCA are physically sound. Therefore VCA needs to be validated on a case-by-case basis.



SQS InGaAs VCA InGaAs

Figure 4.1: Structure of $In_{50}Ga_{50}As$ lattice, simulated using either Special Quasi-random Structures (SQS) or the Virtual Crystal Approximation (VCA), viewed from the [1 0 0]-direction.

In this thesis, both atomistic compound semiconductor simulation methods are explored to simulate $In_{0.53}Ga_{0.47}As$. In the next section, the applicability of the Virtual Crystal Approximation is investigated.

4.2 The virtual crystal approximation

The Virtual Crystal Approximation (VCA) consists of replacing randomly distributed atomic species in an alloy, with virtual atoms which are constructed to mimic the random distribution of the replaced atomic species. Since atoms are represented by their pseudopotentials in DFT, the virtual atom is constructed by mixing the pseudopotentials of the constituent atoms, in the ratio in which they appear in the random alloy. A complication with the VCA arises when using atomically fitted basis-sets like LCAO, as these basis sets are fitted to the specific element which they represent. In QuantumATK this is addressed by using the basis sets of both constituent elements for the virtual atom. While this solution guarantees the inclusion of the physically necessary basis functions, it also leads to issues such as possible over-completeness and increased system-scaling due to the doubled basis set size.

4.2.1 Convergence study

Before discussing the applicability of the VCA to simulate $In_{0.53}Ga_{0.47}As$, it is first necessary to do a basic DFT parameter sweep and convergence study as in Section 3.2.2. The explored parameters are again the k-point sampling density, the occupation function temperature, the density-mesh cutoff, the basis set size and the exchange-correlation functional. The same evaluation criteria are also used, i.e. the total energy, the lattice constant and the band gap. In all calculations, a minimal cubic unit cell of two atoms is constructed with the experimental lattice constant of 5.8687 Å [127]. This unit cell is then relaxed, with the constraint that the Bravais lattice is maintained, until the forces between the atoms are smaller than 0.05 eV $Å^{-1}$ and the stress is lower than 0.1 GPa.

The first parameter to converge, is the k-point sampling density. For the remaining parameters, an occupation function temperature of 1000 K was used, together with a density-mesh cutoff of 105 Ha. HSE was used as the exchange-correlation functional and the basis set was the "Medium" version of the PseudoDojo basis set. The convergence study is shown in Figure 4.2. It is seen that the band gap converges one order of magnitude slower than the total energy. Requiring that the band gap is converged within 1 meV of the final value, thus demands a k-point sampling density of 6 Å, making this semiconductor slightly harder to converge than InP, which only required a k-point sampling density of 4 Å.

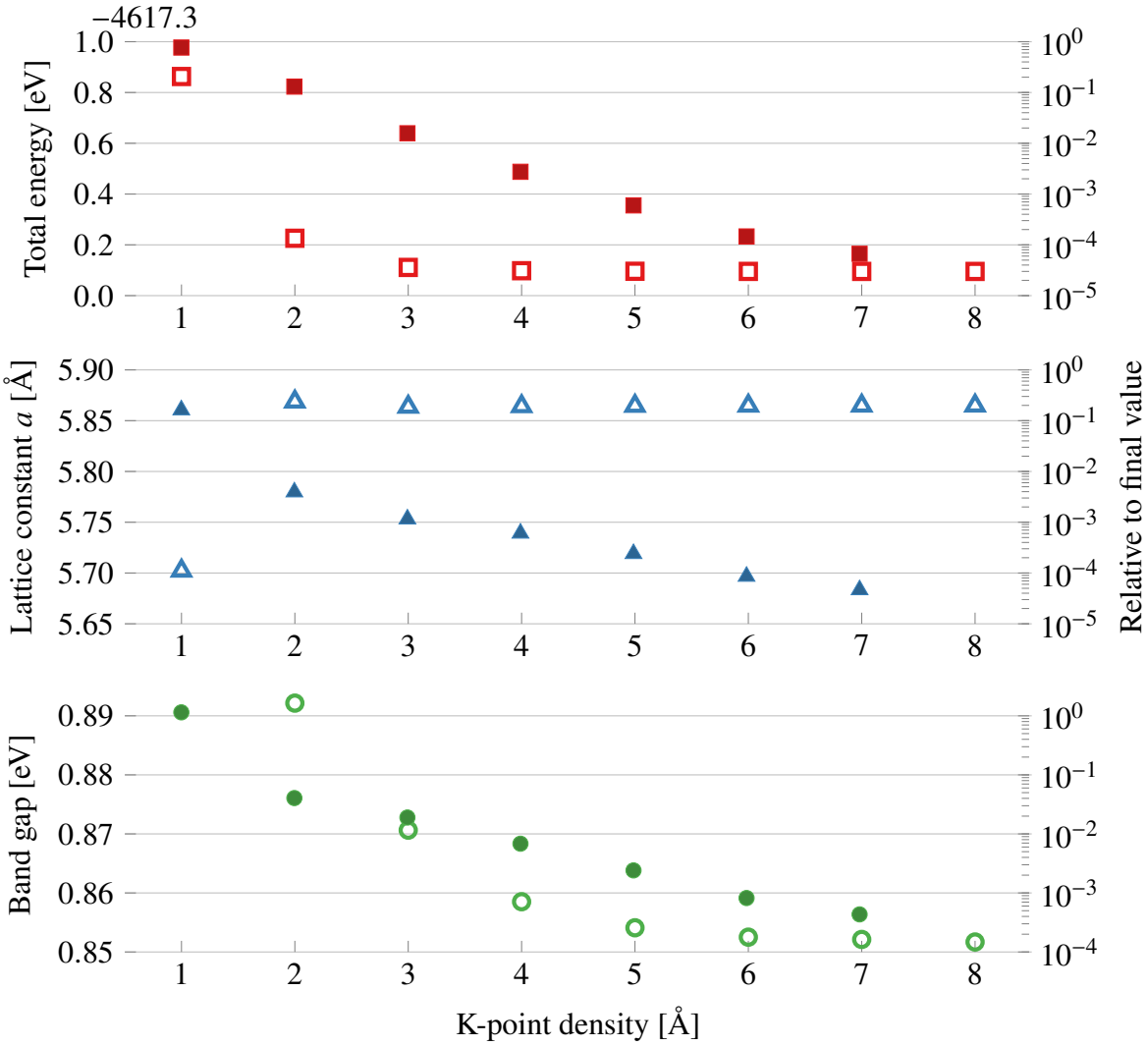


Figure 4.2: Convergence test of Brillouin zone k-point density sampling in $In_{0.53}Ga_{0.47}As$ simulated with the Virtual Crystal Approximation (VCA). Occupation function temperature was 1000 K. Filled symbols indicate values taken relative to the final most accurate value, whereas empty symbols represent the absolute values. The absolute values of the total energies were shifted by −4617.3 eV to facilitate easier reading.

Having converged the k-point sampling density with an occupation function temperature of 1000 K, the second parameter to converge is again the occupation function temperature. The remaining parameters were left unchanged and the second convergence study is shown in Figure 4.3. Like for InP, some convergence behaviour can be seen from the plots relative to the final calculation, but considering the scale of the values, the system appears converged

at all the considered parameter values. The absolute values of the evaluation criteria show numerical differences, which are lower than the trusted accuracy of the calculations, except for perhaps the band gaps, which also does not show any convergence. While the evaluation criteria investigated here does not seem to converge with the occupation function temperature, some calculations such as those carried out in Section 4.5, depends critically on the occupation of states close to the Fermi level. An occupation function temperature of 300 K is therefore chosen for subsequent calculations.

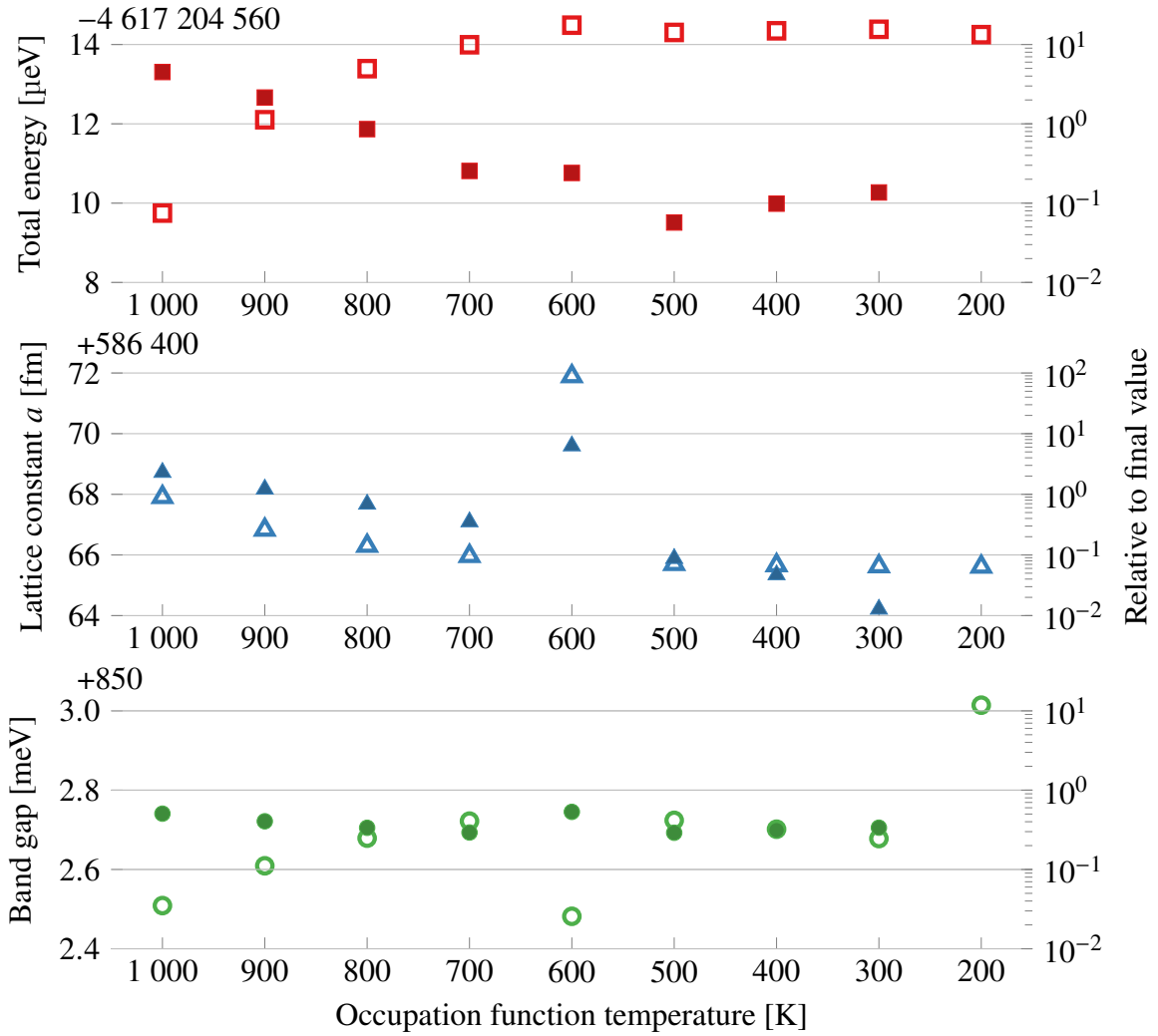


Figure 4.3: Convergence test of occupation function temperature in $In_{0.53}Ga_{0.47}As$ simulated with the Virtual Crystal Approximation (VCA). Filled symbols indicate values taken relative to the final most accurate value, whereas empty symbols represent the absolute values. The absolute values were shifted by −4 617 204 560 µeV, 586 400 fm and 850 meV for the total energies, lattice constants and band gaps respectively, to facilitate easier reading.

Just as in the InP convergence study, it is necessary to check that the k-point sampling density remains converged with the lower occupation function temperature of 300 K. Looking at this convergence check in Figure 4.4, it is seen that the k-point sampling convergence is unchanged by the change in the occupation function temperature, as was the case for InP. Thus the k-point sampling density of 6 Å remains sufficient.

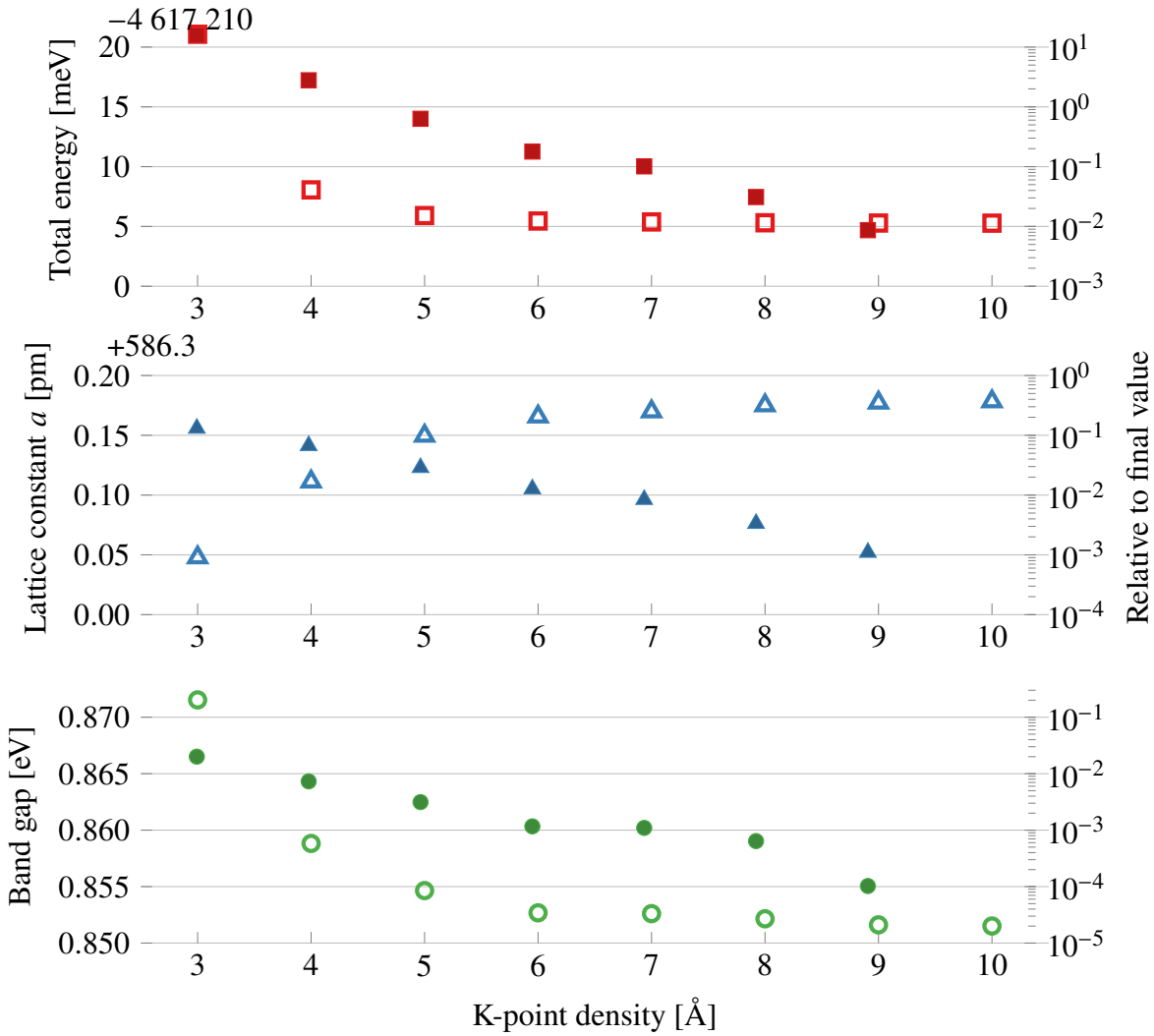


Figure 4.4: Convergence test of Brillouin zone k-point density sampling in $In_{0.53}Ga_{0.47}As$ simulated with the Virtual Crystal Approximation (VCA). Occupation function temperature was 300 K. Filled symbols indicate values taken relative to the final most accurate value, whereas empty symbols represent the absolute values. The absolute values were shifted by −4 617 210 meV and 586.3 pm for the total energies and lattice constants respectively, to facilitate easier reading.

The last parameter to converge is again the density-mesh cutoff, whose convergence check is shown in Figure 4.5. Using the same convergence criteria, of the band gap being converged within 1 meV of the final value, it is seen that the value of 105 Ha used thus far, is exactly sufficiently converged.

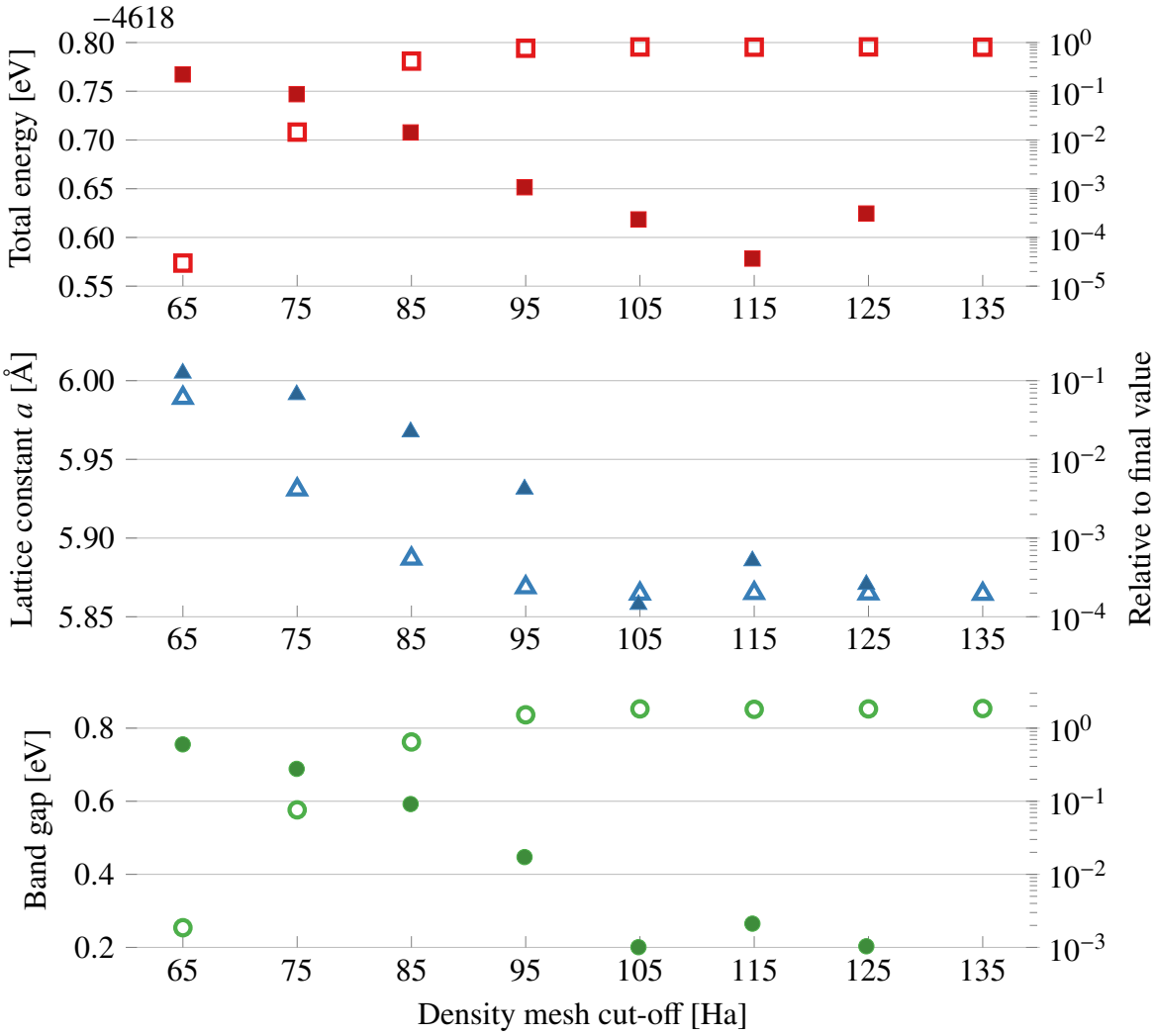


Figure 4.5: Convergence test of density-mesh cutoff in $In_{0.53}Ga_{0.47}As$ simulated with the Virtual Crystal Approximation (VCA). Filled symbols indicate values taken relative to the final most accurate value, whereas empty symbols represent the absolute values. The absolute values of the total energies were shifted by −4618 eV to facilitate easier reading.

Thus far, the accuracy parameters have been converged for a specific combination of exchange-correlation functional and basis set. This combination was chosen, because it was assumed that it offers the best trade-off between computational speed and the accuracy needed for the studies of this chapter, based on previous experiences and the previous convergence study on InP. Alternative combinations include some of those studied in the convergence study of InP in Section 3.2.2: the "High" and "Medium" PseudoDojo basis sets, and the GGA-$^1/_2$ and TB09 MGGA functional. Due to the larger demand for accurate band gaps in the studies of the present chapter, the use of smaller Auxiliary Density Matrix Method (ADMM) [128] basis sets for the exchange part of the HSE06 functional is also tested here. The comparison of the lattice constant, band gap and speed, of these combinations, is illustrated in Figure 4.6, wherein the accuracy parameters converged for the HSE functional were used for all combinations. As in Section 3.2.2, the GGA-$^1/_2$ calculation is relaxed with regular GGA and the TB09 MGGA calculation with the SCAN functional. Considering the lattice constants and band gaps of the GGA and MGGA functionals, it is clear that they are not sufficiently accurate, in terms of replicating the experimental values. The

GGA-$^1/_2$ functional, with both the "Medium" and "High" version of the PseudoDojo basis sets, drastically underestimates the band gap while overestimating the lattice constant. The MGGA functionals considerably improves the lattice constant, although still not sufficiently so, but worsens the band gaps. The drastically overestimated band gaps could be adjusted manually, by tuning the c-parameter around the self-consistently determined value of 1.2, and fixing it for the calculations. However, this could pose problems for the transferability and was therefore not done. All the tested combinations, using the hybrid HSE06 functional, yielded relatively good lattice constants and band gaps. Comparing the "Medium" with the "High" version of the basis set, the "High" version unsurprisingly yields an improved band gap. However, the computation speed is heavily reduced, since the calculation of the exact exchange dominates these calculations with small unit cells and since the exact exchange scales as $\mathcal{O}(N^6)$ with unique basic functions. For the same reason, the usage of ADMM basis sets speed up the calculations by three to five times, for the "Medium" and "High" basis sets respectively, with only minor reductions to accuracy. Unfortunately, the combination of VCA with ADMM basis sets were unsupported in QuantumATK version U-2022.12 for more complicated calculations, and was therefore not used for the remaining calculations in this thesis. Instead, the combination of HSE06 with the "Medium" basis set was used, as it was deemed to offer the best trade-off in terms of accuracy versus computation time.

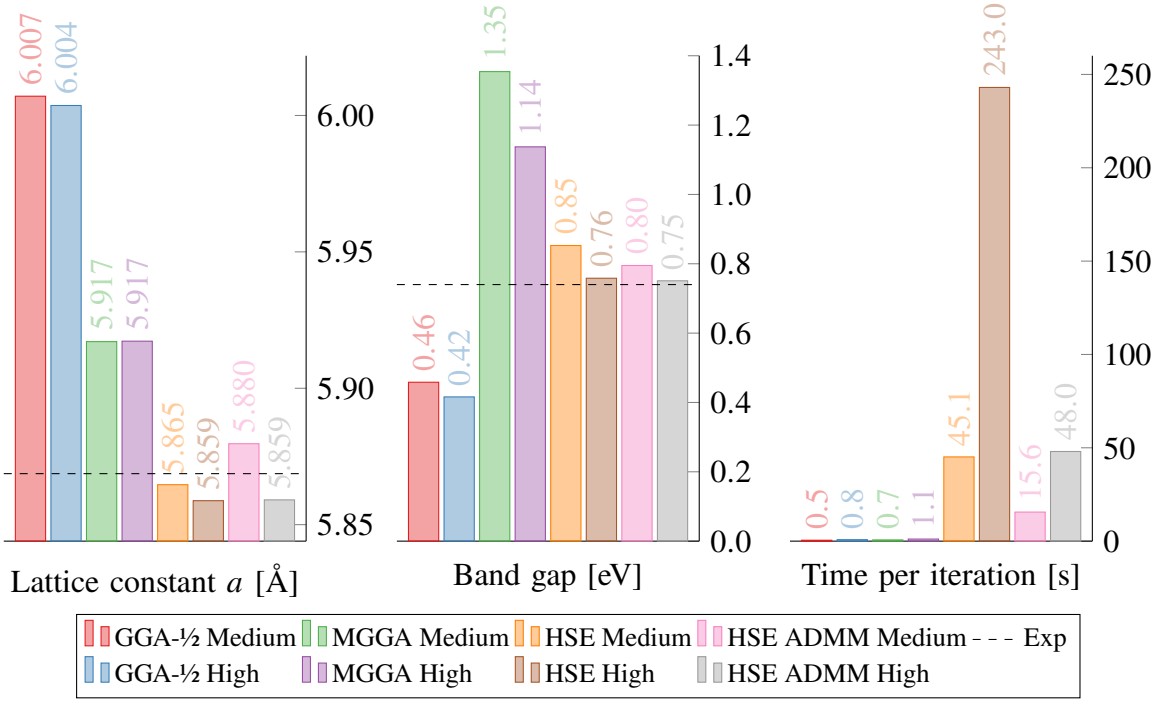


Figure 4.6: Comparison of combinations of exchange-correlation functionals and basis sets to simulate $In_{0.53}Ga_{0.47}As$ using the Virtual Crystal Approximation (VCA). For the GGA-$^1/_2$ functional, the relaxations were performed with regular GGA. For the MGGA functional, relaxations were performed with SCAN while band gaps were calculated with TB09. Experimental values are taken from the materials database Ioffe [102].

4.2.2 $In_{0.53}Ga_{0.47}As$ properties with VCA

Having converged the accuracy parameters and tested the accuracy of the used exchange-correlation functional and basis set, the usability of the VCA can now be investigated. One major advantage of using VCA, is the simplicity in studying various alloy compositions. In a truly random alloy simulation, the finite number of atoms in the supercell limits the range of available alloy compositions. With VCA, this limit is circumvented, since any alloy composition can be simulated by simply changing the mixture rate in the virtual atom. In non-atomistic simulation methods, such as finite-size methods, alloys like $In_xGa_{x-1}As$, with variable alloy fractions x, are simulated with properties interpolated between its two binary compounds, here InAs and GaAs. Often the interpolation has to be made more accurate, by going beyond the linear interpolation, using properties from alloy fractions such as $In_{0.53}Ga_{0.47}As$. To test if the VCA replicates this non-linearity of lattice constants and band gaps, these are calculated for indium fractions ranging from 0 to 1, using both the "Medium" and "High" basis sets. The results are illustrated in Figure 4.7, compared to the experimental interpolations from the material database Ioffe [102].

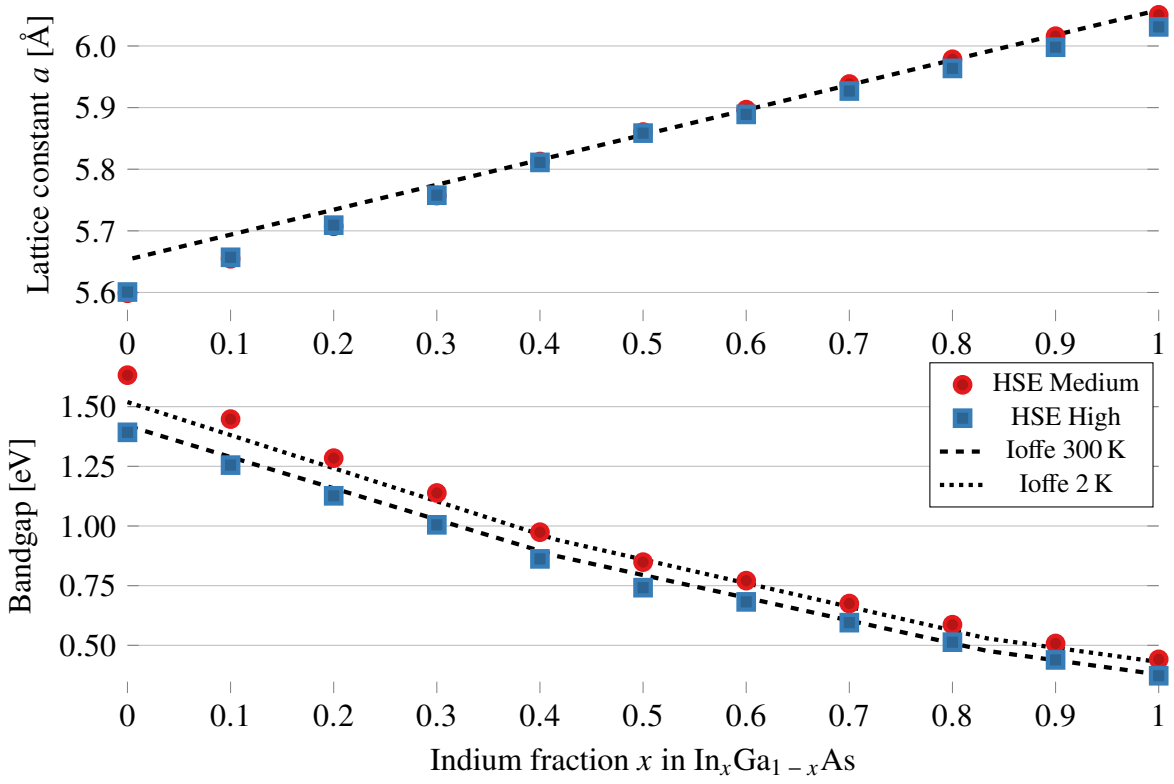


Figure 4.7: VCA simulations of $In_xGa_{1-x}As$ for indium fractions x ranging from 0 to 1. Experimental values are taken from Ioffe [102].

For the lattice constant, both the "Medium" and "High" basis sets replicate the quadratic interpolation fairly well in the intermediate region, but underestimates it slightly when the indium fraction is low (approximately 0.05 Å). For the band gaps, both basis sets follow the interpolation almost exactly, with the "Medium" basis set slightly overestimating the band gaps at low indium fractions, by approximately 0.1 eV for GaAs, and the "High" basis set consistently underestimating the band gaps, by 0.06 eV for InAs and 0.12 eV for GaAs. Since

DFT is in principle a 0 K method, the fact that the "High" basis set so consistently aligns with the 300 K band gaps, is a coincidence rather than a sign of accuracy. The fact that the "Medium" basis set appears to be more accurate than the "High" basis set could be a coincidence, but it could also be a sign of overcompleteness of the "High" VCA basis sets. In either case, the VCA has been shown to be capable at accurately simulate the entire alloy composition range of $In_xGa_{1-x}As$.

As a final investigation of the bulk $In_{0.53}Ga_{0.47}As$ VCA system, the Brilluin zone bandstructure, and the DOS projected on the orbitals of the elements, are calculated and shown in Figure 4.8. Since spin-orbit coupling was not included in the calculations, or spin-polarisation in general, the split-off valence band is degenerate with the heavy- and light-hole bands. Except for this issue, the bandstructure appears sensible and calculating the effective mass of electrons, by fitting a parabola to the conduction band, even yields the correct effective mass of $m_e^* = 0.041\, m_e$ [37].

Looking at the PDOS above the Fermi level, it is seen that the conduction band is dominated by the $In_{0.53}Ga_{0.47}$ virtual atom's p-orbitals, shown in orange, with strong contributions from its s-orbital and the arsenic's p-orbitals, shown in purple and blue respectively. In the valence band the heavy-hole and split-off bands are seen to be dominated by the p-orbitals, whereas the light-hole valence band is dominated by the $In_{0.53}Ga_{0.47}$ s-orbital. However, close to the band edges the situation is quite different, the conduction band is here mostly dominated by the $In_{0.53}Ga_{0.47}$ s-orbital, whereas the valence bands are dominated only by the As p-orbitals.

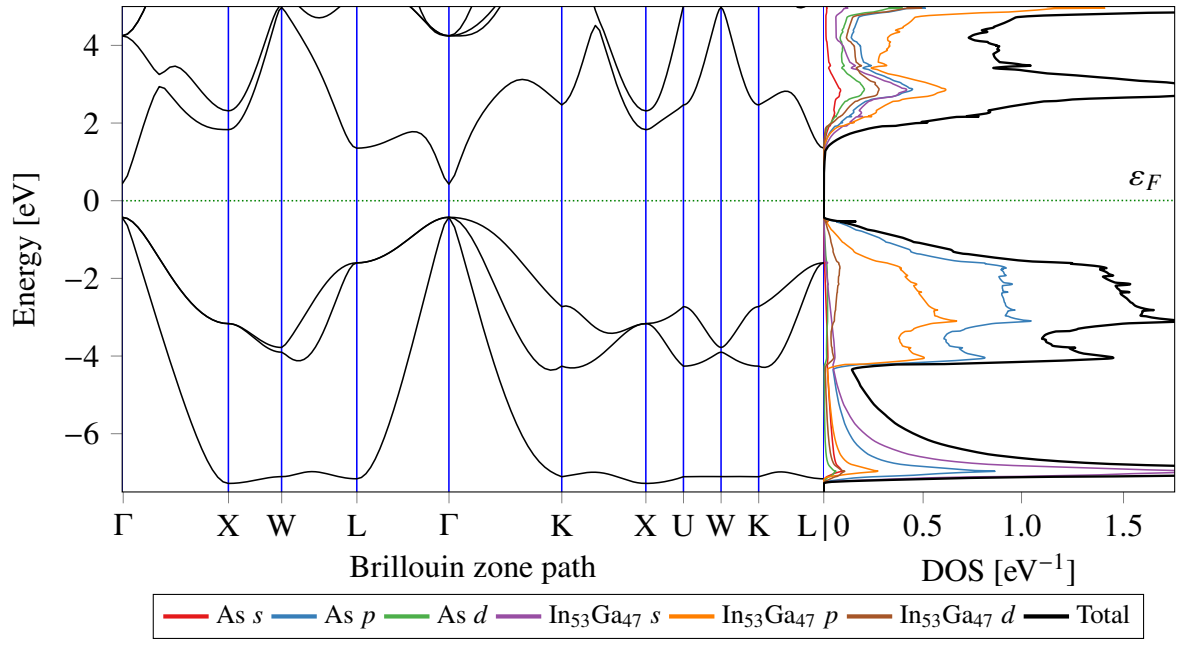


Figure 4.8: Brillouin zone bandstructure of $In_{0.53}Ga_{0.47}As$ simulated with the Virtual Crystal Approximation (VCA), together with the associated Density Of States (DOS) projected on the orbitals of the elements.

Having evaluated the VCA to simulate $In_{0.53}Ga_{0.47}As$, the alternative, which is using Special Quasi-random Structures (SQSs), is studied in the next section.

4.3 Special quasi-random structures

Special Quasi-random Structures (SQSs) are structures generated to capture the true randomness of the real physical alloy. They mimick the statistical ensemble of many random configurations, by yielding the same properties as the average of the ensemble. To investigate the necessity of modelling the entire statistical ensemble of random alloys, effective bandstructures can be calculated and examined.

4.3.1 Effective bandstructure of random alloy $In_{0.53}Ga_{0.47}As$

An effective bandstructure is a bandstructure of a repeated supercell, unfolded back into the simple unit cell. To unfold the bandstructure, each eigenstate is given a weight according to $\left|\langle e^{i\mathbf{k}\cdot\mathbf{r}}, \psi_{j,K}\rangle\right|$[129–131], where $|\psi_{j,K}\rangle$ is the eigenstate of band j in the supercell at k-point K. If the supercell is simply a repetition of the simple unit cell, Bloch's theorem holds which means that each weight becomes 0 or 1 and the bandstructure of the simple unit cell is recovered. If the supercell is a random alloy such as $In_{0.53}Ga_{0.47}As$, each simple unit cell in the supercell is different and the weights can be anywhere in between 0 and 1. This results in finite widths of the unfolded bands in the effective bandstructure, where the widths are a measure of the importance of the crystal disorder and thereby the necessity of modelling the random alloy ensemble.

To calculate effective bandstructures of $In_{0.53}Ga_{0.47}As$, 20 random configurations were prepared by repeating the simple 2-atom cubic unit cell 5x5x5 times, with an experimental lattice constant of $a = 5.8687$ Å. For the DFT parameters, it was assumed that $In_{0.53}Ga_{0.47}As$ would behave similarly, whether simulated with VCA or as a random alloy supercell. Therefore, the converged accuracy parameters of Section 4.2 were used for these calculations, but since ADMM was only unsupported with VCA, and since it yielded huge speed-ups with only minimal loss of accuracy (see Figure 4.6), ADMM was employed in the calculations of this section. The mean of the effective bandstructures across the 20 random samples, together with the related DOS, projected onto the orbitals of the elements, is shown in Figure 4.9. It is seen that the conduction band is narrow around the Γ-point, which means that it is well-defined and unaffected by the random crystal disorder effects. Conversely, along the path to the L-point, the conduction band has a width of approximately 0.2 eV, which means that for transport in this region, crystal disorder effects are highly important and could be the dominant current-limiting scattering mechanism. It is also seen that the upper valence bands have a well-defined width of approximately 0.08 eV in the entire Brillouin zone. While smaller than the width of the conduction band at the Γ-L path, this width is still significant and cannot be neglected, especially since it even occurs at the Γ-point. It highlights the necessity of including crystal disorder effects when simulating $In_{0.53}Ga_{0.47}As$, even for properties which only depend on the Γ-point.

Comparing the DOS in Figure 4.9 with the DOS calculated from VCA in Figure 4.8, yields further insight into the physics behind the bands and of VCA. With VCA the DOS contributions from the group III elements were compounded in the $In_{0.53}Ga_{0.47}$ virtual atom,

in the random alloy it is split into the separate Ga and In contributions, revealing the difference between the group III-elements impact on the random alloy. It is seen that in the valence bands, indium and gallium have near identical DOSs, but in the conduction band it is revealed that the large contribution from $In_{0.53}Ga_{0.47}$ *s*-orbitals in VCA, actually stems mostly from the Ga *s*-orbitals. Additionally, with VCA the valence bands had a small contribution from the $In_{0.53}Ga_{0.47}$ *d*-orbitals, which is not the case in the random alloy, and the contribution from the As *d*-orbital to the conduction band, is also much more significant in the random alloy than the VCA. At the valence bad edge the DOS is also dominated solely by the As *p*-orbitals in the random alloy, but at the conduction band edge the As *p*- and *s*-orbitals show a similar contribution as the *s*-orbitals of Ga and In, in contrast to the situation in the VCA.

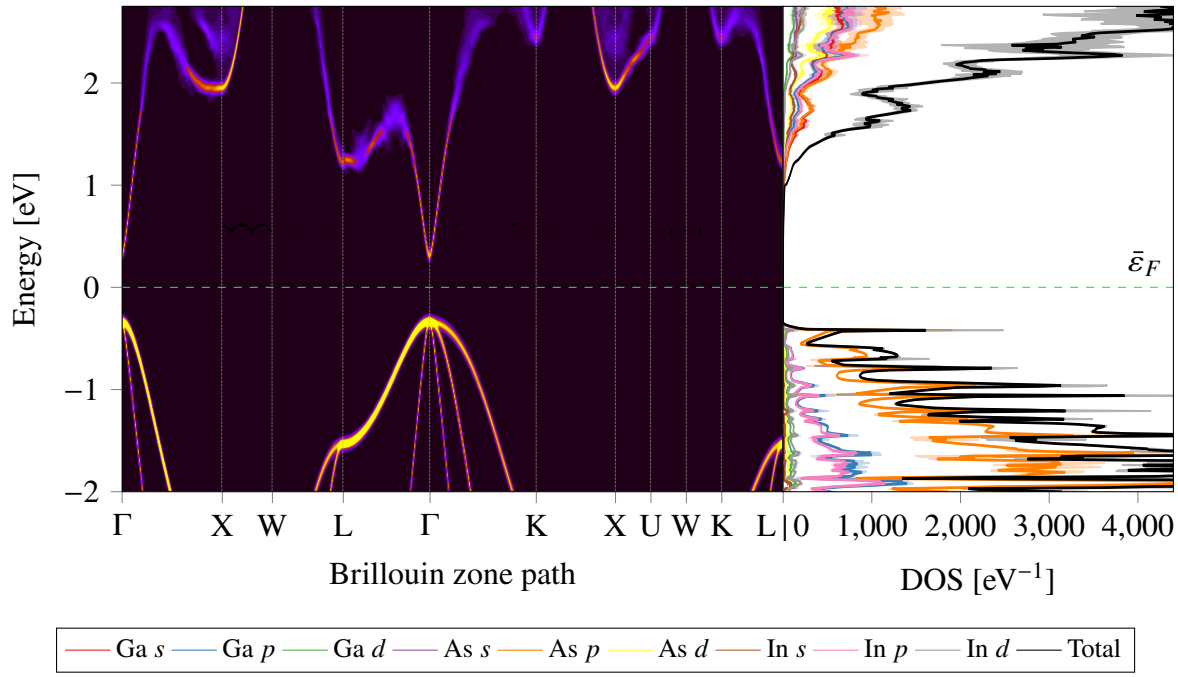


Figure 4.9: Effective bandstructure and associated Density Of States (DOS), projected onto the atomic orbitals, of 20 samples of the statistical ensemble of random alloy $In_{0.53}Ga_{0.47}As$ structures. The configurations were constructed from unrelaxed 5x5x5 repetitions of the simple unit cell, using an experimental lattice constant of $a = 5.8687$ Å. The illustrated bandstructure, DOS and Fermi level is the average across the 20 samples, with the shaded width of the DOS representing the standard deviation.

4.3.2 Special quasi-random structure generation

For the calculations in Figure 4.9, the statistical ensemble of $In_{0.53}Ga_{0.47}As$ random alloys was sampled 20 times and the results were averaged across the samples. In order to make simulations of the ensemble more tractable, Special Quasi-Random Structures (SQSs) are used instead of the group of random ensemble samples. In essence, SQSs are simply good representative structures of the ensemble that they mimick. In order to find these good representatives, a measure of the SQS's quality is imperative. For this measure, the objective

function in Eq. (3) of [132] is used

$$Q = -L + \sum_{\alpha} |\Delta\rho_{\alpha}(\boldsymbol{\sigma})| \tag{4.1}$$

where $\Delta\rho_{\alpha}(\boldsymbol{\sigma}) = \rho_{\alpha}(\boldsymbol{\sigma}) - \rho_{\alpha}(\boldsymbol{\sigma})^{\text{rnd}}$ is the difference between the correlations of cluster α of the SQS candidate and an ideal random structure and L is the size of the largest cluster for which $\Delta\rho_{\alpha}(\boldsymbol{\sigma}) = 0$. The correlation function of cluster α in the candidate SQS, is given by

$$\rho_{\alpha}(\boldsymbol{\sigma}) = \Big\langle \prod_{i} \gamma_{\alpha_i}(\sigma_i) \Big\rangle_{\alpha} \tag{4.2}$$

where $\boldsymbol{\sigma}$ is the vector of elemental occupations σ_i at sites i, $\gamma_{\alpha_i}(\sigma_i)$ is a cluster function which returns 1, if i is **not** included in the cluster, or ± 1 otherwise, depending on the elemental occupation σ_i. The cluster average $\langle \cdots \rangle_{\alpha}$ in Eq. (4.2), is taken over all symmetrically identical clusters (or interactions), which could for example be all the nearest-neighbour clusters. Since the majority of material properties depends mostly on the local environment, matching the smaller clusters (such as nearest-neighbours) is far more important than the larger clusters. To facilitate this, the $-L$ is included in the objective function in Eq. (4.1). If the size of a cluster is determined by the largest distance between two atoms in it, then L is the size of the largest cluster for which $\Delta\rho_{\alpha}(\boldsymbol{\sigma}) = 0$. Finally, the correlations of cluster α in the ideal random structure $\rho_{\alpha}(\boldsymbol{\sigma})^{\text{rnd}}$, can for a binary SQS such as $In_xGa_{1-x}As$ be determined as $(2x-1)^{\alpha_n}$, where α_n is the number of atoms in the cluster.

Having defined a measure of the SQS's quality, the objective function in Eq. (4.1), the best SQS possible can be found by minimising this objective function. This is done by using an evolutionary algorithm, wherein three repeating steps are performed n_{gen} times, on an initial population of n_{pop} randomly generated structures:

1. The objective function, Eq. (4.1), is calculated for each structure in the population.

2. A new population of $n_{\text{pop}} - n_{\text{pro}}$ SQS candidates are generated randomly, by using permutation and hereditary operators on the current population, favouring the structures in the population with lower objective functions.

3. The n_{pro} configurations with the lowest objective function in the current population are added to the new population, the rest are discarded.

Having evolved the initial population using these three steps n_{gen} times, the resulting configuration with the lowest objective function is then the best available SQS. This evolutionary algorithm is illustrated schematically in Figure 4.10.

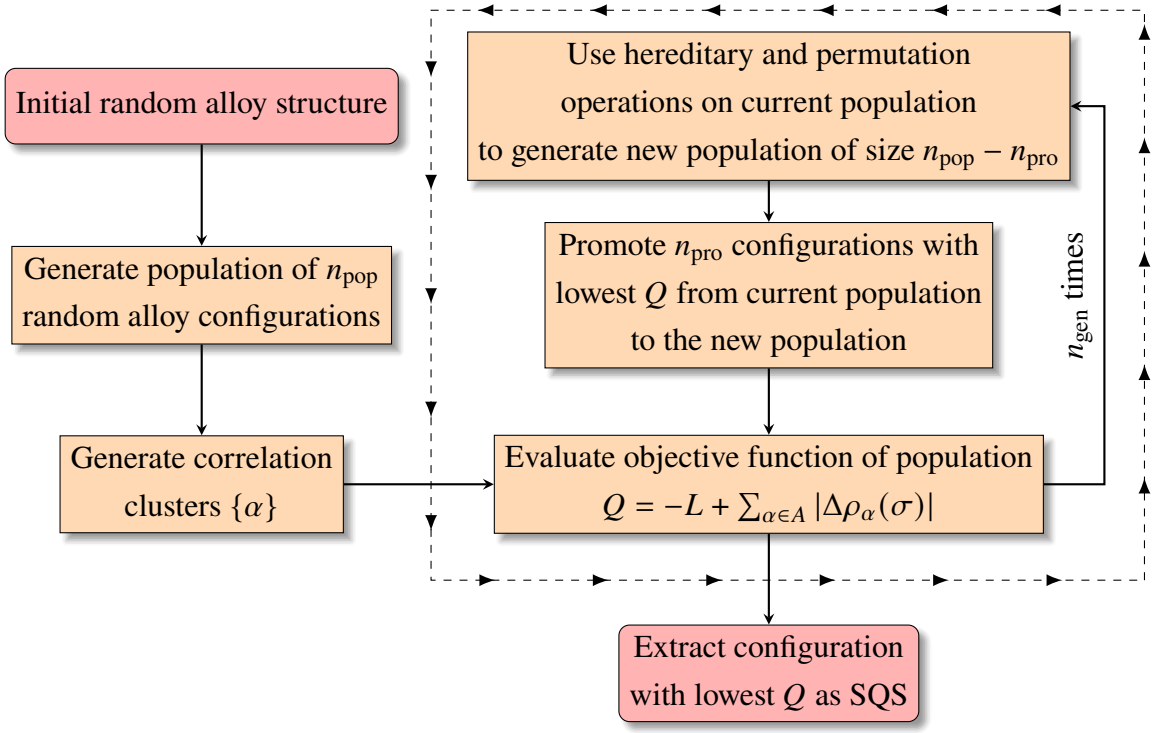


Figure 4.10: Flowchart of the evolutionary algorithm used to find the best available SQS by minimising the objective function Q.

When the size of the SQS increases, the correlation clusters can be better matched to the ideal random alloy, due to the increased degree of freedom in the alloy sublattice sites. Therefore, better SQSs can be achieved with larger structures, but since larger structures are more computationally demanding, it is necessary to compromise to keep simulations feasible. To find the best compromise between size and quality of the SQS, the evolutionary algorithm illustrated in Figure 4.10, is used to generate $In_{0.53}Ga_{0.47}As$ SQSs of increasing sizes, using three different set of algorithm parameters:

$$\begin{aligned} &\text{Set 1:} && n_{pop} = 100, n_{pro} = 10, n_{gen} = 200 \\ &\text{Set 2:} && n_{pop} = 200, n_{pro} = 20, n_{gen} = 400 \\ &\text{Set 3:} && n_{pop} = 300, n_{pro} = 30, n_{gen} = 600 \end{aligned} \tag{4.3}$$

These parameters are proportioned to consistently increase the accuracy of the evolutionary algorithm. The total correlation function error (i.e. Eq. (4.1) without the $-L$) of $In_{0.53}Ga_{0.47}As$ SQSs, with sizes corresponding to 3x3x3 (54 atoms) to 6x6x6 (432 atoms) repetitions of the minimal 2-atom cubic unit cell, is shown in Figure 4.11. A large improvement of the SQSs is seen when increasing the supercell size from 3x3x3 to 4x4x4. While the SQSs are still improving when the size is further increased, the improvements are far less extreme. This suggests that a good compromise between the SQSs quality and their size, is achieved at this supercell size. However, at this 4x4x4 supercell size, the algorithm is struggling to find the best SQS, as seen by the large variation between the resulting SQSs from the different evolutionary algorithm runs. Considering that the algorithm run with the first set of parameters managed to produce the best SQS, it is likely that this SQS is an extreme

outlier, which was happened upon purely by coincidence in this algorithm run. This issue is not observed with the larger structures, where the SQSs are more similar and the third set of algorithm parameters yield better SQSs as expected.

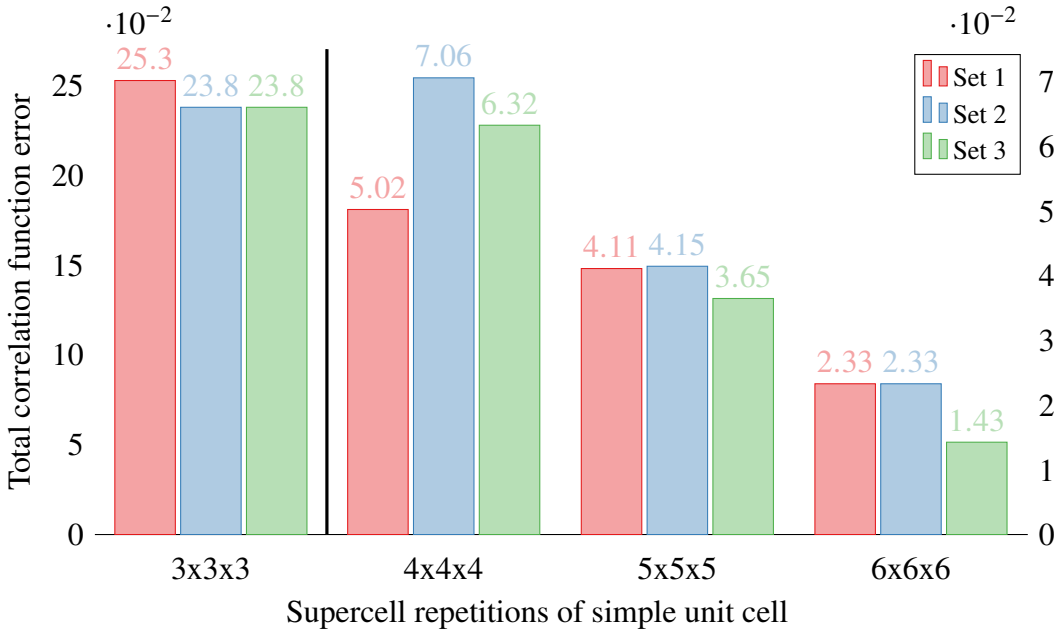


Figure 4.11: Quality of $In_{0.53}Ga_{0.47}As$ Special Quasi-random Structures (SQSs) of various supercell sizes and evolutionary algorithm parameters, as defined by the total correlation function error (lower is better). The three sets of evolutionary algorithm parameters are defined in Eq. (4.3).

Considering solely the total correlation function error of the SQSs, it appears that the SQS from the "Low" evolutionary algorithm run with a 4x4x4 supercell size, manages to adequately capture the randomness of the true $In_{0.53}Ga_{0.47}As$ random alloy.

4.3.3 $In_{0.53}Ga_{0.47}As$ properties with SQS

To further test these SQSs' capability to mimic the random alloy ensemble, the band gap of each SQS is calculated and shown in Figure 4.12. To calculate the band gaps, the SQSs geometry were first optimised, using the same DFT parameters as for the effective bandstructures, and the same relaxation parameters as for the VCA calculations in Section 4.2. It is seen that, despite the SQS from the algorithm run with the first set of parameters having a much lower total correlation function error than the other 4x4x4 supercell SQSs, it has a similarly bad band gap. These band gaps are however greatly improved, when the sizes of the SQSs are increased from 4x4x4 to 5x5x5 supercells. Only minimal further improvement is seen, when their size is further increased, suggesting that this supercell size offers the best trade-off between SQS size and accuracy. Taking only the band gaps into account, it also appears that the choice of the different algorithm parameters sets, have no impact on the accuracy of the results. Considering the (for most) relatively low variance of the total correlation error functions in Figure 4.11, this is likely due to the first set of parameters being sufficiently accurate as it is, for the SQSs that are being generated in this thesis.

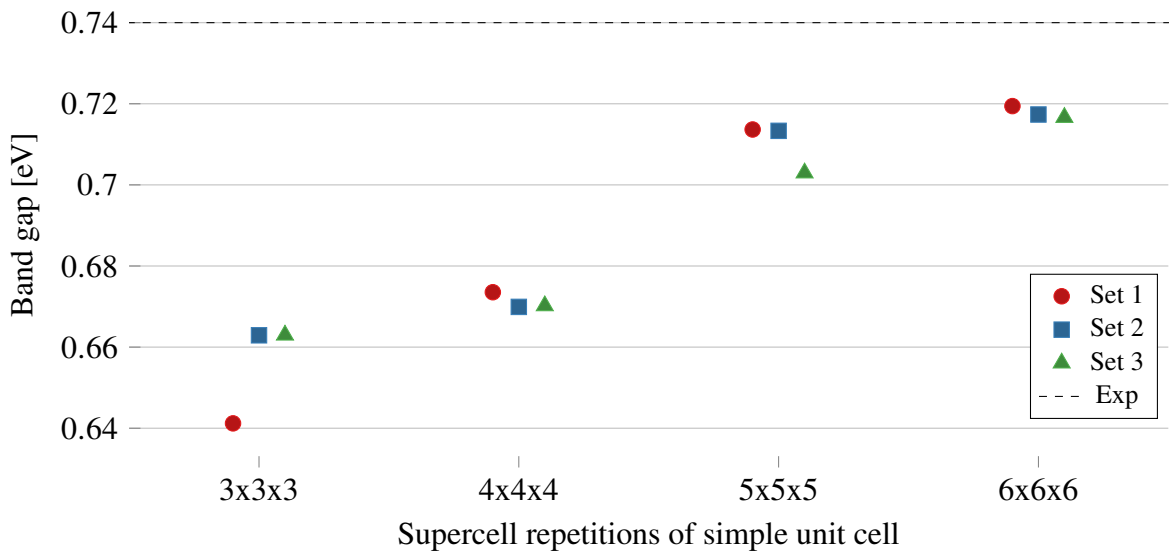


Figure 4.12: Band gaps of relaxed $In_{0.53}Ga_{0.47}As$ Special Quasi-random Structures (SQS) of various supercell sizes and evolutionary algorithm parameters. The three sets of evolutionary algorithm parameters are defined in Eq. (4.3).

So far, the DFT parameters used have been based on the parameter sweep for VCA, shown in Figure 4.6 in Section 4.2. To confirm the accuracy of these parameters for SQS simulations, the combinations of basis set and exchange-correlation functional are tested again on a selected SQS candidate. The chosen SQS is the one resulting from the evolutionary algorithm run with the first set of parameters, with a supercell size of 5x5x5 simple unit cells, as this SQS was converged with respect to size. The comparison of lattice constants, band gaps and computation time between the different combinations is shown in Figure 4.13. The difference between the combinations of basis sets and exchange-correlation functionals, in terms of band gaps and lattice constants, are overall similar to those for the VCA, seen in Figure 4.6. However, the lattice constants with GGA and MGGA are improved in the SQS, whereas for HSE they are worsened to such a degree, that the MGGA functional actually yields better results than the HSE functional, for the "High" PseudoDojo basis set. One large difference between the tests with SQS and the VCA, is the computational speed of each basis set and exchange-correlation functional pair. With VCA the system only contained 2 atoms, in contrast to the 250 atoms of the SQS, and due to the implementation of the VCA, the amount of basis functions were doubled for the alloy species. The HSE functional has a large overhead in computation time and scales incredibly poorly with unique basis functions. Due to this, the HSE functional was significantly slower than the other functionals with the VCA, whereas it is only moderately slower for the SQS. Considering the trade-off between computational speed and accuracy, the "Medium" version of the PseudoDojo basis set with the HSE functional, also appears to be the best option for the SQSs. For the SQSs the ADMM approximation can be used, in contrast to the VCA, for a slight speed-up without loss of accuracy. While the MGGA functionals are faster than HSE with ADMM, and yields similarly good lattice constants, in order to get a good description of the band gaps with MGGA, it is necessary to manually tune the TB09 c-parameter, which can potentially cause a problem with the transferability.

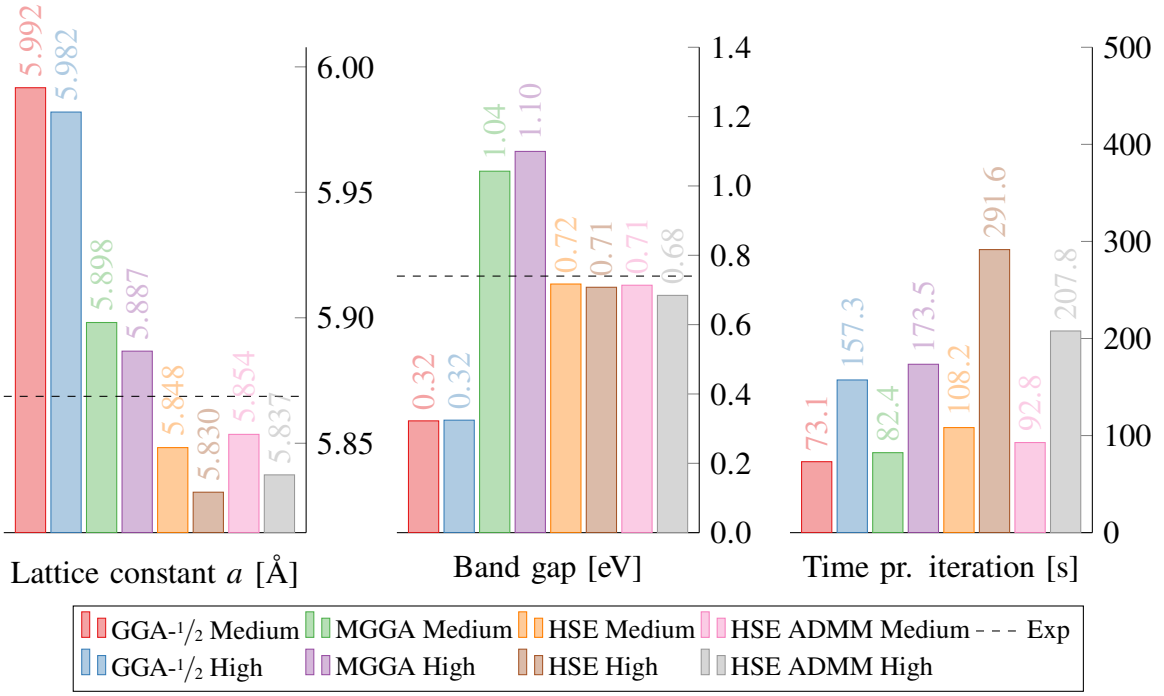


Figure 4.13: Comparison of combinations of exchange-correlation functionals and basis sets to simulate $In_{0.53}Ga_{0.47}As$ using Special Quasi-random Structures (SQSs). For the GGA-$^1/_2$ functional, the relaxations were performed with regular GGA. For the MGGA functional, relaxations were performed with SCAN while band gaps were calculated with TB09. Experimental values are taken from the materials database Ioffe [102].

Having generated a good SQS of $In_{0.53}Ga_{0.47}As$ and tested its convergence with size and choice of DFT parameters, in the next section this SQS will be compared to the VCA, in their ability to simulate point defects. The method which demonstrates the best results, will then be used to simulate an array of experimentally relevant point defects.

4.4 Charged point defects

So far, the two different methods to simulate the random alloy $In_{0.53}Ga_{0.47}As$, the VCA and SQS, have been tested only on pristine bulk structures. Since the simulations to be done in this chapter are simulations of point defects, which depends strongly on the local environment of the defect, it is required to also test the methods on such a calculation. But before this can be done, a description of the theoretical background is obligatory.

4.4.1 Theory

Point defects can be categorised as either intrinsic or extrinsic. Intrinsic point defects are defects which consists solely of elements native to the crystal lattice. Conversely, extrinsic point defects are defects which incorporates foreign elements into the lattice, such as dopants or contaminants. There exists three basic types of point defects, which all more complicated defects can be derived from. The first of these are vacancies. Vacancies are always intrinsic, since they consist of the removal of an atom from the lattice. The second basic type of point defects are interstitials. Interstitials are atoms added to positions in-between lattice sites, as such they can be both intrinsic and extrinsic. The last basic type of point defects are substitutions, which are when an element at a lattice site has been replaced with another element. Substitutions can also be both intrinsic and extrinsic, such as dopants. Intrinsic substitutions are more uncommon and are usually observed in polar materials such as III-Vs, where a group III atom can replace a group V atom, such a substitution is also called an antisite. More complicated defects, so-called defect clusters, are then made up of several of these basic types of defects. Examples of this are Schottky defects, which consists of two or more charged vacancies with zero net charge, or Frenkel defects, which are when an atom moves from a lattice site into an interstitial position, i.e. it consists of a vacancy and an interstitial of the same element. In this thesis, the investigation of charged point defects will be limited to four kinds of defects: substitutions, vacancies, double antisites and Schottky defects. A double antisite is a defect cluster consisting of two neighbouring antisites of opposite elements, i.e. it is simply two neighbouring III-V elements swapped around. These four different types of point defects are illustrated in Figure 4.14, embedded in an $In_{0.53}Ga_{0.47}As$ SQS supercell.

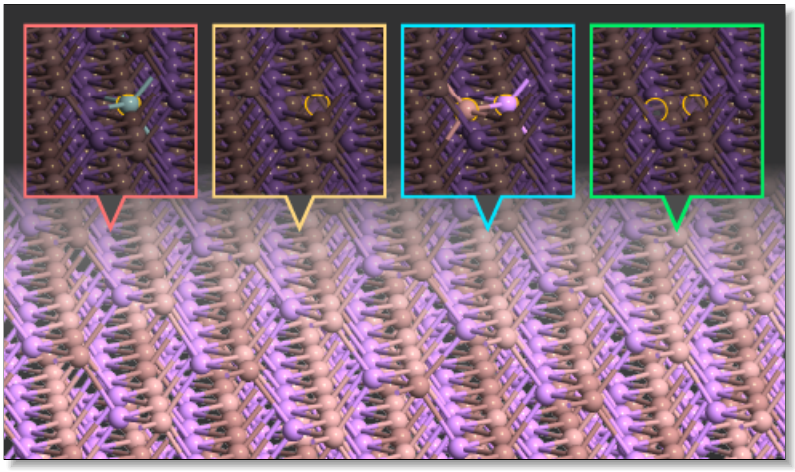


Figure 4.14: The four kinds of point defects studied in this thesis, embedded in an $In_{0.53}Ga_{0.47}As$ SQS. From the left: substitutions, vacancies, double antisites and Schottky defects.

Common for all types of defects, is that it is in general unknown beforehand, which charge state of the defect is the ground state, also called the Stable Charge State (SCS). It is however possible to make a qualified guess, based on the valency of the atoms constituting the defect or its local environment. For example, it is likely, that a zinc substitution of an indium atom in pristine $In_{0.53}Ga_{0.47}As$ is negatively charged, since zinc has one less valency electron than indium and therefore functions as a p-dopant, i.e. it catches an electron from (releases a hole to) the surrounding lattice. If the $In_{0.53}Ga_{0.47}As$ is already heavily p-doped, or positively biased, however, the zinc atom can potentially be unable to catch an electron (release a hole), since the surrounding lattice is already void (full) of electrons (holes). Evidently, for all point defects, there exists a chemical potential for which the charge of its ground state, the SCS, changes. This chemical potential is called the transition level, since it is the level at which the defect undergoes a transition from one charge state to another. Transition levels which lies in the band gap of the host material are important in physical devices, since excited carriers can decay to this transition level. The carriers which have been "caught" by a transition level, no longer conducts current, because the charge of the defect state is localised around the defect. For this reason, transition levels which lies in the band gap of the host material, are sometimes called "trap" levels. Both the transition levels and the charge of the ground state are important properties of a point defect, another important related property is the formation energy of the defect. The formation energy is the energy associated with the formation of the defect from its constituents, to clarify further, it is the energy required to form the defect, disregarding any intermediary energy barriers in the formation process. Formally, the formation energy is defined as

$$E_f^q = E^{\mathrm{def},q} - E^{\mathrm{bulk},q} - \sum_i \Delta n_i \mu_i + q(\mu_e^{\mathrm{ref}} + \Delta\mu_e) \tag{4.4}$$

where $E^{\mathrm{def},q}$ is the total energy of the defect system in the charge-state q, $E^{\mathrm{bulk},q}$ is the total energy of the bulk system without the defect, Δn_i is the difference in number of atomic species i in the defect and bulk systems, μ_i is the atomic chemical potential of the species, μ_e^{ref} is the chemical potential reference level, which will be taken to be the valence band edge and $\Delta\mu_e$

is the electronic chemical potential relative to it. For realistic calculations of E_f^q, the atomic chemical potentials μ_i should be calculated from the precursor molecules (or other origins) where the atomic species i originates from. This is however very difficult, as the origin of an atomic species is not always known, and even when it is known, it usually contains many different species, such as in the case of a precursor molecule, making it hard to untangle the atomic chemical potentials of each separate species. To avoid this issue, and to make comparison of the formation energies of different defects more straightforward, the atomic chemical potentials are calculated from elemental configurations.

Once the formation energies of various charge states in a defect have been calculated, finding the transition levels of the defect is straightforward. In Figure 4.15, the formation energies of charge states −1 through +3 of a defect in $In_{0.53}Ga_{0.47}As$, are plotted for electronic chemical potentials $\Delta\mu_e$, ranging from slightly below the valence band maximum, to slightly above the conduction band minimum. In this energy range, two transition levels are observed, as the +2 charge ground state crosses the +1 state at $\Delta\mu_e = 0.45\,\text{eV}$, which then itself crosses the neutral state at $\Delta\mu_e = 0.66\,\text{eV}$. The defect whose formation energies are shown here, is the $As_{In/Ga}$ antisite. The point defect notation A_B describes an atomic species A replacing an atomic species B. For a vacancy, "V" is used for the replacing atomic species A.

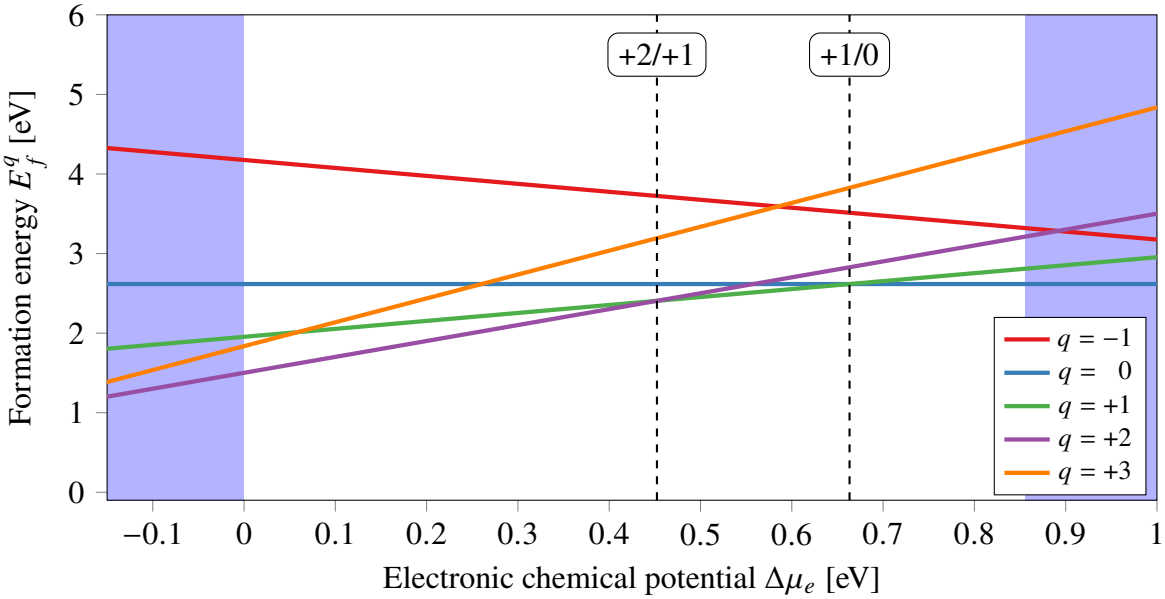


Figure 4.15: Formation energies of the $As_{In/Ga}$ antisite defects charge states in a 5x5x5 supercell of $In_{0.53}Ga_{0.47}As$ simulated with the VCA.

Calculating the formation energies of charge states using Eq. (4.4) in DFT, has one major issue. When simulating defects in materials it is often the dilute limit which is of interest, which means that it is assumed that the defects are so far apart, that they do not interact and can be considered isolated. In DFT, Bloch's theorem is used to limit the simulation scope to a small unit cell, which is then simply repeated periodically to create the full system. If a defect is embedded in this small unit cell, it will then also be repeated and it will interact with its repeated images. To correct for this spurious interaction, one can use large supercells instead of minimal unit cells for the simulation domain. However, for the charged states,

the electrostatic interaction is far-reaching and would require prohibitively large supercells to dampen out. Instead of using large supercells to minimise the interactions, another possibility is to correct for the interactions. To do that, Freysoldt, Neugebauer and Van de Walle came up with the FNV correction scheme[133–135], which corrects the formation energy by adding two terms to Eq. (4.4):

$$E_f^q = E^{\mathrm{def},q} + E_q^{\mathrm{latt}} - E^{\mathrm{bulk},q} - \sum_i \Delta n_i \mu_i + q(\mu_e^{\mathrm{ref}} - \Delta V_{q/0} + \Delta \mu_e) \tag{4.5}$$

The first of these corrections E_q^{latt} account for the interactions between the periodic images of the charge, by subtracting the difference between an isolated model charge and the same model charge repeated periodically as in the system. The second correction $\Delta V_{q/0}$ accounts for the shift of the electronic bands due to the presence of the defect, by using the potential difference between the pristine bulk and the defect system, evaluated far from the defect, where the periodic interaction is removed using the repeated model charge potential. Since the electrostatic interaction between the images of the defect system is screened by the system, screening is also accounted for in the model charge system, using a chosen static dielectric constant.

With the methodology of charged point defect simulations described, it is now possible to compare the VCA with SQSs, in their ability to accurately simulate embedded defects.

4.4.2 VCA versus SQS

To compare the VCA and SQS methods to simulate point defects in $In_{0.53}Ga_{0.47}As$, the As_{In} antisite defect is used, since this defect has earlier been proposed to be the most stable of the intrinsic defects in $In_{0.53}Ga_{0.47}As$[119]. One of the advantages with the VCA over SQSs, was that the VCA enables the use of minimal unit cells, thereby allowing for a straightforward up-scaling of the simulation cell size. As such, it is possible with the VCA to investigate the impact of finite size effects on formation energies and transition levels in the defect. To do this, VCA calculations were done for supercells consisting of 2x2x2 to 5x5x5 repetitions of the simple 2-atom unit cell, and the resulting formation energies and transition levels are shown in Figure 4.16. The value of the transition levels shown in Figure 4.16, the calculated transition levels which falls outside of the $In_{0.53}Ga_{0.47}As$ band gap, as well as the formation energy of the defect in the stable charge state, at the chemical potential of the intrinsic neutral bulk, are given in Table 4.1.

Looking at the formation energies, the neutral charge state of the defect seems to have converged already for the 3x3x3 supercell (54 atoms). The charged states however, barely looks converged for the 5x5x5 supercell (250 atoms), except for the +2 charge state which appears converged already with the 4x4x4 supercell (128 atoms). The reason why the +2 charge state converges more easily than the other charge states, may be due to it being the ground state at the zero chemical potential considered here. The fact that the differences between the corrected and uncorrected formation energies are not reduced for the supercell

sizes considered here, highlights the long range and slow decay of the screened electrostatic interaction of the charged point defect. Considering the magnitude of the electrostatic correction for the +2 and +3 charge states, it is evident that this correction is necessary, in order to achieve the correct transition levels involving these charge states.

Examining the transition levels, it is seen that of the four transition levels calculated in the As_{In} antisite defect, two of them lies in the $In_{0.53}Ga_{0.47}As$ band gap. Considering the fluctuations of these transition levels for the various supercell sizes, it becomes even more clear that the charged defect may still not be fully converged in the 5x5x5 supercell. The +2/+1 transition level shifts approximately 0.1 eV upwards in energy with each increase in the system size, not yet showing signs of convergence even in the 5x5x5 supercell. At this supercell size, the +2/+1 transition level of the defect has only just shifted above the Fermi level of the intrinsic neutral bulk $In_{0.53}Ga_{0.47}As$, thereby confirming that the +2 state is the stable charge state of the defect, rather than the +1 charge state, as predicted by the defect valency. In contrast to the +2/+1 transition, the +1/0 transition appears converged in the 5x5x5 supercell, as the downwards shift of the transition decreases to only approximately 0.03 eV, when the system size is increased to the 5x5x5 supercell. Despite the uncertainty in the convergence of the +2/+1 transition level at the 5x5x5 supercell size, calculations with larger supercells were not carried out, due to computational limitations and the increased size-scaling caused by the VCA implementation.

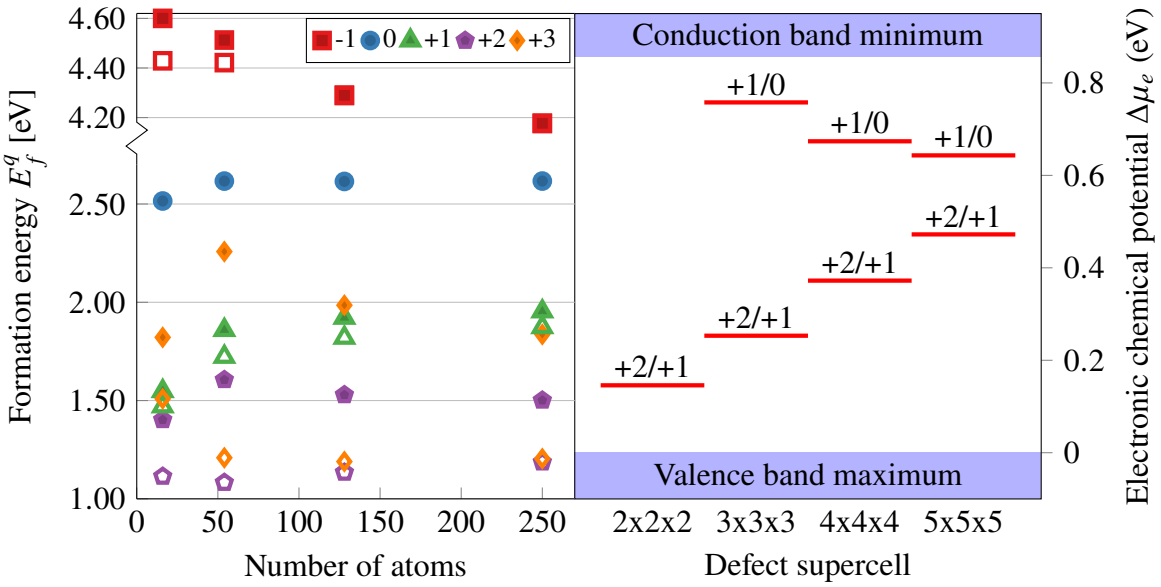


Figure 4.16: Finite size study of formation energies and transition levels in an As_{In} antisite defect embedded in $In_{0.53}Ga_{0.47}As$ simulated with the Virtual Crystal Approximation (VCA). Formation energies are calculated at $\Delta\mu_e = 0$, using Eq. (4.5) with the FNV correction (solid symbols) and using Eq. (4.4) without the correction (empty symbols). Transition levels are calculated using the FNV corrected formation energies.

Table 4.1: Full transition level data of study shown in Figure 4.16, including transition levels which lie outside of the band gap of the defect host material. The given formation energy, is the formation energy of the stable charge state at the chemical potential of the intrinsic neutral bulk, calculated with the FNV correction.

	Defect supercell			
Transition [eV]	2x2x2	3x3x3	4x4x4	5x5x5
+3/+2	-0.4195	-0.6523	-0.4563	-0.3351
+2/+1	0.1459	0.2531	0.3923	0.4523
+1/0	0.9674	0.7578	0.6937	0.6633
0/−1	2.0848	1.8961	1.6741	1.5591
q^{SCS} [e]	+1	+1	+1	+2
E_f^{SCS} [eV]	1.9761	2.2870	2.3493	2.3571

With the SQS method it is possible to increase the system size to a 6x6x6 supercell, to further investigate the finite-size convergence of the charged point defect study. However, due to the stochastic nature of the SQS, it is necessary to include a few different SQSs of each size in the study, to ascertain that the differences between various supercell sizes are indeed caused by the enlarged supercell and not merely random fluctuations. For this, the three 5x5x5 and 6x6x6 supercell SQSs generated in Section 4.3 are utilised and the transition levels of the As_{In} antisite defect embedded in these SQSs are shown in Figure 4.17. As for the VCA finite-size study, the numerical values of the transition levels are given in in Table 4.2, for all calculated transition levels, together with the formation energies of the stable charge state at the chemical potential of the intrinsic neutral bulk. All transition levels specified here are calculated with the FNV corrected formation energies, which will also be the case for all future calculations.

Whereas the band gap, a bulk property, was converged with 5x5x5 SQS supercells, it is clear from Figure 4.17 and Table 4.2, that transition levels of point defects require 6x6x6 SQS supercells in order to be properly converged. It is seen, that the +2/+1 and +1/0 transition levels vary substantially between the different 5-repetition SQS supercells, to the degree that even the stable charge state is different among the SQSs, whereas they exhibit barely any fluctuations in the 6-repetition supercells and the stable charge state is consistently predicted to be the +2 state. Thus, despite the average of the transition levels across the SQSs being converged with 5x5x5 supercells, using 6x6x6 SQS supercells seemingly allows for the use of a single SQS for calculations, rather than a small ensemble.

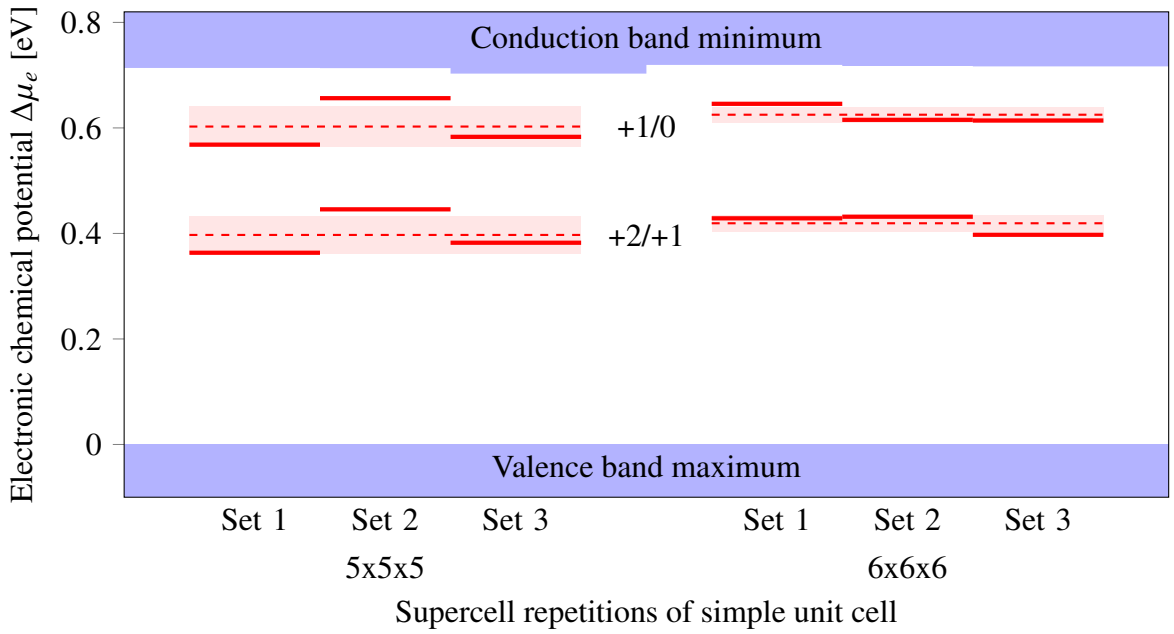


Figure 4.17: Transition levels of the As_{In} antisite defect in the SQS $In_{0.53}Ga_{0.47}As$ supercells consisting of 5x5x5 and 6x6x6 repetitions of the simple unit cell. The mean of the transition levels across the SQSs resulting from the three different set of SQS parameters, are given by the dashed line, and the standard deviation by the shaded area. The SQS parameter sets are given in Eq. (4.3) and the transition levels are calculated with the FNV corrected formation energies.

Table 4.2: Full transition level data of study shown in Figure 4.17, including transition levels which fall outside of the band gap of the defect host material. The given formation energy, is the formation energy of the stable charge state at the chemical potential of the intrinsic neutral bulk, calculated with the FNV correction.

[eV]	5x5x5 SQS supercell					6x6x6 SQS supercell				
Transition	Set 1	Set 2	Set 3	Mean	STD	Set 1	Set 2	Set 3	Mean	STD
+3/+2	-0.323	-0.335	-0.339	-0.332	0.007	-0.261	-0.249	-0.244	-0.251	0.007
+2/+1	0.363	0.446	0.382	0.397	0.035	0.429	0.432	0.398	0.419	0.012
+1/0	0.568	0.656	0.583	0.603	0.039	0.646	0.615	0.614	0.625	0.015
0/−1	1.455	1.410	1.401	1.422	0.023	1.329	1.331	1.372	1.344	0.020
q^{SCS} [e]	+1	+2	+2	-	-	+2	+2	+2	-	-
E_f^{SCS} [eV]	2.241	2.096	2.155	2.164	0.059	2.123	2.047	2.193	2.130	0.049

Considering that it was possible to converge the transition levels using SQSs, whereas it was uncertain if the VCA transition levels were converged, using SQSs to simulate charged point defects in the random alloy $In_{0.53}Ga_{0.47}As$, appears to be the best option. The fact that the mid-gap transition levels in the 5x5x5 VCA supercell, are less than 0.04 eV higher than the mid-gap transition levels in the 6x6x6 SQS supercells, lends credibility to the convergence of the VCA transition levels. Despite this, even if the VCA transition levels are converged in the 5x5x5 supercell, the main benefit of using the VCA is the ability to use minimal unit cells. Having to use supercells of similar sizes as those of the SQSs, renders this benefit ineffective,

which further reinforces SQSs as the better option to simulate defects in $In_{0.53}Ga_{0.47}As$. Considering also the scientific risk with using the unphysical virtual atom in the VCA, it is concluded, that using SQSs is the correct choice for the calculations to be carried out in this thesis. Considering the minimal fluctuations of the transition levels in the 6-repetitions SQS supercells, a single SQS will be sufficient for the subsequent calculations. The SQS to be utilised for the remaining calculations, is chosen to be the one resulting from the evolutionary algorithm with the first set of algorithm parameters.

Having found the optimal $In_{0.53}Ga_{0.47}As$ model wherein to calculate transition levels of point defects, it will now be used to simulate a wide array of experimentally feasible point defects.

4.4.3 Point defect study

Experimentally feasible point defects include substitutions with common contaminants or dopants. The contaminants to be studied are carbon and silicon, which could originate from the ambient atmosphere in the growth chamber, and from the growth substrate or oxide, respectively. The dopants to be studied are tin and zinc, which could be incorporated intentionally or have diffused into the host material from adjacent doped regions, such as from the n- and p-InP regions in the heterojunction PIN diode in Section 1.4. In principle, carbon and silicon could also be used as dopants, but tin and zinc are more commonly used for $In_{0.53}Ga_{0.47}As$. All substitutions just mentioned will be examined substituting any of the bulk species. The computed transition levels of these extrinsic point defects, around their expected stable charge states, are shown in Figure 4.18. The values of all the calculated transition levels, including those which fall outside of the limits of Figure 4.18 are given in Table 4.3, along with the formation energies of the defects in their calculated stable charge states.

Examining the stable charge states of the dopants, it is seen that the group IV elements have +1 charge states when replacing the group III elements, meaning they have released an electron to the bulk, and -1 charge states when replacing the group V elements, meaning they have released a hole to the bulk, i.e. they work as either n- or p-dopants, depending on the element which they substitute, as expected from their valencies. Which situation is more probable, can be estimated from their formation energies. It is seen that the silicon and tin dopants favour replacing gallium and indium respectively, with formation energies approximately 1 eV lower than by replacing arsenic, in intrinsic $In_{0.53}Ga_{0.47}As$, thereby making these n-type dopants. Considering that their +1/0 transition levels lie high up in the conduction band, they will keep functioning as n-dopants for extremely high doping levels. In contrast, it can be seen that carbon favours replacing arsenic over indium or gallium, with a formation energy approximately 1.2 eV lower, thereby making carbon a p-type dopant in $In_{0.53}Ga_{0.47}As$. However, its 0/-1 transition level lies 0.074 eV above the valence band maximum, which thus defines the maximum doping level that can be achieved with carbon. Considering that carbon is a much smaller atom than indium, gallium and arsenic, it might

be more energetically favourable in an interstitial site, rather than as a substitution. As a group IV element, carbon in an interstitial site is not expected to act as a dopant, which was confirmed in a small validation calculation, making its study as a charged recombination center, the overall theme of this chapter, uninteresting, which is why it was not studied in this thesis. The zinc dopant is seen to favour replacing indium over arsenic, in a stable charge state of -1, with a substantial formation energy difference of 2.5 eV, making zinc a p-dopant in $In_{0.53}Ga_{0.47}As$. Like the carbon dopant, it has an upper doping level defined by its 0/-1 transition level, located slightly above the valence band maximum. It is important to note, that the doping-type of the dopants considered here, are specific to the $In_xGa_{x-1}As$ stoichiometry of $x = 0.53$. It is well-known, in the case of the carbon dopant, that its dopant-type will change, depending on the stoichiometry of the InGaAs alloy [136].

Transition levels in $In_{0.53}Ga_{0.47}As$ have been widely studied by the experimental scientific community, using an array of techniques such as temperature dependent Hall measurements [137], Scanning-Tunneling Microscopy (STM) [26], the conductance method [121, 122, 138] and Deep-Level Transient Spectroscopy (DLTS) [33, 139–141]. Comparing the transition levels of the examined extrinsic point defects with the experimental data shown in Figure 4.18, it is seen that the 0/-1 transitions of the C_{As}, Si_{As}, Zn_{In} and Zn_{Ga} defects, all approximately align with the lower experimentally determined transition level, marked with a circle. The circle transition level has a donor-like nature [121], which invalidates these extrinsic defects as the origin of the experimentally observed transition level.

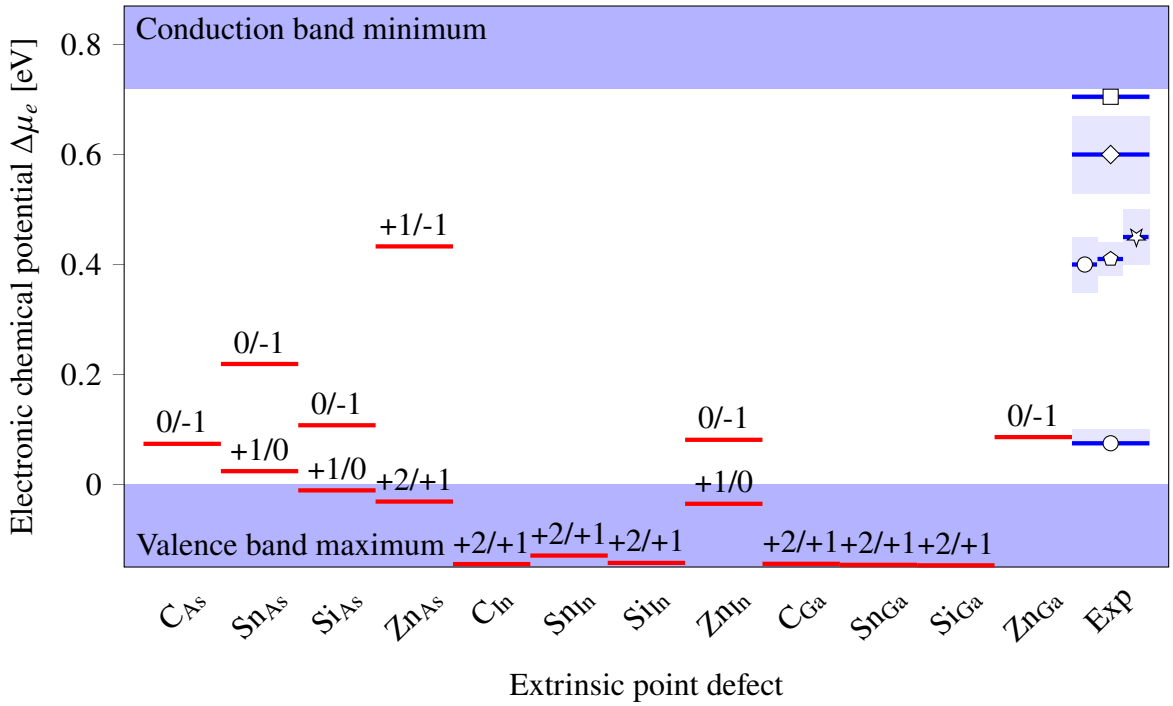


Figure 4.18: Transition levels of extrinsic point defects in $In_{0.53}Ga_{0.47}As$. The full study is given in Table 4.3. The experimental transition levels are measured on InGaAs/oxide interfaces and are determined with various methods (square [137], diamond [26], circle [121], polygon [33], star [138]). Uncertainties in the experimental values are shown by the blue shaded area.

It is also seen in Figure 4.18, that the +1/-1 transition level of Zn_{As} approximately aligns with the middle transition level, observed in multiple experiments with multiple techniques, marked with a circle, polygon and a star, respectively [33, 121, 138]. Considering that this transition level is observed experimentally in both doped and undoped samples, it is likely a transition level associated with an intrinsic defect, rather than a dopant. Adding to this the unlikelihood of the Zn_{As} defect ever forming, due to the substantial formation energy difference between the different zinc substitutions, it can be safely concluded that the Zn_{As} defect is not the origin of the experimentally observed transition level.

Table 4.3: Transition levels and formation energies of studied extrinsic point defects in $In_{0.53}Ga_{0.47}As$. The given formation energy, is the formation energy of the stable charge state at the chemical potential of the intrinsic neutral bulk. Grey-shaded numbers are transition levels which are not shown in Figure 4.18.

Defect	+3/+2 [eV]	+2/+1 [eV]	+1/0 [eV]	0/-1 [eV]	-1/-2 [eV]	-2/-3 [eV]	q^{SCS} [e]	E_f^{SCS} [eV]
C_{As}	-	-	-	0.074	1.447	-	-1	3.598
Sn_{As}	-	-	0.024	0.219	1.482	-	-1	1.794
Si_{As}	-	-	-0.011	0.108	1.474	-	-1	1.641
Zn_{As}	-	-0.031	0.433		1.285	-	+1	3.038
C_{In}	-	-0.145	0.968	1.221	-	-	+1	4.994
Sn_{In}	-0.282	-0.129	1.250	1.535	-	-	+1	0.449
Si_{In}	-	-0.142	1.222	1.532	-	-	+1	0.674
Zn_{In}	-	-	-0.035	0.081	1.471	-	-1	0.474
C_{Ga}	-	-0.144	0.995	-	-	-	+1	4.802
Sn_{Ga}	-	-0.146	1.234	-	-	-	+1	0.525
Si_{Ga}	-	-0.147	1.243	-	-	-	+1	0.623
Zn_{Ga}	-	-	-	0.086	1.503	-	-1	0.546

In addition to the aforementioned extrinsic defects, experimentally feasible point defects also include all intrinsic defects, which can occur due to the random nature of material growth, which is not perfectly controlled, or due to strain relaxations imposed by lattice-mismatched substrates, interfaces or oxides. In the case of $In_{0.53}Ga_{0.47}As$, it is specifically hypothesised that the interface oxidation induce intrinsic defects at and below the interface [121, 122], which are the defects primarily occurring in $In_{0.53}Ga_{0.47}As$. The intrinsic defects to be studied include the As_{In} antisite, which was the defect used for the VCA versus SQS comparison, as well as the other antisites: As_{Ga}, In_{As}, Ga_{As} and the double antisites In_{As}-As_{In} and Ga_{As}-As_{Ga}. The list of defects to study will also comprise the vacancies V_{In}, V_{Ga}, V_{As} and the Schottky defects V_{As}-V_{In}, V_{As}-V_{Ga} at two neighbouring III-V sites. The calculated transition levels of these intrinsic point defects are illustrated in Figure 4.19. The values of all the calculated transition levels, including those which fall outside the limits of Figure 4.19, are given in Table 4.4, along with the formation energies of the defects in their calculated stable charge states.

Inspecting the various intrinsic defects involving either the gallium site or the indium site, it is seen that the transition levels and formation energies of these defects are almost identical in most cases. In fact, the same was observed for the extrinsic defects in Figure 4.18. This can be explained by the almost identical local environment on the sites and the chemical similarity between the two species. The only observed exceptions are the Schottky defects, wherein the defect states populated in the -1 charge states, de-hybridise with the conduction bands upon their occupation, causing large energy-shifts of the states, which differ in the two investigated defects. This phenomena is shown in Appendix A. The difference between these energy-shifts in the two defects could be caused by the difference in the group III dangling bonds in the defects, i.e. the difference between having a surplus gallium (V_{As}-V_{In}) or indium (V_{As}-As_{Ga}) dangling bond in the defect. It could also be caused by the Schottky defects' sensitivity to the local and near-local environment, enhanced by the interaction of the dangling bond pairs between the two constituting vacancies. Despite this apparent enhanced sensitivity to the local environment, or to the species atomic chemistry, the formation energies of the stable charge states remain similar, regardless of the change in the SCS.

To properly account for the effect of the nearest neighbour stoichiometry, one should in principle make calculations for all possible types of local environments, and then compute weighted averages according to the fractional occurrences of each local environment in the system. Based on tests on a few of the studied defects, including the V_{As}-V_{Ga} Schottky defect, it was found that the effect of the stoichiometry of the local environment, was generally small. For this reason, it was not studied further and a local environment, representative of the most commonly occurred environment in the alloy, was used instead. Considering that the transition levels of the V_{As}-V_{Ga} Schottky defect does not change substantially, when the local environment stoichiometry is changed, it would appear that the differences between the two Schottky defects stem from the chemistry of the involved species or the surplus dangling bond, rather than the local environment. This hypothesis is further strengthened when inspecting the arsenic site antisites, In_{As} and Ga_{As}. The arsenic site for these defects were carefully chosen, to be one with two of each group III species as nearest neighbours, ensuring consistent and representative results. Despite this identical and unbiased local environment, the transition levels of the two antisites exhibit non-negligible differences, which must then necessarily originate from the chemical differences between the species. Whereas the Schottky defects illustrate a substantial effect from the choice of the removed group III species, and the arsenic antisites exhibit a non-negligible effect from the chosen group III substitution species, this choice appears to be unimportant in the double antisites, In_{As}-As_{In} and Ga_{As}-As_{Ga}. It is likely, that the effect which is causing the difference between the arsenic antisites, which is possibly the difference between having Ga-Ga bonds or In-In bonds, is damped out by the other interactions in the double antisites, which themselves are independent of the chosen group III species, thereby causing the two double antisites to be near identical.

Comparing the experimentally determined trap levels with the calculated transition levels in Figure 4.19, the experimentally observed defects can be attempted to be identified. The mid gap states align with the -1/-2 transitions of the $V_{In/Ga}$ vacancies and the +2/+1 transition

of the $As_{In/Ga}$ antisites, and considering that the circle transition level has a donor-like nature [121], the antisites are the more probable origin of these states. The transition levels close to the conduction band aligns with the -2/-3 transition of the $V_{In/Ga}$ vacancies and the +1/0 transition of the $As_{In/Ga}$ antisites for the lower level (diamond) and the the 0/-1 transition of the Ga_{As} antisite for the upper level (square). The transition level close to the valence band aligns with many of the defects, and it has even been shown to align with the As dangling bond [142]. It is however noteworthy, that it aligns with the +1/0 transition in the In_{As} antisite, which is a donor-like transition like the observed experimental state. According to the calculated formation energies shown in Table 4.4, the antisites are the most likely occurring of these defects. Additionally, it has been suggested that a charge compensating effect produces antisites of one kind ($As_{In/Ga}$), in response to antisites of the other kind ($(In/Ga)_{As}$), in polar materials like InGaAs [143]. This charge compensation rule can be used to further strengthen the identification of the square trap level as the Ga_{As} antisite 0/-1 transition level. The density of the square trap level defect was found to increase with the excess arsenic during the growth [137]. Under As-rich conditions, the indium and gallium site antisites are more likely to form, due to their formation energies being lowered [119], and because of the charge compensating effect, more arsenic site antisites should also form as a consequence, thereby explaining the increased density of the square trap level.

From the alignment with calculated transition levels, the calculated defect formation energies and the described compensation rule and its application, the antisites are concluded to be the most likely source of the experimentally observed transition levels.

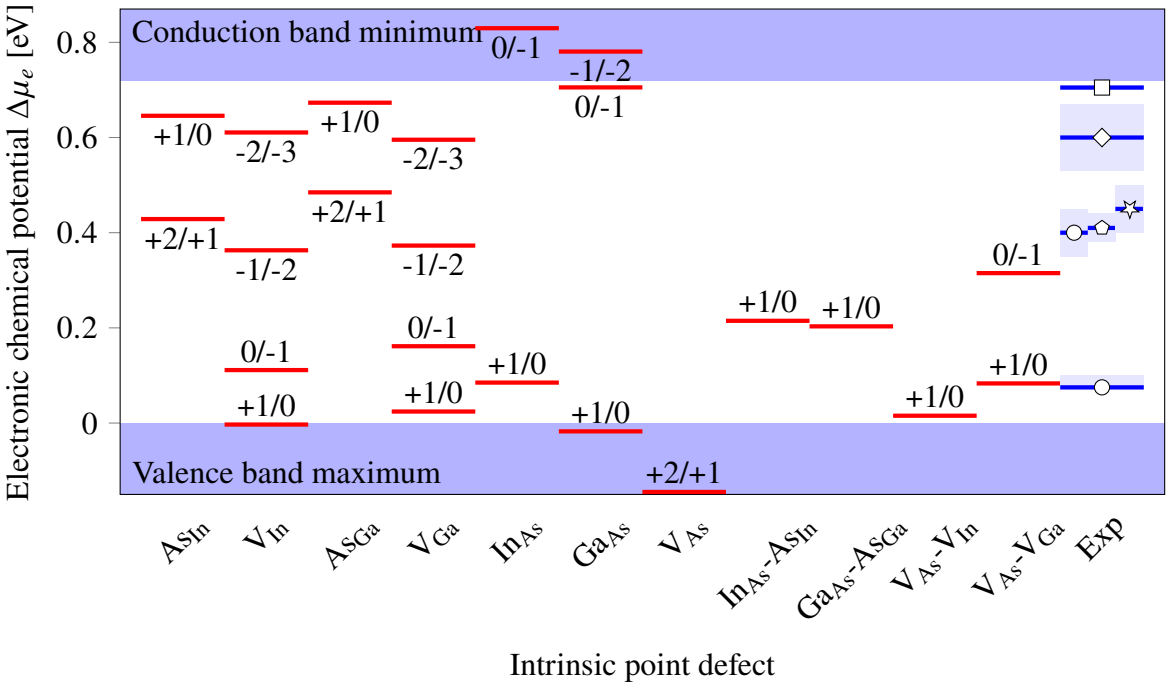


Figure 4.19: Transition levels of intrinsic point defects in $In_{0.53}Ga_{0.47}As$. The full study is given in Table 4.4. The experimental transition levels are measured on InGaAs/oxide interfaces and are determined with various different methods (square [137], diamond [26], circle [121], polygon [33], star [138]). Uncertainties in the experimental values are shown by the blue shaded area.

While the double antisites In_{As}-As_{In} and Ga_{As}-As_{Ga} are not directly detected in the experiments, this is not conclusive evidence of them not forming at all. Since they consist of the experimentally found singular antisites, and taking into account that their combination is energetically favoured over the sum of the constituting single antisites, it is reasonable to assume that they should also form, albeit possibly in a much smaller density, owing to the low likelihood of two singular antisites coinciding in space.

Table 4.4: Transition levels and formation energies of intrinsic point defects in $In_{0.53}Ga_{0.47}As$. The given formation energy, is the formation energy of the stable charge state at the chemical potential of the intrinsic neutral bulk. Grey-shaded numbers are transition levels which are not shown in Figure 4.19.

Defect	+3/+2 [eV]	+2/+1 [eV]	+1/0 [eV]	0/-1 [eV]	-1/-2 [eV]	-2/-3 [eV]	q^{SCS} [e]	E_f^{SCS} [eV]
As_{In}	-0.261	0.429	0.646	1.329	-	-	+2	2.123
V_{In}	-	-	-0.003	0.111	0.363	0.611	−2	4.121
As_{Ga}	-0.251	0.485	0.673	1.348	-	-	+2	2.116
V_{Ga}	-	-	0.024	0.161	0.373	0.595	−2	4.108
In_{As}	-	-	0.085	0.830	0.948	-	0	2.524
Ga_{As}	-	-	-0.017	0.705	0.781	-	0	2.392
V_{As}	-0.260	-0.145	0.922		-	-	+1	2.667
In_{As}-As_{In}	-	-	0.215	1.151	-	-	0	2.921
Ga_{As}-As_{Ga}	-	-	0.203	1.141	-	-	0	2.879
V_{As}-V_{In}	-	-	0.015	1.003	-	-	0	5.165
V_{As}-V_{Ga}	-	-	0.083	0.315	-	-	−1	5.056

When investigating transition levels, it is important to distinguish between "shallow trap levels", which are transition states in the band gap near the band edges, and "deep trap levels", which lies far above (deep below) the valence band maximum (conduction band minimum). Shallow trap levels often behave like dopants, depending on the valency of the defect and the charge states involved in the transition level. Deep trap levels act as carrier recombination centres and thus have a detrimental impact on the current conducting properties of the host material.

In this thesis, the transition levels are partitioned at 0.2 eV ($\approx 8k_BT$ at 300 K) above (below) the valence band maximum (VBM) (conduction band minimum (CBM)). This threshold was chosen, because defect states associated with transition levels below this threshold, often hybridise with the conduction (valence) band states and thereby de-localise, and for a trap level to act as a recombination centre, having a localised defect state is a necessity. Examining the transition levels of the extrinsic defects in Figure 4.18, it is seen that the 0/-1 level of the Sn_{As} substitution ($\approx 8k_BT$ above VBM) and the +1/-1 level of the Zn_{As} substitution ($\approx 11k_BT$ below CBM) can be categorised at deep trap levels. Inspecting the transition levels of the intrinsic defects in Figure 4.19, reveals that the +2/+1 transition levels of the $As_{In/Ga}$ antisites ($\approx 9-11k_BT$ below CBM), the -1/-2 levels of the $V_{In/Ga}$ vacancies ($\approx 14k_BT$ above (below)

VBM (CBM)), the +1/0 levels of the double antisites ($\approx 8k_BT$ above VBM) and the 0/-1 level of the V_{As}-V_{Ga} Schottky defect ($\approx 12k_BT$ above VBM) can all also be categorised as deep trap levels. To learn how detrimental these specific defects are on the current transport properties of $In_{0.53}Ga_{0.47}As$, the capture coefficients and trap-assisted recombination rates associated with these deep trap levels, will be calculated in the following section.

4.5 Shockley-Read-Hall recombination

One physical process which can drastically reduce the efficiency of semiconductor devices, is the recombination of excited or injected carriers. In semiconductors, excited electrons can occasionally recombine with their empty valence band states, i.e. holes, effectively eliminating both carriers. The three most important recombination processes are illustrated schematically in Figure 4.20. The first process is the radiative direct band-to-band recombination, wherein photons are emitted with a total energy corresponding to the band gap of the semiconductor. This process is the direct opposite to photon absorption and is the desired recombination mechanism of photon emitters and lasers. The second process is the non-radiative trap-assisted recombination mechanism, called the Shockley-Read-Hall (SRH) recombination [144]. In SRH recombination, the excited electron relaxes first to an intermediate trap level E_t, before relaxing further down to the valence band and recombines with the hole. The released energy from the recombination is here usually dissipated as phonons, instead of photons. The last illustrated recombination process is the Auger recombination, wherein the direct band-to-band recombination is facilitated by exciting a carrier to a higher energy state, rather than by emitting a photon.

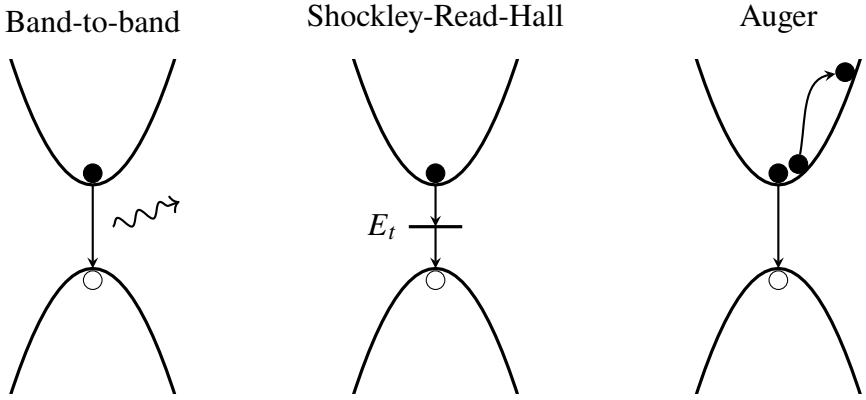


Figure 4.20: The three main recombination mechanisms in semiconductors.

While all three recombination processes illustrated in Figure 4.20 are important in various semiconductor devices under various conditions, the purpose of this thesis chapter is the study of the SRH recombination mechanism, as it is thought to be the performance-limiting process in *p*-InP/*i*-InGaAs/*n*-InP PIN photodiodes [61, 63]. The SRH recombination in the deep trap level of a defect, such as those found in the previous section, depends upon both of the two separate electron relaxation processes, also known as carrier capture processes, since a free carrier is captured by the defect in both processes. These carrier capture processes occur at

a rate $R_{n/p}$ given by

$$\begin{aligned} R_n &= C_n N_D^A n \\ R_p &= C_p N_D^A p \end{aligned} \tag{4.6}$$

where n and p are the density of electrons and holes respectively and N_D^A is the density of the subset of the defects in their accepting charge state, which is the more negative (less positive) charge state of a transition for hole capture and the more positive (less negative) charge state for electron capture. The last coefficient $C_{n/p}$ is a trap-carrier specific capture coefficient with units $[C_{n/p}] = \mathrm{cm}^3\,\mathrm{s}^{-1}$, which is a numerical representation of the probability of an excited electron or hole to be "caught" in the trap level of the defect. The capture coefficient is related to the capture cross section $\sigma_{n/p}$, which is often the value presented in the literature, by the characteristic carrier velocity:

$$C = \langle v \rangle \sigma \tag{4.7}$$

The characteristic carrier velocity $\langle v \rangle$ in Eq. (4.7) is the average thermal velocity for non-degenerate statistics. The calculation of the electron and hole capture coefficients of the various trap levels found in the previous section, is the goal of the current section.

4.5.1 Capture coefficients calculation method

To calculate the SRH recombination capture coefficients, the methodology developed by Alkauskas et al. [145] is used. In this method, it is assumed that the screened defect potential range λ, is much shorter than the average distance between carriers $\lambda \ll n^{-1/3}$. Here the defect potential is the Coulomb potential for charged defects, and the short-range potential for neutral defects. The consequence of this assumption, is that each defect only interacts with one carrier at a time, thereby simplifying the calculations of the capture coefficient to only consider a single carrier at a single defect

$$C = Vr \tag{4.8}$$

where V is the volume of the system and r is the capture coefficient of a single carrier in a single defect in the entire volume V. This single-carrier capture coefficient, or capture rate, can be calculated from Fermi's golden rule:

$$r = \frac{2\pi}{\hbar} g \sum_m w_m \sum_j |\Delta H_{im;fj}^{\mathrm{e-ph}}|^2 \delta(E_{im} - E_{fj}) \tag{4.9}$$

In Eq. (4.9), w_m is the thermal occupation of the vibrational state m of the initial electronic state i, $\Delta H_{im;fj}^{\mathrm{e-ph}}$ is the electron-phonon coupling facilitating the transition between this state, and the final electronic state f and its vibrational state j, E_{im} and E_{fj} is the total energies of these states and g is the degeneracy of the final state. This degeneracy could for example be of structural nature, such as if in the final state, the defect has four energetically identical orientations.

To continue with the derivation, further approximations are made. Firstly, the static

approximation is employed. In the static approximation, the wave function of the system can be described as the product of an electronic wave function $\Psi(\{Q_0\}, \{x\})$, calculated in some initial ionic configuration Q_0, and the ionic wave function $\chi(\{Q\})$. Here $\{x\}$ represents all electronic degrees of freedom and $\{Q\}$ represents all ionic coordinates. In this approximation, the electron-phonon coupling becomes [146] $\Delta H^{\mathrm{e-ph}} = H(\{Q\}, \{x\}) - H(\{Q_0\}, \{x\})$. Taylor expanding this around Q_0 and keeping only the linear term, an approximation also called the linear coupling approximation, yields the following expression for the electron-phonon coupling matrix

$$\Delta H^{\mathrm{e-ph}}_{im;fj} = \sum_k \underbrace{\langle \Psi_i | \, \partial H / \partial Q_k \, | \Psi_f \rangle}_{w^k_{if}} \langle \chi_{im} | \, Q_k - Q_{0;k} \, | \chi_{fj} \rangle \tag{4.10}$$

where the sum runs over all phonon modes Q_k, $Q_{0;k}$ is the projection of the initial ionic configuration unto phonon mode k and w^k_{if} is the electron-phonon coupling of the phonon mode.

Calculating the electron-phonon coupling according to Eq. (4.10), quickly becomes unfeasible for large systems, such as those used for the charged point defect calculations in Section 4.4.3. To simplify the calculation further, the special phonon mode approximation is used. In this approximation, the sum over phonon modes in Eq. (4.10) is replaced with a single effective phonon mode. This approximation is motivated by the large energy difference between the highest energy phonon, i.e. the longitudinal optical phonon at 34 meV in $In_{0.53}Ga_{0.47}As$ [102], and the typical energy of the deep trap levels, which is at minimum 200 meV, with the chosen threshold. This large energy difference thus necessitates multi-phonon emission (MPE) to facilitate the transition. Additionally, the second term in Eq. (4.10) ensures that only phonon modes which couples to the distortion, contribute to the electron-phonon coupling. As such, using a single 1-dimensional effective phonon mode, constructed from the lattice distortion of the transition, is a reasonable approximation [147]. Formally, this effective phonon mode is constructed as a displacement along the difference in ionic coordinates of the equilibrium configuration, in the initial (excited) state $R_{i:\alpha t}$, and the final (ground) state $R_{f:\alpha t}$, of the involved charge transition. This effective phonon mode is a 1-dimensional mode along the generalised configuration coordinate Q axis, which is defined as

$$Q^2 = \sum_{\alpha,t} m_\alpha \left(R_{\alpha t} - R_{f:\alpha t} \right)^2 \tag{4.11}$$

where m_α is the atomic mass of atom α and $t = x, y, z$. In this generalised configuration coordinate axis, $Q = 0$ when the ionic configuration is that of the final (ground) state, and $Q = \Delta Q$ when it is that of the initial (excited) state.

The frequencies of the effective phonon mode in the two charge charge states in the harmonic approximation, are then given by the second order derivative with regards to the

generalised configuration coordinate, around the charge states' equilibrium configurations:

$$\Omega_{i/f}^2 = \frac{\partial^2 E_{i/f}}{\partial Q^2} \tag{4.12}$$

This 1-dimensional effective phonon model is illustrated in the configuration coordinate diagram in Figure 4.21. Here, the higher energy excited initial state includes the free carrier to be caught $e^{\pm}$ (either an electron or a hole), and the trap-level defect in its corresponding accepting charge state D^A. The lower energy final ground state then represents the defect in the corresponding charge state after having caught the carrier D^F. In principle, the carrier capture process of SRH recombination has two steps, shown by the arrows in the diagram. The first step is the energy-conserving charge-transition, and the second step is the lattice relaxations. The lattice relaxations usually occur at much faster timescales than the charge-transition [148], which means that they can be neglected for calculations of the SRH recombination rate.

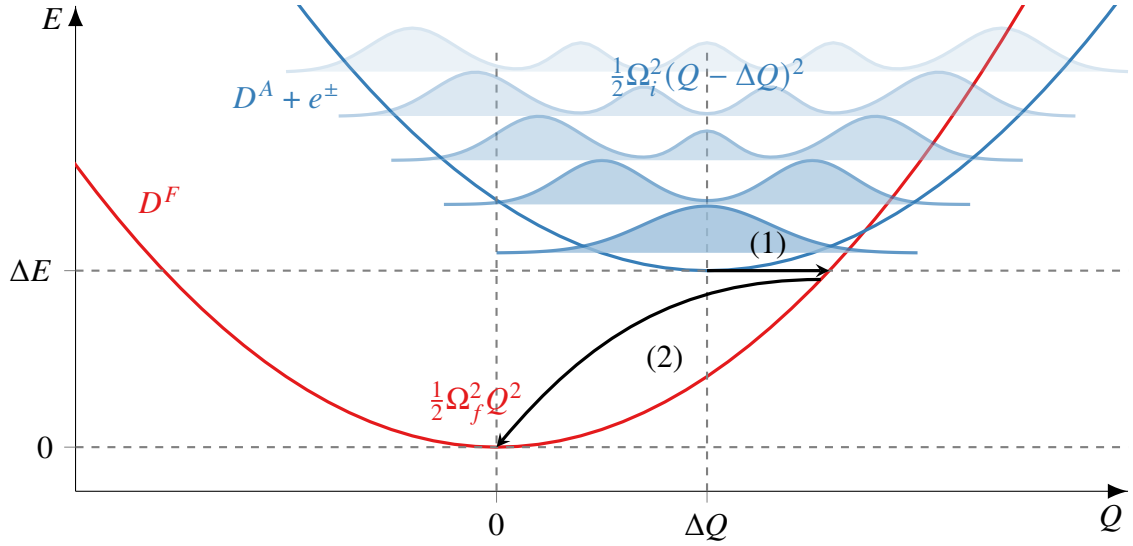


Figure 4.21: Configuration coordinate diagram of carrier capture, with the effective phonon mode model. The upper initial state represents the accepting defect and the free carrier, whereas the lower final state represents the satisfied defect. The two sub-processes of carrier capture is also shown, with the first being the energy-conserving charge-state transition, and the second being the relaxations of the lattice. In the upper initial state, the first 5 harmonic oscillator modes are illustrated, with their fade representing their thermal occupancy.

A useful dimensionless quantity, the Huang-Rhys factor S, which describes the strength of the electron-phonon interaction, is also calculated from the generalised coordinate displacement [149]:

$$S_{i/f} = \frac{1}{2\hbar}(\Delta Q)^2 \Omega_{i/f} \tag{4.13}$$

Using the 1-dimensional effective phonon model, together with the independent particle picture of DFT, to replace the many-body Hamiltonian H and its wave functions Ψ, with their corresponding single-particle counterparts h and ψ, the calculation of the electron-phonon

coupling in the static and linear-coupling approximation becomes

$$\tilde{W}_{if} = \langle\tilde{\psi}_i|\frac{\partial\hat{h}}{\partial Q}|\psi_f\rangle = (\varepsilon_f - \varepsilon_i)\left\langle\tilde{\psi}_i\left|\frac{\partial\psi_f}{\partial Q}\right.\right\rangle \tag{4.14}$$

where ε_i and ε_f are the energies of the single-particle states ψ_f and $\tilde{\psi}_i$, which are then respectively, the localised defect state, and the bulk state that corresponds to a hole (electron) in the valence (conduction) band, perturbed by the defect.

Putting together the electron-phonon coupling in Eq. (4.14), with Eq. (4.10), Fermi's golden rule in Eq. (4.9), the volume scaling in Eq. (4.8) and the described effective phonon model, one arrives at the following expression for the SRH capture coefficient:

$$\tilde{C}_{n/p} = \tilde{V}\frac{2\pi}{\hbar}g\tilde{W}_{if}^2\sum_m w_m\sum_n\left|\langle\chi_{im}|Q - Q_0|\chi_{fn}\rangle\right|^2 \times \delta(\Delta E + m\hbar\Omega_i - n\hbar\Omega_f) \tag{4.15}$$

Here, $\tilde{V}$ is the volume of the supercell, g is the degeneracy of the final state, which is 1 for all the trap levels found in Section 4.4.3, ΔE is the energy difference between the initial and final states, i.e. the trap level, w_m is the thermal occupation of the vibrational state m, $|\chi_{im/fn}\rangle$ is the ionic wave function of the initial (final) state in vibrational state m (n) and Q_0 is the initial configuration. The electron-phonon coupling $\tilde{W}_{if}$ can then be calculated in either the initial (excited) state or the final (ground) state of the capture process, or a combination of the charge state and configuration thereof, providing four options in total. In this work, the configuration and charge state of the initial (excited) state was chosen for all transitions, since this state is the physical state which captures carriers to the local defect level.

Similarly to the defect transition levels, the capture coefficient calculated with Eq. (4.15) has to be corrected for finite-size effects, caused by the presence of a localised charge in the limited size of the supercell. Additionally, since the single-particle bulk carrier state has to be taken from the defect configuration, in order for it to belong to the same Hamiltonian as the single-particle defect state, the capture coefficient also has to be corrected for the spurious Coulombic interaction between the charged defect and the delocalised bulk carrier:

$$C_{n/p} = f(T)\tilde{C}_{n/p} \tag{4.16}$$

The dimensionless scaling factor $f(T)$ in Eq. (4.16), is the product of the supercell scaling correction [150], and the second correction, known in the literature as the Sommerfeld parameter [151]. The finite-size correction can be estimated from the inverse ratio of the integrated radial charge density of the defect system, to the integrated radial charge density of a homogeneous distribution, which is simply $(4/3)\pi R^3/V$. The Sommerfeld parameter, which describes the enhancement or suppression of a delocalised wave function near a charge, is given as

$$s(\mathbf{k}) = -\frac{2\pi Z}{a_b|\mathbf{k}|}\frac{1}{1 - \mathrm{e}^{2\pi Z/a_b|\mathbf{k}|}} \tag{4.17}$$

where Z is the charge and $a_b = 4\pi\epsilon_0\hbar^2/m_b e^2$ is the effective Bohr radius. To include

this momentum-dependent Sommerfeld parameter in the scaling factor, Eq. (4.17) is first temperature averaged either numerically or analytically, the latter of which requires the approximation $|\mathbf{k}| \ll 2\pi|Z|/a_b$.

4.5.2 Recombination capture coefficients

In order to use Eq. (4.15) to calculate SRH capture coefficients, it is first necessary to identify the single-particle states associated with the localised defect and the delocalised carriers. To do this, the electronic Inverse Participation Ratios (IPRs) of the defect systems were calculated, for the charge states involved in deep trap level transitions. The IPR of a state is a measure of the degree of localisation of the state. For state n, it is defined as [152]

$$p_n^{-1} = \sum_{i=1}^{N} |q_i^{(n)}|^2, \qquad q_i^{(n)} = \sum_{\mu \in i, \nu} S_{\mu\nu} c_{\mu n} c_{\nu n} \tag{4.18}$$

where $q_i^{(n)}$ is the contribution from site i to the electronic IPR p_n^{-1} of state n, $c_{\mu n}$ is the nth state's basis set coefficient for orbital μ, $S_{\mu\nu}$ is the overlap matrix element between orbital μ and ν and μ is a summation index running over only the atomic orbitals of site i, whereas ν runs over all orbitals on all sites. For a fully delocalised state, the IPR is $1/N$, where N is the number of atoms in the system.

To identify the defect and carrier states from the IPRs, a state close to the Fermi level with a comparatively high IPR, is identified as the localised defect state. The conduction and valence band states are then subsequently identified, as states with low IPR on either side of the defect state, with an energy difference close to the pristine band gap. An example of this state identification is shown in Figure 4.22, for the +2 and +1 charge states of the As_{In} defect.

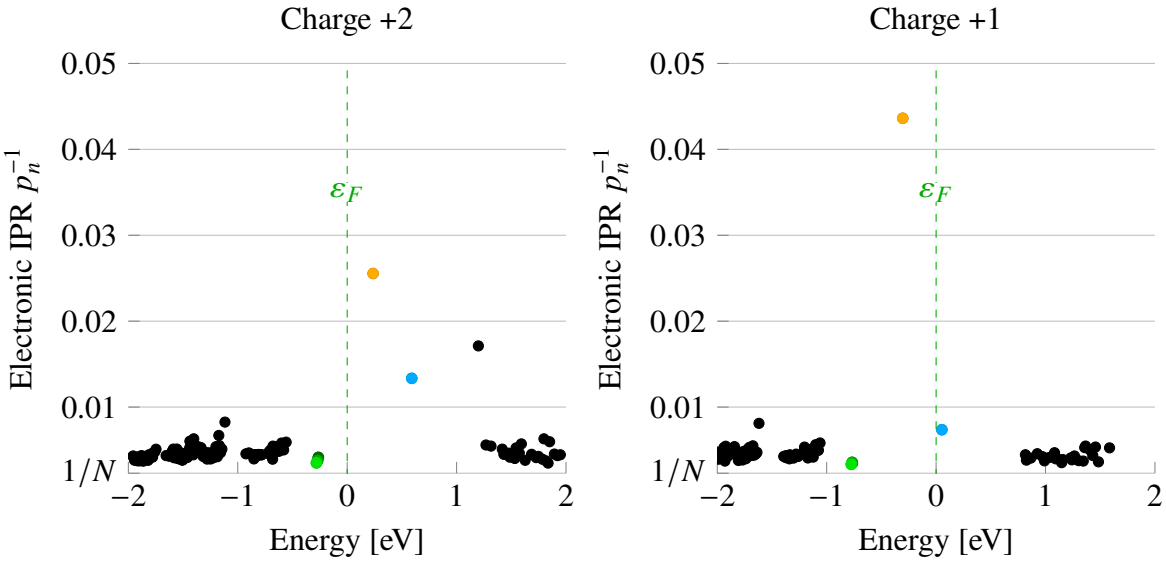


Figure 4.22: Electronic inverse participation ratio (IPR) of states close to the Fermi level, in the system consisting of the defect As_{In} in $In_{0.53}Ga_{0.47}As$, for the charge states +2 and +1. The defect, conduction and valence band states are marked in orange, blue and shades of green, respectively, with the first, second and third shade of green, representing the three different valence bands.

This state identification method is backed up by band structure calculations, such as those shown in Appendix A, wherein the shape of the band corresponding to a specific state, can be compared with the relevant band in the pristine $In_{0.53}Ga_{0.47}As$ configuration, in order to ascertain the truth of the identification.

The state identification was carried out for all the deep trap levels found in Section 4.4.3. For the Zn_{As} substitution and the V_{As}-V_{Ga} Schottky defect, a localised single-particle state representing the defect, could not be univocally identified in both charge states of the transition. As shown for the Schottky defect in Appendix A, the energy-level of the supposed defect states varies greatly depending on the charge states of the systems. In these defects, since a well-defined defect level does not simply become occupied upon the charge transition, the SRH formalism defined in this section is not applicable to them, and the SRH capture coefficient and rate of these defects, were therefore not calculated in this thesis.

For the remaining defects exhibiting deep trap levels, the SRH capture coefficients for both electron and hole capture were calculated, using Eq. (4.15) as implemented in QuantumATK version U-2022.12. Herein, the derivatives in Eq. (4.14) were evaluated with finite-differences. For the hole capture coefficient, all three energy-degenerate valence band states were used for the calculation of the electron-phonon coupling, yielding a separate coupling element for each of them. The total electron-phonon coupling used for the calculation of the capture coefficient, was then taken as the Euclidean norm of the three separate electron-phonon coupling elements. For the numeric Sommerfeld parameter in Eq. (4.17), effective masses were used, which were calculated from parabolic fits to the band structure of the pristine $In_{0.53}Ga_{0.47}As$ SQS in Figure A.1. The resulting effective masses of $0.33m_0$ and $0.036m_0$, were used for the electron and hole capture respectively. The calculated capture coefficients can be seen in Figure 4.23, as a function of temperature. The hole capture coefficients of the V_{In} and V_{Ga} vacancies are 7 and 12 orders of magnitude lower than the Sn_{As} substitution respectively, and are therefore not included in the figure.

Considering that the SRH recombination scales exponentially with the trap level energy [153], it would be expected that the As_{In} and As_{Ga} antisites have the highest electron capture coefficients and lowest hole capture coefficients, since their deep trap levels are the closest ones to the conduction band, out of all the investigated defects exhibiting deep trap levels. Conversely, the deep trap levels of the In_{As}-As_{In} and Ga_{As}-As_{Ga} double antisites are the closest ones to the valence band, and their capture coefficients would therefore be expected to be the largest for hole capture and the smallest for electron capture. While this assumption is satisfied in regards to the highest capture coefficients, it fails in regards to the smallest capture coefficients and even for the expected intermediate capture coefficients, i.e. the V_{In} and V_{Ga} vacancies, which that have deep trap levels exactly in the middle of the pristine $In_{0.53}Ga_{0.47}As$ band gap. Rather than having intermediate capture coefficients for both electron and hole capture, the vacancies turn out to have hole capture coefficients many orders of magnitude lower than the other defects, and electron capture coefficients among the lowest of the investigated defects.

Additionally, while nearly all of the defects exhibit similar capture coefficients, whether

the defect involves an indium or a gallium atom, which is as expected from their similar deep trap levels and the chemical similarity of the species, the electron capture coefficients of the double antisites, display differences of several orders of magnitude, depending on their group III species. Finally, based purely on the trap level energies, the Sn_{As} substitution should have capture coefficients similar to those of the double antisites, in particular, it should exhibit a comparatively large hole capture coefficient, which is also not the case, as seen in Figure 4.23. Evidently, estimating SRH carrier capture mainly from the position of the deep trap level, can lead to very wrong conclusions about the strength of the capture process, and thereby on the importance of the defect in a physical device.

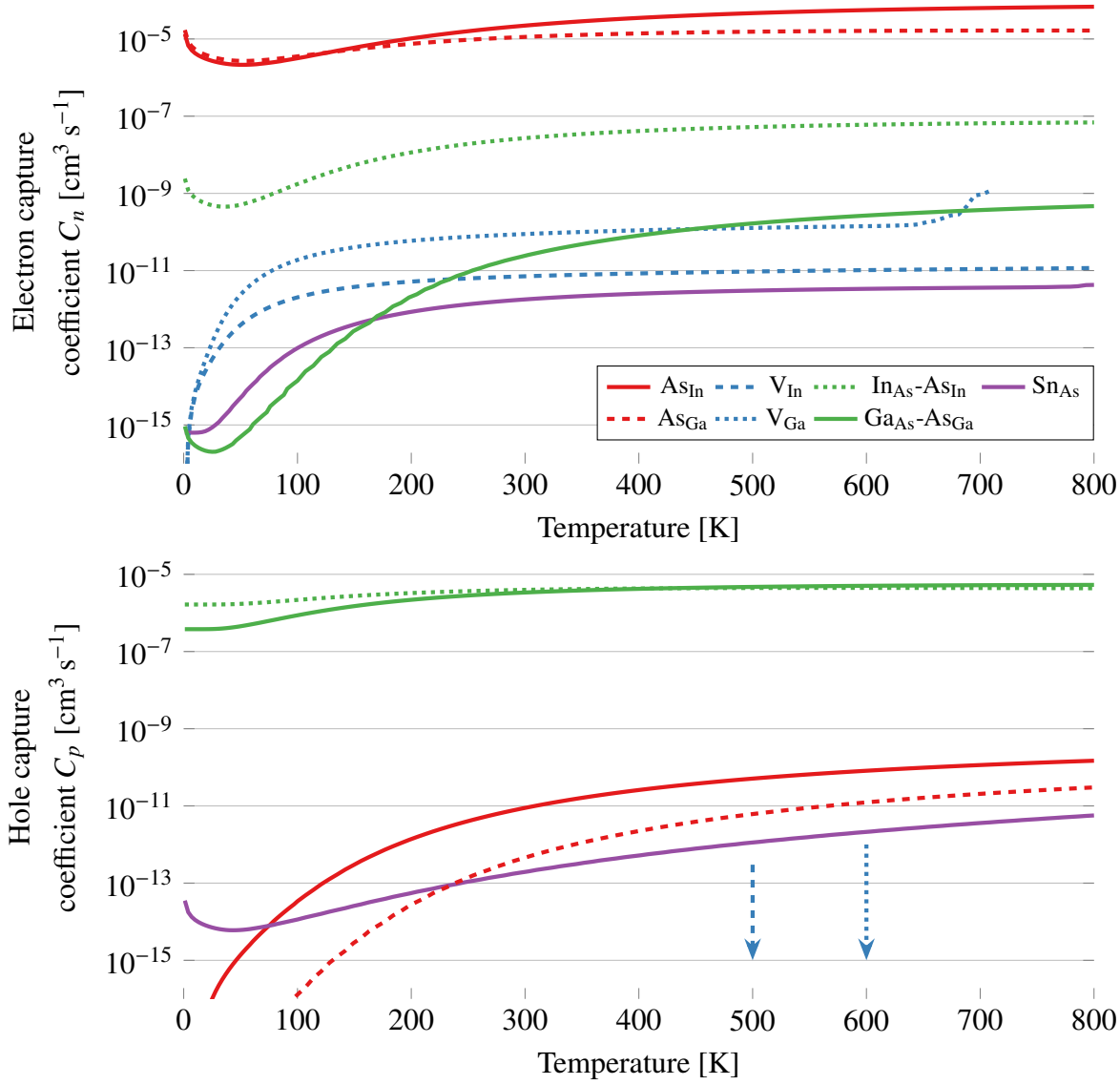


Figure 4.23: Shockley-Read-Hall capture coefficients for electron and hole capture, in the deep trap levels of the point defects in $In_{0.53}Ga_{0.47}As$.

In order to achieve a better understanding of the origins behind the magnitude of the various capture coefficients in Figure 4.23, the main parameters behind the calculations are collected in Table 4.5. Since SRH recombination is a phonon-mediated recombination mechanism, it would make sense if the capture coefficients primarily depended on the calculated electron-phonon couplings. Looking back at the expression for the capture coefficients in

Eq. (4.15), it is indeed seen that the capture coefficients have a quadratic dependency on the electron-phonon couplings. Comparing the electron-phonon couplings in Table 4.5 with the capture coefficients in Figure 4.23, a clear direct proportionality is observed between the electron capture coefficients C_n and the conduction band electron-phonon couplings $\tilde{W}_c$. For the hole capture coefficients, such a direct relationship is unfortunately not apparent, as can be seen, for example, with the Sn_{As} substitution and the Ga_{As}-As_{Ga} double antisite. Here, the substitution defect has larger electron-phonon couplings for all three valence bands, yet its hole capture coefficient is many orders of magnitude lower than the double antisite's.

The reason for this lack of a direct relationship can be explained by considering the Hyang-Rhys factors S of the involved charge states. The Sn_{As} substitution and the V_{In} and V_{Ga} vacancies all have much lower Hyang-Rhys factors than the antisites, as a consequence of their lower generalised configuration coordinate displacements ΔQ and phonon frequencies Ω (recall the definition of the Hyang-Rhys factor in Eq. (4.13)). The Hyang-Rhys factors of the vacancies are particularly low, thereby explaining their particularly low capture coefficients. Ostensibly, the SRH carrier capture coefficients have a strong dependence on both the Hyang-Rhys factors of the charge states involved in the transition, and on the electron-phonon couplings between the defect and the carrier states.

This dependence of the capture coefficients on the Hyang-Rhys factors can be understood by considering the configuration coordinate diagram in Figure 4.21. If the configuration coordinate displacement or the phonon frequencies of the transition are very small, the change in the configuration coordinate Q of sub-process (1) is very large. What this means, is that in such a case, a large sudden structural change of the system is required, for the carrier capture to occur. The relaxations of sub-process (2) usually occur in many small steps, and the magnitude of the change in the configuration coordinate therefore has little impact on it.

Table 4.5: Calculation parameters of Shockley-Read-Hall capture coefficients of the point defects exhibiting deep trap levels in $In_{0.53}Ga_{0.47}As$. The $^{\cdot}/_{\bullet}$ ($^{\bullet}/_{\cdot}$) symbol refer to the more(less) negative charge state of the deep trap transition in the specific defect. The electron-phonon coupling constants are given for all three involved valence band states $\tilde{W}_{v1-v3}$ as well as the conduction band state $\tilde{W}_c$.

Defect	$\hbar\Omega^{\bullet/\cdot}$	$\hbar\Omega^{\cdot/\bullet}$	$S^{\bullet/\cdot}$	$S^{\cdot/\bullet}$	ΔQ	$\tilde{W}_{v1}$	$\tilde{W}_{v2}$	$\tilde{W}_{v3}$	$\tilde{W}_c$
	[meV]				[Å $\sqrt{\text{amu}}$]	[eV/(Å $\sqrt{\text{amu}}$)]			
As_{In}	17.2	15.5	8.26	7.45	2.01	7.31e-4	6.30e-4	1.22e-5	7.14e-2
As_{Ga}	15.8	14.3	8.08	7.30	2.07	1.84e-4	1.25e-4	1.11e-3	7.15e-2
V_{In}	5.8	16.7	1.52	4.35	1.48	5.73e-3	1.92e-2	1.95e-3	1.63e-4
V_{Ga}	4.6	13.3	1.30	3.74	1.53	6.53e-3	1.16e-2	4.03e-3	5.20e-4
Sn_{As}	8.5	16.1	3.24	6.16	1.79	4.47e-2	2.69e-3	9.11e-4	6.86e-5
In_{As}-As_{In}	14.3	17.2	9.06	10.9	2.30	3.75e-2	3.68e-3	5.29e-3	5.58e-3
Ga_{As}-As_{Ga}	11.8	13.4	8.53	9.74	2.46	3.81e-2	4.08e-4	8.32e-5	5.66e-4

4.5.3 Recombination rates

Having calculated the carrier capture coefficients, for both electron and hole capture, for each of the discovered deep trap level defects in $In_{0.53}Ga_{0.47}As$, it is now possible to estimate the InGaAs device performance degradation, caused by the SRH recombination in these defects. This is done by calculating the non-equilibrium steady-state recombination rate, according to the original SRH relation [144, 154]

$$R^{\text{SRH}} = \frac{np - n_i^2}{\tau_n(p + p_t) + \tau_p(n + n_t)} \tag{4.19}$$

where n and p are the usual carrier densities, $n_i = p_i$ is the intrinsic carrier density, n_t and p_t are the carrier densities when the quasi-Fermi level aligns with the trap level and τ_n and τ_p are the single electron and hole lifetimes. These lifetimes can be calculated from the capture coefficients as $\tau_{n/p} = 1/C_{n/p}N_t$, where N_t is the density of the related defect.

The derivation of Eq. (4.19) starts with the rate equations for the four different carrier processes of SRH recombination. The two carrier capture rate equations were given in simple terms in Eq. (4.6). Besides recombining with an oppositely charged carrier, it is also possible for a captured carrier, to be re-emitted back into the conduction or valence bands. The rate equations of these emission processes resemble the expressions for the capture rates, when slightly re-written. The rate equations for the capture processes in this re-written form, as well as the rate equations for the emission processes, are given in Eq. (4.20):

$$\begin{aligned} R_n^C &= C_n(1 - f_t)N_t \int_{E_c}^{\infty} f(E)D(E)dE \quad & R_n^{\mathcal{E}} &= \mathcal{E}_n f_t N_t \int_{E_c}^{\infty} [1 - f(E)]D(E)dE \\ R_p^C &= C_p f_t N_t \int_{\infty}^{E_v} [1 - f(E)]D(E)dE \quad & R_p^{\mathcal{E}} &= \mathcal{E}_p(1 - f_t)N_t \int_{\infty}^{E_v} f(E)D(E)dE \end{aligned} \tag{4.20}$$

Here, R_n^C, R_p^C, $R_n^{\mathcal{E}}$ and $R_p^{\mathcal{E}}$ are the rates for electron capture, hole capture, electron emission and hole emission respectively, and C_n, C_p, $\mathcal{E}_n$ and $\mathcal{E}_p$ are the corresponding rate coefficients. The density of carrier accepting defects N_D^A has been replaced with the general density of trap level defects N_t and a trap level occupation f_t, so that $f_t N_t$ ($[1 - f_t]N_t$) is the density of hole accepting or full (electron accepting or empty) defects. The trap level occupation is given by a Fermi-Dirac distribution at the trap level energy E_t with a trap-specific quasi-Fermi level F_t. Similarly, the conduction band electrons and valence band holes are assumed to be in thermal equilibrium among themselves. This means that their densities n and p can be expressed in terms of integrals over the density of states $D(E)$, and Fermi-Dirac occupation functions $f(E)$, with quasi-Fermi levels F_n and F_p. Due to the different quasi-Fermi levels of the two carriers, their corresponding integrals start (end) at the conduction band minimum E_c (valence band maximum E_v).

An expression for the emission rate coefficients are then derived by considering total system thermal equilibrium $F_t = F_n = F_p$. In thermal equilibrium, the capture and emission rate of a carrier balances out, which leads to an expression for the emission coefficients in

terms of the capture coefficients:

$$\mathcal{E}_{n/p} = C_{n/p}\frac{f(E)[1-f_t]}{f_t[1-f(E)]} = C_{n/p}\mathrm{e}^{\frac{E_t-E}{k_BT}} \tag{4.21}$$

Inserting this expression for the emission coefficients in the rate equations in Eq. (4.20) and evaluating the integrals, using the assumption that the doping levels are at most intermediate, so that $F_n < E_c$ and $F_p > E_v$

$$\begin{aligned} R_n^C &= C_n(1-f_t)N_t n \qquad & R_n^{\mathcal{E}} &= C_n f_t N_t n_t \\ R_p^C &= C_p f_t N_t p \qquad & R_p^{\mathcal{E}} &= C_p(1-f_t)N_t p_t \end{aligned} \tag{4.22}$$

where $n_t = \int_{E_c}^{\infty} f(E_t)D(E)dE$ is the electron density, when the electron quasi-Fermi level is at the trap level, and similarly for p_t.

The final step of the derivation is to consider steady-state conditions, such as a system under bias or a constantly (in the relevant time-scale) illuminated diode. In steady-state, the net capture of electrons is exactly balanced with the net capture of holes, where the net capture is the capture rate minus the corresponding emission rate: $R_n^C - R_n^{\mathcal{E}} = R_p^C - R_p^{\mathcal{E}}$:

$$C_n(1-f_t)n - C_n f_t n_t = C_p f_t p - C_p(1-f_t)p_t \tag{4.23}$$

Using that the product $n_t p_t$ is simply the intrinsic carrier density n_i^2, Eq. (4.23) turns into the SRH relation in Eq. (4.19), with a little re-arrangement of the terms and the introduction of the lifetimes $\tau_{n/p}$.

To calculate the SRH recombination rate with Eq. (4.19), a bias is applied to the defect-free $In_{0.53}Ga_{0.47}As$ structure along the [1 1 1]-direction, and injected excess carrier densities are then calculated from the occupations with the resulting electric field Fermi-shift. For the capture coefficients, the results in Figure 4.23 at a temperature of 300 K are used, together with a trap density of $N_t = 1 \times 10^{14}\,\mathrm{cm}^{-3}$. The resulting recombination rates are shown in Figure 4.24, as a function of the applied voltage and the related excess carrier densities. The non-linear inconsistent relationship between the excess carrier density and the applied voltage in Figure 4.24, is due to the complex density of states of the pristine $In_{0.53}Ga_{0.47}As$ structure, as shown in Figure 4.9. The results in Figure 4.24 can be explained by considering the process of SRH recombination. For SRH recombination to occur, the defects need to capture both an electron and a hole, in order for the recombination to take place. As such, if there is a large difference between the two capture rates of the defect, the slower capture rate will dominate the SRH recombination rate. This is clearly exemplified by the defect exhibiting the highest recombination rate in Figure 4.24, the In_{As}-As_{In} double antisite. This defect has a nearly identical hole capture coefficient as the Ga_{As}-As_{Ga} double antisite, which is also approximately as large as the electron capture coefficients of the As_{In} and As_{Ga} antisites, yet the electron capture coefficient is several orders of magnitude larger than that of the Ga_{As}-As_{Ga} double antisite or the hole capture coefficients of the As_{In} and As_{Ga} antisites, which therefore explains the much higher SRH recombination rate.

Since the capture rate is the product of the capture coefficient and the respective density of carriers, which rate is the slower one will depend not only on the details of the defect, in terms of the capture coefficients and the energy of the trap level, but also on the doping and state of operation of the semiconductor, as these aspects determine the free carrier densities. Of the investigated defects, in an undoped $In_{0.53}Ga_{0.47}As$ host at 300 K, only the tin substitution's Sn_{As} recombination rate is not solely determined by its capture coefficients.

In this defect, the two capture coefficients are very similar at 300 K, with the hole capture coefficient being only approximately an order of magnitude smaller than its electron capture coefficient. Therefore, the carrier densities only need to be imbalanced by an order of magnitude or more, in favour of the hole density, in order for the recombination in this defect to be determined by its electron capture coefficient. Due to the specifics of the density of states of $In_{0.53}Ga_{0.47}As$, the injected hole density is orders of magnitude larger than the injected electron density, in the investigated bias range. This imbalance is further reinforced, and even enforced at zero bias, by the tin substitution's trap level being closer to the valence band, thereby making the hole trap level carrier density p_t orders of magnitude larger than the electron trap level carrier density n_t. It is for this reason, that the tin substitution Sn_{As} has a larger recombination rate than the As_{Ga} antisite, as its recombination is determined by its electron capture coefficient, rather than its hole capture coefficient, which is larger than the hole capture coefficient of the As_{Ga} antisite.

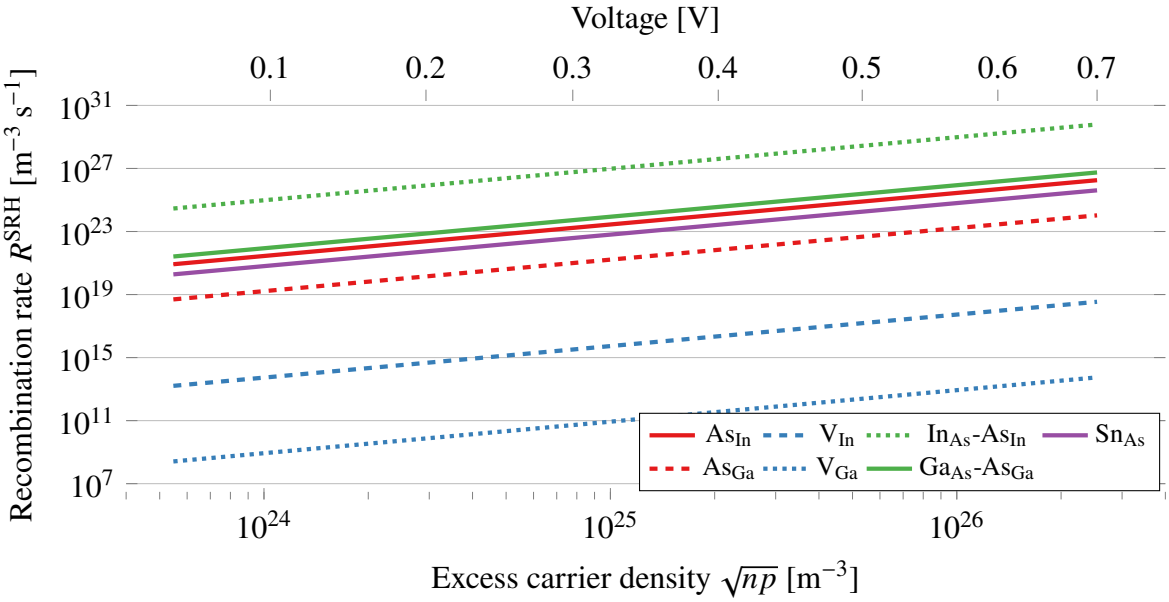


Figure 4.24: Shockley-Read-Hall recombination rate for deep trap levels of investigated point defects in $In_{0.53}Ga_{0.47}As$. Results are calculated in steady-state conditions under an applied bias, with a trap concentration of 1×10^{14} cm^{-3} and a temperature of 300 K.

For defects exhibiting multiple deep trap levels with associated localised single-particle defect states, it is possible to generalise the SRH relation in Eq. (4.19), to include multiple mid-gap transition levels for the recombination rate [155]. In such a model, depending on the speed of the various capture rates, the defects get stuck in either the uppermost or lowermost charge states, making only the nearest transition level relevant in the recombination rate, as

in the single trap level model. If the SRH model is further expanded to also include excited states of the defect, and if these excited states lie within the band gap, then the defects can capture into these levels and more efficiently escape the uppermost or lowermost charge states, thereby making the effects of including multiple trap levels important, which in turn will make the SRH rate increase substantially. Of the defects investigated in this thesis, only the studied vacancies exhibited multiple deep trap levels with distinctly associated local single-particle states. As such, expanding the SRH model to account for multiple trap levels was deemed unnecessary and out-of-scope in this thesis. It could however, for future studies, be interesting to investigate how much the SRH rate will be impacted in the relevant defects, such as the vacancies studied here, when multiple trap levels are accounted for.

4.6 Conclusions

In summary, this thesis chapter delved into the details of the research carried out on SRH recombination and trap levels in InGaAs, as was published in Ref. [112]. Two distinct methods to model random alloys in atomistic simulations were investigated, namely the Virtual Crystal Approximation (VCA) and Special Quasi-random Structures (SQSs), specifically to simulate the III-V compound $In_{0.53}Ga_{0.47}As$. It was found that the superior model was the SQS, as it was more physically correct as well as less computationally demanding, for the research tasks of this thesis. This SQS model was then utilised to study the transition levels in, and formation energies of, 23 different experimentally feasible intrinsic and extrinsic point defects in $In_{0.53}Ga_{0.47}As$. By comparing the calculated transition levels with experimentally measured levels, and by considering the formation energies of the various defects, it was concluded that the commonly occurring defect measured in experiments, must be one of the III-V antisites. To further study the detrimental impact on a device's properties due to these defects, the SRH recombination rate was then calculated in those defects which exhibited transition levels deep in the $In_{0.53}Ga_{0.47}As$ band gap. Of the 23 investigated defects, 9 of them had such transition levels and only 7 of those had well-defined associated localised single-particle states, suitable for the subsequent calculation of the associated SRH recombination. Of these seven defects, it was found that the defect exhibiting the highest SRH recombination rate in intrinsic $In_{0.53}Ga_{0.47}As$, is the In_{As}-As_{In} double antisite. The remaining defects, in descending order of recombination rate, were found to be the Ga_{As}-As_{Ga} double antisite, the As_{In} antisite, the Sn_{As} substitution, the As_{Ga} antisite, the V_{In} vacancy and finally the V_{Ga} vacancy.

Based on these results, it is evident that it is critical to minimise the formation of the antisites in devices incorporating InGaAs, in order to increase device performance which is limited by SRH recombination, such as the PIN photodiode in Section 1.4. This conclusion is further enhanced by the discovery that the SRH recombination is even stronger in the double antisites, which, despite having higher formation energies than the singular antisites, could form from energetically favourable combining events of singular antisites coinciding by defect diffusion.

Chapter 5

Outlook

The research carried out in this thesis can be continued in numerous ways, some of which are more straightforward and some of which holds more scientific potential. The overall topic of the DESIGN project, which this thesis is a part of, was narrowed down substantially from the general topic of defects in III-V semiconductors, by using the PIN diode from Section 1.4 as guidance. As such, a straightforward way to continue the current research, is simply to extend the studies in Chapter 3 and 4 to other scientifically relevant materials. In a similar fashion, the research can also be continued by extending the range of studied defects, or density of defects.

There are also many ways to continue the particular research carried out in this thesis, on polytypic defects in InP and SRH recombination in $In_{0.53}Ga_{0.47}As$. For the research on polytypic defects in InP, the next step would be to calculate the POP scattering rates in Section 3.5.2 using an *ab initio* method, rather than the simple model in Eq. (3.7), to get more accurate mobilities. Another point of interest would be to do other studies which relies on the polar optical phonons at the Γ-point, such as Raman spectra calculations, to further investigate the impact of the non-analytic discontinuity in the phonon dispersions caused by the Wurtzite segments.

The next steps in the $In_{0.53}Ga_{0.47}As$ studies would be to include dangling bonds in the list of studied defects. This can be done by creating a passivated cavity, as in Ref. [142], and then simply removing one of the passivating hydrogen atoms and treat the defect as a vacancy. Another very interesting way to continue this study, is to use the more advanced FNV correction scheme for interfaces [156], and then to study the antisites at interfaces with oxides. This could be particularly useful to study surface recombination and possibly investigate methods to reduce it or remove it completely, to huge benefit for the semiconductor device industry. It is also possible to make the specific SRH recombination studies in this thesis more accurate, by including multiple mid-gap transition levels for the SRH recombination rate [155], as well as excited states of the defect levels, which could be important for the studied vacancies and possibly other point defects. A final particularly interesting way to continue the $In_{0.53}Ga_{0.47}As$ SRH recombination research, is to include the SRH recombination in NEGF methods, in order to calculate the actual reduction in operating device current caused by SRH recombination.

Appendix A

Charged Point Defect Bandstructures

In order to facilitate identification of the pristine bulk bands in the defect configurations' bandstructures, the bandstructure of the pristine $In_{0.53}Ga_{0.47}As$ 6x6x6 SQS host is shown in Figure A.1.

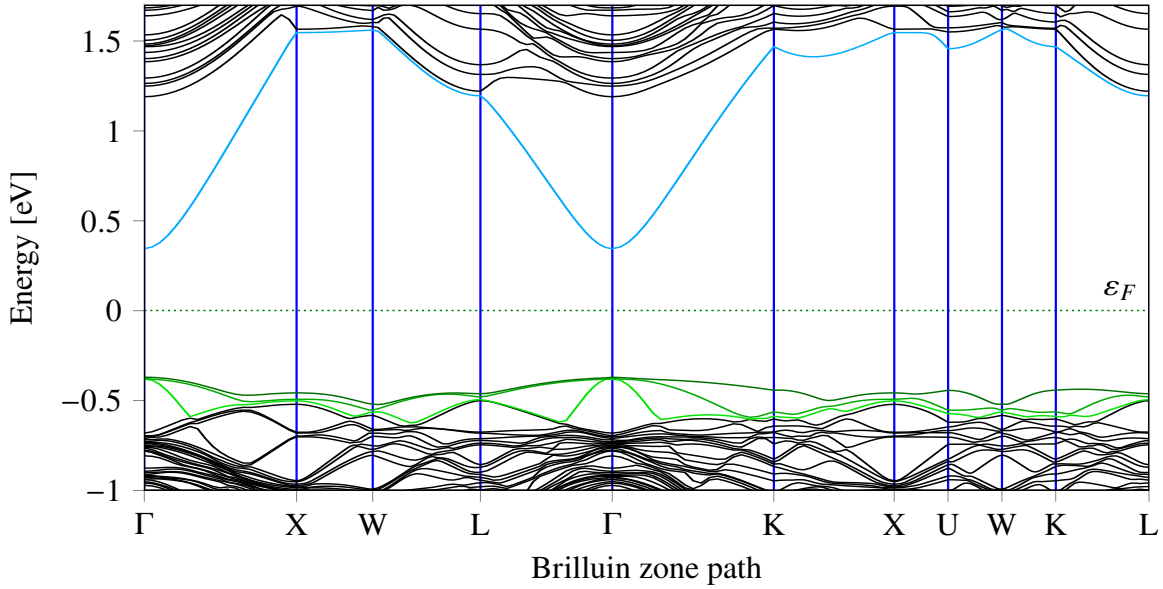


Figure A.1: Bandstructure of pristine 6x6x6 repetitions (432 atoms) SQS. The conduction band is given in blue and the valence bands in shades of green for the first, second and third bands.

In Figure A.2, the bandstructures of the neutral and -1 charge state of the V_{As}-V_{In} Schottky defect are shown. It can be seen that upon its partial occupation, the defect state strongly hybridised with the conduction band, de-hybridises (to some degree) and shifts downwards in energy towards the middle of the band gap. This de-hybridisation allows for the other hybridised defect state to hybridise more strongly with the conduction band, creating a new pseudo-conduction band. Additionally, another band, possibly another defect state, is shifted down below the speculated original conduction band, which is identified by its behaviour around the high-symmetry points, by comparison with the pristine conduction band in Figure A.1. The original valence bands were identified in a similar fashion.

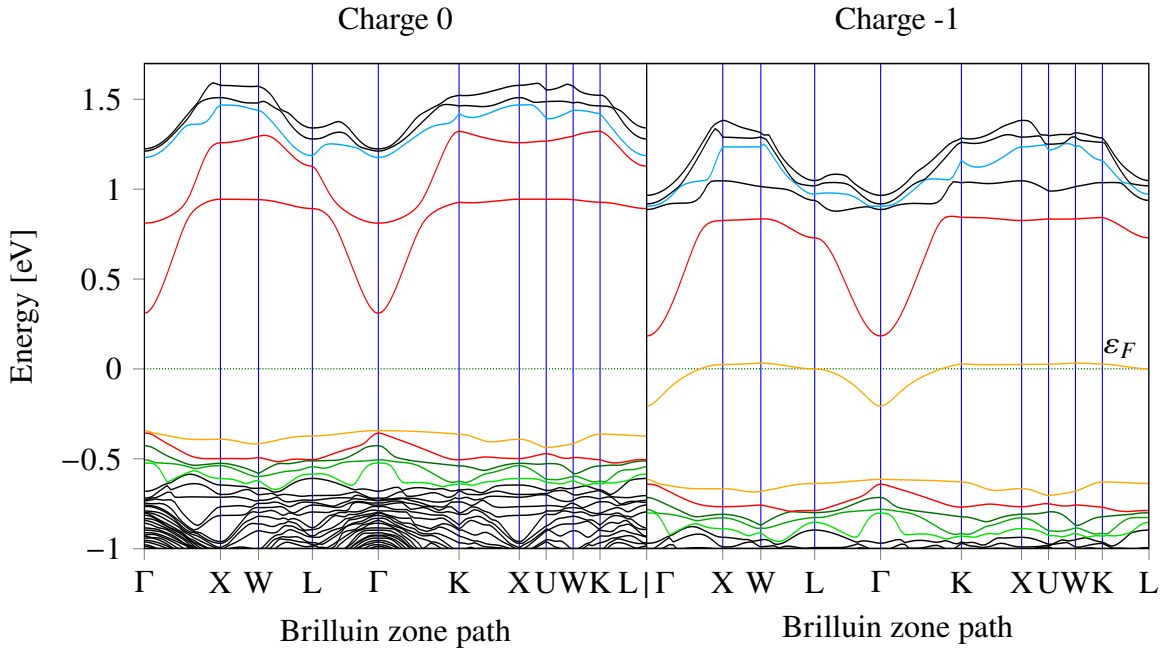


Figure A.2: Bandstructures of the neutral and -1 charge states of the V_{As}-V_{In} Schottky defect, embedded in the 6x6x6 repetitions (432 atoms) SQS. The conduction band is given in blue and the valence bands in shades of green for the first, second and third bands. The orange bands represent well-defined localised defect states and the red bands represent defect states highly hybridised with the periodic bulk bands.

In Figure A.3, the bandstructures of the neutral and -1 charge state of the other Schottky defect, V_{As}-V_{Ga}, are shown. In this defect, the hybridised defect state shifts so far downwards upon being partly occupied, that it disappears into the bulk valence bands and becomes fully occupied. In its stead, the previously occupied well-defined defect state near the valence bands becomes partly occupied. This phenomenological difference between the Schottky defects, explain the difference in their transitions levels observed in Figure 4.19.

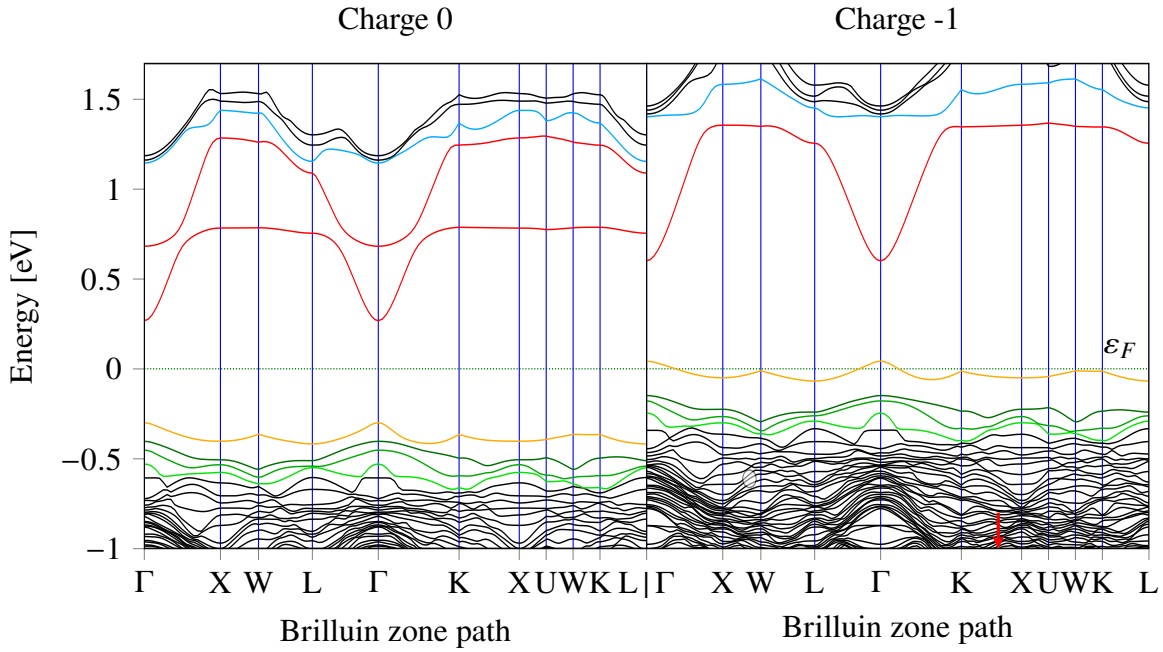


Figure A.3: Bandstructures of the neutral and -1 charge states of the V_{As}-V_{Ga} Schottky defect, embedded in the 6x6x6 repetitions (432 atoms) SQS. The conduction band is given in blue and the valence bands in shades of green for the first, second and third bands. The orange bands represent well-defined localised defect states and the red bands represent defect states highly hybridised with the periodic bulk bands.

Bibliography

[1] *Semiconductor industry association report: "state of the u.s. semiconductor industry 2023"*, https://www.semiconductors.org/2023-state-of-the-u-s-semiconductor-industry/ (visited on 01/14/2024).

[2] *Ieee computer society article: "as 5g approaches, semiconductor industry must combat friction points to make world of 'smart everything' a reality"*, https://www.computer.org/publications/tech-news/trends/5G-semiconductor-industry-drives-smart-everything-iot (visited on 01/14/2024).

[3] *Applied energy systems article: "navigating complexities: the semiconductor shortage's effect on the auto industry"*, https://www.appliedenergysystems.com/semiconductor-shortages-effect-on-auto-industry/ (visited on 01/14/2024).

[4] *Kpmg article: "russia-ukraine war: impact on the semiconductor industry"*, https://kpmg.com/ua/en/home/insights/2022/05/russia-ukraine-war-impact-semiconductor-industry.html (visited on 01/14/2024).

[5] *Jabil article: "why the chips are down: navigating the global chip shortages and beyond"*, https://www.jabil.com/blog/global-chip-shortages.html (visited on 01/14/2024).

[6] *Public discourse article: "the semiconductor industry and the future of the world economy (part i)"*, https://www.thepublicdiscourse.com/2023/02/87628/ (visited on 01/14/2024).

[7] *Deloitte article: "how semiconductor plays as a crucial element for economy"*, https://www2.deloitte.com/th/en/pages/about-deloitte/articles/semiconductor-eng.html (visited on 01/14/2024).

[8] *Intel article: "moore's law – now and in the future"*, https://www.intel.ie/content/www/ie/en/newsroom/opinion/moore-law-now-and-in-the-future.html (visited on 01/16/2024).

[9] G. E. Moore, "Cramming more components onto integrated circuits", Electronics **38**, 114 (1965).

[10] P. Gupta and E. Papadopoulou, "Yield analysis and optimization", 10.1201/9781420013481.ch37 (2008).

[11] S. Mahajan, "Growth- and processing-induced defects in semiconductors", Progress in Materials Science **33**, 1–84 (1989).

[12]D. Drabold and S. Estreicher, "Theory of defects in semiconductors", in , Vol. 104 (Jan. 2006), pp. 1–10.

[13]S. Mahajan, "Defects in semiconductors and their effects on devices", Acta Materialia **48**, 137–149 (2000).

[14]R. Geiger, J. Frigerio, M. J. Süess, D. Chrastina, G. Isella, R. Spolenak, J. Faist, and H. Sigg, "Excess carrier lifetimes in Ge layers on Si", Applied Physics Letters **104**, 062106 (2014).

[15]R. J. Roedel, A. R. V. Neida, R. Caruso, and L. R. Dawson, "The effect of dislocations in ga1-xalxas:si light-emitting diodes", Journal of The Electrochemical Society **126**, 637 (1979).

[16]T. Teraji, A. Fiori, N. Kiritani, S. Tanimoto, E. Gheeraert, and Y. Koide, "Mechanism of reverse current increase of vertical-type diamond Schottky diodes", Journal of Applied Physics **122**, 135304 (2017).

[17]M.-K. Lee and C.-F. Yen, "Characteristics of liquid phase deposited sio2 on (nh4)2s-treated gaas with an ultrathin si interface passivation layer", Japanese Journal of Applied Physics **53**, 056502 (2014).

[18]L. T. M. Hoa, T. Ouisse, D. Chaussende, M. Naamoun, A. Tallaire, and J. Achard, "Birefringence microscopy of unit dislocations in diamond", Crystal Growth & Design **14**, 5761–5766 (2014).

[19]M. A. Moram, R. A. Oliver, M. J. Kappers, and C. J. Humphreys, "The spatial distribution of threading dislocations in gallium nitride films", Advanced Materials **21**, 3941–3944 (2009).

[20]Y. Bioud, A. Boucherif, M. Myronov, S. Ali, G. Patriarche, N. Braidy, M. Jellite, D. Drouin, and R. Arès, "Uprooting defects to enable high-performance iii–v optoelectronic devices on silicon", Nature Communications **10**, 1–12 (2019).

[21]Y. Chen, T. Burgess, X. An, Y.-W. Mai, H. H. Tan, J. Zou, S. P. Ringer, C. Jagadish, and X. Liao, "Effect of a high density of stacking faults on the young's modulus of gaas nanowires", Nano Letters **16**, PMID: 26885570, 1911–1916 (2016).

[22]U. Krishnamachari, M. Borgstrom, B. J. Ohlsson, N. Panev, L. Samuelson, W. Seifert, M. W. Larsson, and L. R. Wallenberg, "Defect-free inp nanowires grown in [001] direction on inp (001)", Applied Physics Letters **85**, 2077–2079 (2004).

[23]H. J. Joyce, Q. Gao, H. H. Tan, C. Jagadish, Y. Kim, X. Zhang, Y. Guo, and J. Zou, "Twin-free uniform epitaxial gaas nanowires grown by a two-temperature process", Nano Letters **7**, PMID: 17335270, 921–926 (2007).

[24]H. Tahini, A. Chroneos, S. Murphy, U. Schwingenschlögl, and R. Grimes, "Vacancies and defect levels in iii–v semiconductors", Journal of Applied Physics **114**, 10.1063/1.4818484 (2013).

[25]H. Lee, Y. Cho, S. Jeon, and M. Shin, "First-principles-based quantum transport simulations of interfacial point defect effects on inas nanowire tunnel fets", IEEE Transactions on Electron Devices **68**, 5901–5907 (2021).

[26]B. Grandidier, H. Chen, R. M. Feenstra, D. T. McInturff, P. W. Juodawlkis, and S. E. Ralph, "Scanning tunneling microscopy and spectroscopy of arsenic antisites in low temperature grown InGaAs", Applied Physics Letters **74**, 1439–1441 (1999).

[27]J. Hoyt, H. Nayfeh, S. Eguchi, I. Aberg, G. Xia, T. Drake, E. Fitzgerald, and D. Antoniadis, "Strained silicon mosfet technology", in Digest. international electron devices meeting, (2002), pp. 23–26.

[28]T. Skotnicki and F. Boeuf, "How can high mobility channel materials boost or degrade performance in advanced cmos", in 2010 symposium on vlsi technology (2010), pp. 153–154.

[29]P. Ong and L. Teugels, "5 - cmp processing of high mobility channel materials: alternatives to si", in *Advances in chemical mechanical planarization (cmp)*, edited by S. Babu (Woodhead Publishing, 2016), pp. 119–135.

[30]C. Convertino, C. Zota, H. Schmid, D. Caimi, M. Sousa, K. Moselund, and L. Czornomaz, "Ingaas finfets directly integrated on silicon by selective growth in oxide cavities", Materials **12**, 10.3390/ma12010087 (2019).

[31]D.-H. Kim, J. A. del Alamo, P. Chen, W. Ha, M. Urteaga, and B. Brar, "50-nm e-mode in0.7ga0.3as phemts on 100-mm inp substrate with fmax > 1 thz", in 2010 international electron devices meeting (2010), pp. 30.6.1–30.6.4.

[32]D.-H. Kim and J. A. del Alamo, "30-nm inas phemts with f_T = 644 GHz and $f_{\max}$ = 681 GHz", IEEE Electron Device Letters **31**, 806–808 (2010).

[33]N. Sghaier, S. Bouzgarrou, M. B. Salem, A. Souifi, A. Kalboussi, and G. Guillot, "I-v anomalies on inalas/ingaas/inp hfets and deep levels investigations", Materials Science and Engineering: B **121**, 178–182 (2005).

[34]J. Alamo, "Nanometre-scale electronics with iii-v compound semiconductors", Nature **479**, 317–23 (2011).

[35]J. A. del Alamo, D. A. Antoniadis, J. Lin, W. Lu, A. Vardi, and X. Zhao, "Iii-v mosfets for future cmos", in 2015 ieee compound semiconductor integrated circuit symposium (csics) (2015), pp. 1–4.

[36]M. Heyns, "Ultimate scaling of cmos logic devices with ge and iii-v materials", MRS Bulletin **34**, 485–492 (2009).

[37]I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for III-V compound semiconductors and their alloys", Journal of Applied Physics **89**, 5815–5875 (2001).

[38]I. Vurgaftman and J. R. Meyer, "Band parameters for nitrogen-containing semiconductors", Journal of Applied Physics **94**, 3675–3696 (2003).

[39]A. De and C. E. Pryor, "Predicted band structures of iii-v semiconductors in the wurtzite phase", Phys. Rev. B **81**, 155210 (2010).

[40]S. Mokkapati and C. Jagadish, "Iii-v compound sc for optoelectronic devices", Materials Today **12**, 22–32 (2009).

[41]S. Sato and S. Satoh, "1.21 µm continuous-wave operation of highly strained gainas quantum well lasers on gaas substrates", Japanese Journal of Applied Physics **38**, L990 (1999).

[42]K. Wakita, Y. Kawamura, Y. Yoshikuni, H. Asahi, and S. Uehara, "High-speed long-wavelength optical modulation in ingaas/inalas multiple quantum wells", English, Electronics Letters **21**, 951–953(2) (1985).

[43]I. Fraunhofer and I. Fraunhofer, "Develops the world's most efficient solar cell with 47.6 percent efficiency", Fraunhofer Institute for Solar Energy Systems ISE: Freiburg im Breisgau, Germany (2022).

[44]R. R. King, D. C. Law, K. M. Edmondson, C. M. Fetzer, G. S. Kinsey, H. Yoon, R. A. Sherif, and N. H. Karam, "40% efficient metamorphic GaInP/GaInAs/Ge multijunction solar cells", Applied Physics Letters **90**, 183516 (2007).

[45]L. Shen, E. Y. B. Pun, and J. C. Ho, "Recent developments in iii–v semiconducting nanowires for high-performance photodetectors", Mater. Chem. Front. **1**, 630–645 (2017).

[46]S. Pleasants, "Leds: overcoming the 'green gap'", Nature Photonics **7**, 585– (2013).

[47]W. Liu, A. Y. Chang, R. D. Schaller, and D. V. Talapin, "Colloidal insb nanocrystals", Journal of the American Chemical Society **134**, PMID: 23198950, 20258–20261 (2012).

[48]P. Reiss, M. Carrière, C. Lincheneau, L. Vaure, and S. Tamang, "Synthesis of semiconductor nanocrystals, focusing on nontoxic and earth-abundant materials", Chemical Reviews **116**, PMID: 27391095, 10731–10819 (2016).

[49]G. Stillman, V. Robbins, and N. Tabatabaie, "Iii-v compound semiconductor devices: optical detectors", IEEE Transactions on Electron Devices **31**, 1643–1655 (1984).

[50]G. E. Stillman, C. M. Wolfe, A. G. Foyt, and W. T. Lindley, "Schottky barrier InxGa1-xAs alloy avalanche photodiodes for 1.06 µm", Applied Physics Letters **24**, 8–10 (2003).

[51]B. Bahari, A. Ndao, F. Vallini, A. E. Amili, Y. Fainman, and B. Kanté, "Nonreciprocal lasing in topological cavities of arbitrary geometries", Science **358**, 636–640 (2017).

[52]H. Zhao, P. Miao, M. H. Teimourpour, S. Malzard, R. El-Ganainy, H. Schomerus, and L. Feng, "Topological hybrid silicon microlasers", Nature Communications **9**, `10.1038/s41467-018-03434-2` (2018).

[53]N. Materise, M. C. Dartiailh, W. M. Strickland, J. Shabani, and E. Kapit, "Tunable capacitor for superconducting qubits using an inas/ingaas heterostructure", Quantum Science and Technology **8**, 045014 (2023).

[54]*Design-eid project webpage*, `https://www.gla.ac.uk/research/az/design-eid/` (visited on 01/16/2024).

[55]G. Dummer, *Electronic inventions and discoveries: electronics from its earliest beginnings to the present day*, Pergamon International Library of Science, Technology, Engin (Elsevier Science, 2013).

[56]N. Storey, *Electronics: a systems approach*, Electronic systems engineering series (Pearson, 2013).

[57]K. K. Ng, "Pin photodiode", in *Complete guide to semiconductor devices* (2002), pp. 431–437.

[58]I. Kimukin, N. Biyikli, B. Butun, O. Aytur, S. Unlu, and E. Ozbay, "Ingaas-based high-performance p-i-n photodiodes", IEEE Photonics Technology Letters **14**, 366–368 (2002).

[59]Z. Li, J. Allen, M. Allen, H. H. Tan, C. Jagadish, and L. Fu, "Review on iii-v semiconductor single nanowire-based room temperature infrared photodetectors", Materials **13**, 10.3390/ma13061400 (2020).

[60]C. Martinez-Oliver, K. E. Moselund, and V. P. Georgiev, "Evaluation of material profiles for iii-v nanowire photodetectors", in 2021 international conference on numerical simulation of optoelectronic devices (nusod) (2021), pp. 39–40.

[61]T. K. Oanh Vu, M. T. Tran, N. X. Tu, N. T. Thanh Bao, and E. K. Kim, "Electronic transport mechanism and defect states for p-inp/i-ingaas/n-inp photodiodes", Journal of Materials Research and Technology **19**, 2742–2749 (2022).

[62]Y. Ito, R. Yokogawa, O. Ueda, N. Sawamoto, K. Ide, L. Men, and A. Ogura, "Analysis of InGaAs/InP p-I-n Photodiode Failed by Electrostatic Discharge", Journal of Electronic Materials **52**, 5150–5158 (2023).

[63]X. D. Wang, W. D. Hu, X. S. Chen, W. Lu, H. J. Tang, T. Li, and H. M. Gong, "Dark current simulation of inp/ingaas/inp p-i-n photodiode", in 2008 international conference on numerical simulation of optoelectronic devices (nusod) (2008), pp. 31–32.

[64]J. Kohanoff, *Electronic structure calculations for solids and molecules: theory and computational methods* (Cambridge University Press, 2006).

[65]M. Born and R. Oppenheimer, "Zur quantentheorie der molekeln", Annalen der Physik **389**, 457–484 (1927).

[66]F. Bloch, "Über die quantenmechanik der elektronen in kristallgittern", Zeitschrift für Physik **52**, 555–600 (1929).

[67]P. Hohenberg and W. Kohn, "Inhomogeneous electron gas", Phys. Rev. **136**, B864–B871 (1964).

[68]W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects", Phys. Rev. **140**, A1133–A1138 (1965).

[69]J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems", Phys. Rev. B **23**, 5048–5079 (1981).

[70]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple", Phys. Rev. Lett. **77**, 3865–3868 (1996).

[71]L. G. Ferreira, M. Marques, and L. K. Teles, "Approximation to density functional theory for the calculation of band gaps of semiconductors", Phys. Rev. B **78**, 125116 (2008).

[72]J. Sun, A. Ruzsinszky, and J. P. Perdew, "Strongly constrained and appropriately normed semilocal density functional", Phys. Rev. Lett. **115**, 036402 (2015).

[73]F. Tran and P. Blaha, "Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential", Phys. Rev. Lett. **102**, 226401 (2009).

[74]A. D. Becke and M. R. Roussel, "Exchange holes in inhomogeneous systems: a coordinate-space model", Phys. Rev. A **39**, 3761–3767 (1989).

[75]J. Heyd, G. E. Scuseria, and M. Ernzerhof, "Hybrid functionals based on a screened Coulomb potential", The Journal of Chemical Physics **118**, 8207–8215 (2003).

[76]S. Lehtola, C. Steigemann, M. J. Oliveira, and M. A. Marques, "Recent developments in libxc — a comprehensive library of functionals for density functional theory", SoftwareX **7**, 1–5 (2018).

[77]N. Marzari, D. Vanderbilt, A. De Vita, and M. C. Payne, "Thermal contraction and disordering of the al(110) surface", Phys. Rev. Lett. **82**, 3296–3299 (1999).

[78]H. Bruus and K. Flensberg, *Many-body quantum theory in condensed matter physics - an introduction*, English (Oxford University Press, United States, 2004).

[79]M. Paulsson, *Non equilibrium green's functions for dummies: introduction to the one particle negf equations*, 2006.

[80]M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, "Density-functional method for nonequilibrium electron transport", Physical Review B **65**, 165401 (2002).

[81]G. Onida, L. Reining, and A. Rubio, "Electronic excitations: density-functional versus many-body green's-function approaches", Rev. Mod. Phys. **74**, 601–659 (2002).

[82]C. Vedel, S. Smidstrup, and V. Georgiev, "First-principles investigation of polytypic defects in inp", Scientific Reports **12**, 10.1038/s41598-022-24239-w (2022).

[83]S. Smidstrup, T. Markussen, P. Vancraeyveld, J. Wellendorff, J. Schneider, T. Gunst, B. Verstichel, D. Stradi, P. A. Khomyakov, U. G. Vej-Hansen, et al., "Quantumatk: an integrated platform of electronic and atomic-scale modelling tools", J. Phys: Condens. Matter **32**, 015901 (2020).

[84]M. van Setten, M. Giantomassi, E. Bousquet, M. Verstraete, D. Hamann, X. Gonze, and G.-M. Rignanese, "The pseudodojo: training and grading a 85 element optimized norm-conserving pseudopotential table", Computer Physics Communications **226**, 39–54 (2018).

[85]D. Stradi, U. Martinez, A. Blom, M. Brandbyge, and K. Stokbro, "General atomistic approach for modeling metal-semiconductor interfaces using density functional theory and nonequilibrium green's function", Phys. Rev. B **93**, 155302 (2016).

[86]T. Enoki, K. Arai, A. Kohzen, and Y. Ishii, "Design and characteristics of ingaas/inp composite-channel hfet's", IEEE Transactions on Electron Devices **42**, 1413–1418 (1995).

[87]M. Stepanenko, I. Yunusov, V. Arykov, P. Troyan, and Y. Zhidik, "Multi-parameter optimization of an inp electro-optic modulator", Symmetry **12**, 10.3390/sym12111920 (2020).

[88]J. Klamkin, H. Zhao, B. Song, Y. Liu, B. Isaac, S. Pinna, F. Sang, and L. Coldren, "Indium phosphide photonic integrated circuits: technology and applications", in 2018 ieee bicmos and compound semiconductor integrated circuits and technology symposium (bcicts) (2018), pp. 8–13.

[89]M. Smit, X. Leijtens, H. Ambrosius, E. Bente, J. van der Tol, B. Smalbrugge, T. de Vries, E.-J. Geluk, J. Bolk, R. van Veldhoven, L. Augustin, P. Thijs, D. D'Agostino, H. Rabbani, K. Lawniczuk, S. Stopinski, S. Tahvili, A. Corradi, E. Kleijn, D. Dzibrou, M. Felicetti, E. Bitincka, V. Moskalenko, J. Zhao, R. Santos, G. Gilardi, W. Yao, K. Williams, P. Stabile, P. Kuindersma, J. Pello, S. Bhat, Y. Jiao, D. Heiss, G. Roelkens, M. Wale, P. Firth, F. Soares, N. Grote, M. Schell, H. Debregeas, M. Achouche, J.-L. Gentner, A. Bakker, T. Korthorst, D. Gallagher, A. Dabbs, A. Melloni, F. Morichetti, D. Melati, A. Wonfor, R. Penty, R. Broeke, B. Musk, and D. Robbins, "An introduction to InP-based generic integration technology", Semiconductor Science and Technology **29**, 083001 (2014).

[90]Y. Jiao, N. Nishiyama, J. van der Tol, J. van Engelen, V. Pogoretskiy, S. Reniers, A. A. Kashi, Y. Wang, V. D. Calzadilla, M. Spiegelberg, Z. Cao, K. Williams, T. Amemiya, and S. Arai, "InP membrane integrated photonics research", Semiconductor Science and Technology **36**, 013001 (2020).

[91]M. Smit, K. Williams, and J. van der Tol, "Past, present, and future of InP-based photonic integration", APL Photonics **4**, 050901 (2019).

[92]Y. Han and L. Lin, "New insight into the origin of twin and grain boundary in inp", Solid State Communications **110**, 403–406 (1999).

[93]J. Cheng, L. Largeau, G. Patriarche, P. Regreny, G. Hollinger, and G. Saint-Girons, "Twin formation during the growth of InP on SrTiO3", Applied Physics Letters **94**, 231902 (2009).

[94]P. Staudinger, S. Mauthe, K. E. Moselund, and H. Schmid, "Concurrent zinc-blende and wurtzite film formation by selection of confined growth planes", Nano Letters **18**, PMID: 30427685, 7856–7862 (2018).

[95]C. Stampfl and C. G. Van de Walle, "Energetics and electronic structure of stacking faults in aln, gan, and inn", Phys. Rev. B **57**, R15052–R15055 (1998).

[96]Y. Rebane, Y. Shreter, and M. Albrecht, "Stacking faults as quantum wells for excitons in wurtzite gan", physica status solidi (a) **164**, 141–144 (1997).

[97]P. Corfdir, C. Hauswald, J. K. Zettler, T. Flissikowski, J. Lähnemann, S. Fernández-Garrido, L. Geelhaar, H. T. Grahn, and O. Brandt, "Stacking faults as quantum wells in nanowires: density of states, oscillator strength, and radiative efficiency", Phys. Rev. B **90**, 195309 (2014).

[98]K. Korona, A. Reszka, M. Sobanska, P. Perkowska, A. Wysmołek, K. Klosek, and Z. Zytkiewicz, "Dynamics of stacking faults luminescence in gan/si nanowires", Journal of Luminescence **155**, 293–297 (2014).

[99]M. Albrecht, S. Christiansen, G. Salviati, C. Zanotti Fregonara, Y. Rebane, Y. Shreter, M. Mayer, A. Pelzmann, M. Kamp, K. Ebeling, M. Bremser, R. Davis, and H. Strunk, "Luminescence related to stacking faults in heterepitaxially grown wurtzite gan", MRS Online Proceeding Library Archive **468**, 293 (1997).

[100]J. Bao, D. C. Bell, F. Capasso, J. B. Wagner, T. Mårtensson, J. Trägårdh, and L. Samuelson, "Optical properties of rotationally twinned inp nanowire heterostructures", Nano Letters **8**, PMID: 18275163, 836–841 (2008).

[101]L. Pavesi, F. Piazza, A. Rudra, J. F. Carlin, and M. Ilegems, "Temperature dependence of the inp band gap from a photoluminescence study", Phys. Rev. B **44**, 9052–9055 (1991).

[102]*Ioffe material parameters database*, `http://www.ioffe.ru/SVA/NSM/Semicond/` (visited on 07/25/2023).

[103]D. Kriegner, E. Wintersberger, K. Kawaguchi, J. Wallentin, M. T. Borgström, and J. Stangl, "Unit cell parameters of wurtzite inp nanowires determined by x-ray diffraction", Nanotechnology **22**, 425704 (2011).

[104]G. Signorello, E. Lörtscher, P. Khomyakov, S. Karg, D. Dheeraj, B. Gotsmann, H. Weman, and H. Riel, "Inducing a direct-to-pseudodirect bandgap transition in wurtzite gaas nanowires with uniaxial stress", Nature communications **5**, 3655 (2014).

[105]Y. Li, W. Kerr, and N. Holzwarth, "Continuity of phonon dispersion curves in layered ionic materials", Journal of Physics: Condensed Matter **32**, `10.1088/1361-648X/ab4cc1` (2019).

[106]y. Wang, S.-L. Shang, H. Fang, Z.-K. Liu, and L.-Q. Chen, "First-principles calculations of lattice dynamics and thermal properties of polar solids", npj Computational Materials **2**, 16006 (2016).

[107]A. Fox, *Optical properties of solids*, Oxford master series in condensed matter physics (Oxford University Press, 2001).

[108]L. Dacal and A. Cantarero, "An ab initio study of the polytypism in inp", Scientific Reports **6**, 33914 (2016).

[109]Heinz Schmid, Personal communication, Senior Engineer at IBM Research Europe - Zürich, 2020.

[110]M. Lundstrom, *Fundamentals of carrier transport*, 2nd ed. (Cambridge University Press, 2000).

[111]G. K. Madsen and D. J. Singh, "Boltztrap. a code for calculating band-structure dependent quantities", Computer Physics Communications **175**, 67–71 (2006).

[112]C. D. Vedel, T. Gunst, S. Smidstrup, and V. P. Georgiev, "Shockley-read-hall recombination and trap levels in $In_{0.53}Ga_{0.47}As$ point defects from first principles", Phys. Rev. B **108**, 094113 (2023).

[113]T. Pearsall and M. Pollack, "Chapter 2 compound semiconductor photodiodes", in *Lightwave communications technology*, Vol. 22, edited by W. Tsang, Semiconductors and Semimetals (Elsevier, 1985), pp. 173–245.

[114]N. V. Kryzhanovskaya, E. I. Moiseev, Y. S. Polubavkina, M. V. Maximov, M. M. Kulagina, S. I. Troshkov, Y. M. Zadiranov, A. A. Lipovskii, N. V. Baidus, A. A. Dubinov, Z. F. Krasilnik, A. V. Novikov, D. A. Pavlov, A. V. Rykov, A. A. Sushkov, D. V. Yurasov, and A. E. Zhukov, "Electrically pumped ingaas/gaas quantum well microdisk lasers directly grown on si(100) with ge/gaas buffer", Opt. Express **25**, 16754–16760 (2017).

[115]F. Dimroth, M. Grave, P. Beutel, U. Fiedeler, C. Karcher, T. N. D. Tibbits, E. Oliva, G. Siefer, M. Schachtner, A. Wekkeli, A. W. Bett, R. Krause, M. Piccin, N. Blanc, C. Drazek, E. Guiot, B. Ghyselen, T. Salvetat, A. Tauzin, T. Signamarcheix, A. Dobrich, T. Hannappel, and K. Schwarzburg, "Wafer bonded four-junction gainp/gaas//gainasp/gainas concentrator solar cells with 44.7% efficiency", Progress in Photovoltaics: Research and Applications **22**, 277–282 (2014).

[116]J. Robertson, Y. Guo, and L. Lin, "Defect state passivation at III-V oxide interfaces for complementary metal–oxide–semiconductor devices", Journal of Applied Physics **117**, 112806 (2015).

[117]L. Zhou, B. Bo, X. Yan, C. Wang, Y. Chi, and X. Yang, "Brief review of surface passivation on iii-v semiconductor", Crystals **8**, 10.3390/cryst8050226 (2018).

[118]W. Melitz, E. Chagarov, T. Kent, R. Droopad, J. Ahn, R. Long, P. C. McIntyre, and A. C. Kummel, "Mechanism of dangling bond elimination on as-rich ingaas surface", in 2012 international electron devices meeting (2012), pp. 32.4.1–32.4.4.

[119]H.-P. Komsa and A. Pasquarello, "Comparison of vacancy and antisite defects in gaas and ingaas through hybrid functionals", Journal of Physics: Condensed Matter **24**, 045801 (2012).

[120]H.-P. Komsa and A. Pasquarello, "Intrinsic defects in gaas and ingaas through hybrid functional calculations", Physica B: Condensed Matter **407**, 26th International Conference on Defects in Semiconductors, 2833–2837 (2012).

[121]G. Brammertz, H. Lin, K. Martens, A.-R. Alian, C. Merckling, J. Penaud, D. Kohen, W.-E. Wang, S. Sioncke, A. Delabie, M. Meuris, M. R. Caymax, and M. Heyns, "Electrical properties of iii-v/oxide interfaces", ECS Transactions **19**, 375 (2009).

[122]É. O'Connor, B. Brennan, V. Djara, K. Cherkaoui, S. Monaghan, S. B. Newcomb, R. Contreras, M. Milojevic, G. Hughes, M. E. Pemble, R. M. Wallace, and P. K. Hurley, "A systematic study of (NH4)2S passivation (22%, 10%, 5%, or 1%) on the interface properties of the Al2O3/In0.53Ga0.47As/InP system for n-type and p-type In0.53Ga0.47As epitaxial layers", Journal of Applied Physics **109**, 024101 (2011).

[123]W. Wang, C. Gong, B. Shan, R. M. Wallace, and K. Cho, "Sulfur passivation effect on HfO2/GaAs interface: A first-principles study", Applied Physics Letters **98**, 232113, 10.1063/1.3597219 (2011).

[124]L.-Y. Deng and D. Bowman, "Developments in pseudo-random number generators", WIREs Computational Statistics **9**, e1404 (2017).

[125]Y. Li, F. Yangyang, W. Wang, X. Meng, H. Wang, and Q. Duan, "Quantum random number generator using a cloud superconducting quantum computer based on source-independent protocol", Scientific Reports **11**, 10.1038/s41598-021-03286-9 (2021).

[126]A. Zunger, S.-H. Wei, L. G. Ferreira, and J. E. Bernard, "Special quasirandom structures", Phys. Rev. Lett. **65**, 353–356 (1990).

[127]T. P. Pearsall, R. Bisaro, R. Ansel, and P. Merenda, "The growth of GaxIn1-xAs on (100) InP by liquid-phase epitaxy", Applied Physics Letters **32**, 497–499 (1978).

[128]M. Guidon, J. Hutter, and J. VandeVondele, "Auxiliary density matrix methods for hartree-fock exchange calculations", Journal of Chemical Theory and Computation **6**, PMID: 26613491, 2348–2364 (2010).

[129]M. W. Haverkort, I. S. Elfimov, and G. A. Sawatzky, *Electronic structure and self energies of randomly substituted solids using density functional theory and model calculations*, 2011.

[130]V. Popescu and A. Zunger, "Effective band structure of random alloys", Phys. Rev. Lett. **104**, 236403 (2010).

[131]V. Popescu and A. Zunger, "Extracting E versus $\vec{k}$ effective band structure from supercell calculations on alloys and impurities", Phys. Rev. B **85**, 085201 (2012).

[132]A. van de Walle, P. Tiwary, M. de Jong, D. Olmsted, M. Asta, A. Dick, D. Shin, Y. Wang, L.-Q. Chen, and Z.-K. Liu, "Efficient stochastic generation of special quasirandom structures", Calphad **42**, 13–18 (2013).

[133]C. Freysoldt, J. Neugebauer, and C. G. Van de Walle, "Fully ab initio finite-size corrections for charged-defect supercell calculations", Phys. Rev. Lett. **102**, 016402 (2009).

[134]C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, and C. G. Van de Walle, "First-principles calculations for point defects in solids", Rev. Mod. Phys. **86**, 253–305 (2014).

[135]S. E. Taylor and F. Bruneval, "Understanding and correcting the spurious interactions in charged supercells", Phys. Rev. B **84**, 075155 (2011).

[136]D. Cui, S. M. Hubbard, D. Pavlidis, A. Eisenbach, and C. Chelli, "Impact of doping and mocvd conditions on minority carrier lifetime of zinc- and carbon-doped ingaas and its applications to zinc- and carbon-doped inp/ingaas heterostructure bipolar transistors", Semiconductor Science and Technology **17**, 503 (2002).

[137]H. Künzel, J. Böttcher, R. Gibis, and G. Urmann, "Material properties of Ga0.47In0.53As grown on InP by low-temperature molecular beam epitaxy", Applied Physics Letters **61**, 1347–1349 (1992).

[138]Y. Hwang, R. Engel-Herbert, N. G. Rudawski, and S. Stemmer, "Analysis of trap state densities at HfO2/In0.53Ga0.47As interfaces", Applied Physics Letters **96**, 102910 (2010).

[139]S. Kugler, K. Steiner, U. Seiler, K. Heime, and E. Kuphal, "Low-frequency noise measurements on n-InGaAs/p-InP junction field-effect transistor structures", Applied Physics Letters **52**, 111–113 (1988).

[140]Y. Takanashi and N. Kondo, "Deep trap in InGaAs grown by gas source molecular beam epitaxy", Journal of Applied Physics **85**, 633–634 (1999).

[141]P. S. Whitney, W. Lee, and C. G. Fonstad, "Capacitance transient analysis of molecular-beam epitaxial n-In0.53Ga0.47As and n-In0.52Al0.48 As", Journal of Vacuum Science & Technology B: Microelectronics Processing and Phenomena **5**, 796–799 (1987).

[142]H.-P. Komsa and A. Pasquarello, "Dangling bond charge transition levels in AlAs, GaAs, and InAs", Applied Physics Letters **97**, 191901, 10.1063/1.3515422 (2010).

[143]J. Robertson and L. Lin, "Bonding principles of passivation mechanism at III-V-oxide interfaces", Applied Physics Letters **99**, 222906, 10.1063/1.3665061 (2011).

[144]W. Shockley and W. T. Read, "Statistics of the recombinations of holes and electrons", Phys. Rev. **87**, 835–842 (1952).

[145]A. Alkauskas, Q. Yan, and C. G. Van de Walle, "First-principles theory of nonradiative carrier capture via multiphonon emission", Phys. Rev. B **90**, 075202 (2014).

[146]R. Pässler, "Calculation of nonradiative multiphonon capture coefficients and ionization rates for neutral centres according to the static coupling scheme: i. theory", physica status solidi (b) **68**, 69–79 (1975).

[147]A. Alkauskas, J. L. Lyons, D. Steiauf, and C. G. Van de Walle, "First-principles calculations of luminescence spectrum line shapes for defects in semiconductors: the example of gan and zno", Phys. Rev. Lett. **109**, 267401 (2012).

[148]A. Matulionis, J. Liberis, I. Matulionienė, M. Ramonas, E. Šermukšnis, J. H. Leach, M. Wu, X. Ni, X. Li, and H. Morkoç, "Plasmon-enhanced heat dissipation in GaN-based two-dimensional channels", Applied Physics Letters **95**, 192102 (2009).

[149]K. Huang, A. Rhys, and N. F. Mott, "Theory of light absorption and non-radiative transitions in f-centres", Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **204**, 406–423 (1950).

[150]M. E. Turiansky, A. Alkauskas, M. Engel, G. Kresse, D. Wickramaratne, J.-X. Shen, C. E. Dreyer, and C. G. Van de Walle, "Nonrad: computing nonradiative capture coefficients from first principles", Computer Physics Communications **267**, 108056 (2021).

[151]R. Pässler, "Relationships between the nonradiative multiphonon carrier-capture properties of deep charged and neutral centres in semiconductors", physica status solidi (b) **78**, 625–635 (1976).

[152]S. I. Simdyankin, S. R. Elliott, Z. Hajnal, T. A. Niehaus, and T. Frauenheim, "Simulation of physical properties of the chalcogenide glass As_2S_3 using a density-functional-based tight-binding method", Phys. Rev. B **69**, 144202 (2004).

[153]V. Abakumov, V. Perel, and I. Yassievich, *Nonradiative recombination in semiconductors*, ISSN (Elsevier Science, 1991).

[154]R. N. Hall, "Electron-hole recombination in germanium", Phys. Rev. **87**, 387–387 (1952).

[155]A. Alkauskas, C. E. Dreyer, J. L. Lyons, and C. G. Van de Walle, "Role of excited states in shockley-read-hall recombination in wide-band-gap semiconductors", Phys. Rev. B **93**, 201304 (2016).

[156]C. Freysoldt and J. Neugebauer, "First-principles calculations for charged defects at surfaces, interfaces, and two-dimensional materials in the presence of electric fields", Phys. Rev. B **97**, 205425 (2018).